Extended Formulation Lower Bounds for Combinatorial Optimization

Jonah Brown-Cohen

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2018-94
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2018/EECS-2018-94.html

July 30, 2018

Extended Formulation Lower Bounds for Combinatorial Optimization

by

Jonah Brown-Cohen

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Prasad Raghavendra, Chair
Professor Satish Rao
Professor Luca Trevisan
Assistant Professor Nikhil Srivastava

Summer 2018

Extended Formulation Lower Bounds for Combinatorial Optimization

Abstract

Extended Formulation Lower Bounds for Combinatorial Optimization

by

Jonah Brown-Cohen

Doctor of Philosophy in Computer Science

University of California, Berkeley

Associate Professor Prasad Raghavendra, Chair

Linear and semidefinite programs are fundamental algorithmic tools, often providing conjecturally optimal results for a variety of combinatorial optimization problems. Thus, a natural question is to understand the limitations of linear and semidefinite programming relaxations. In particular, the goal is to prove unconditional lower bounds on the size of any linear or semidefinite programming relaxation for a given problem.

In this dissertation, I will give two results of this flavor. First, I will show that any linear programming relaxation for refuting random instances of constraint satisfaction problems (e.g. k-SAT) requires super-polynomial size. This theorem can be understood as evidence that refuting CSPs is hard, since it rules out a broad class of algorithms. Second, I will show that any symmetric semidefinite programming relaxation for the matching problem in general graphs requires exponential size. Since there is a polynomial time algorithm for the matching problem, this result provides an example of the limitations of semidefinite programming relaxations.

To my family.

Contents

Acknowledgments

This dissertation, along with all the rest of my work as a graduate student, would not have been possible without the support, advice and friendship of many wonderful people. I would like to thank and acknowledge:

My advisor, Prasad Raghavendra, for teaching me how to think about theory, as well as for his patience, kindness and generosity. His advice provided a guide and his work was an inspiration.

My thesis committee, Satish Rao, Nikhil Srivastava and Luca Trevisan, for serving on my committee and for everything I have learned from them over the years. In particular, I would like to thank Satish for our many conversations about teaching, and about everything else.

The other faculty members of the Berkeley theory group, who created such a great environment for my graduate studies. I would especially like to thank Christos Papadimitriou, for advising me during my first semester and for convincing me to come to Berkeley in the first place.

My collaborators, Gabor Braun, Arefin Huq, Sebastian Pokutta, Prasad Raghavendra, Aurko Roy, Benjamin Weitz, Daniel Zink, Arvind Narayanan, Christos Alexandros Psomas, S. Matthew Weinberg and Raghu Meka, for all their hard work.

My fellow students, who were such a large part of my academic and social life during my time at Berkeley. My sincerest thanks to: Piyush Srivastava, Anindya De, Urmila Mahadev, Seung Woo Shin, Di Wang, Sam Wong, Siu Man Chan, Siu On Chan, James Cook, Rafael Frongillo, George Pierrakos Frank Ban, Lynn Chua, Grace Dinh, Arun Ganesh, Fotis Iliopoulos, Tarun Kathuria, Marc Khoury, Rachel Lawrence, Jingcheng Liu, Pasin Manurangsi, Peihan Miao, Jarrod Millman, Chinmay Nirkhe, Manuel Sabin, Aaron Schild, Nick Spooner, Akshayaram Srinivasan, Elizabeth Yang, and Morris Yau.

The other students who began at Berkeley in the same year that I did, Aviad Rubinstein, Benjamin Weitz, Jarett Schwartz, Alexandros Psomas, Tselil Schramm, Ning Tan and Paul Christiano, for going through everything together.

Hrysoula Papadakis, for being with me the whole time, and for always being on the up-and-up and encouraging me to do the same.

My family: my brother and sister Lawry and Moana, for growing up with me, and my parents, Laura and Walter, for raising me and supporting me in everything I do.

Chapter 1

Introduction

One of the most powerful algorithmic paradigms in combinatorial optimization is that of convex relaxation. Algorithms based on convex relaxations have been successfully applied to many fundamental problems including maximum flow, bipartite matching, vertex cover, the traveling salesperson problem, and decoding error correcting codes. In the case of constraint satisfaction problems, semidefinite programming relaxations have been shown to give optimal approximation algorithms assuming the Unique Games Conjecture [Rag08].

In all these cases, a natural goal is to understand the precise power and limits of convex relaxations in solving particular classes of problems. One approach, as exemplified in [Rag08], is to use NP-hardness or Unique Games-hardness reductions to prove some upper limit on the performance of any polynomial time algorithm. Then, if a given convex relaxation achieves this limit, we can indeed say that it performs optimally. This approach relies fundamentally on unproven complexity assumptions, though of course many of these are quite reasonable.

An alternative method for understanding the power of convex relaxations is to restrict the space of potential algorithms to a particular class of relaxations, and then prove that no polynomial-size relaxation in this class can achieve performance above some threshold on a given problem. Such lower bounds on the size of relaxations are *unconditional*, meaning they do not depend on any unproven complexity assumption. Further, they often allow us to more deeply understand the structure of convex relaxations.

A recent line of work beginning with [FKPT12, FMP$^+$12], and based on the original paper of Yannakakis [Yan88], has developed a method to prove such results for the very broad classes of linear or semidefinite programming relaxations. The terminology in these originating papers refers to such results as extended formulation lower bounds. In this thesis I will prove lower bounds on the size of linear and semidefinite programs for a few well-studied classes of problems.

1.1 Convex Relaxations

Convex relaxations generally take the following form. First, choose some combinatorial optimization problem to be solved over a discrete domain (e.g. SAT, knapsack, the traveling salesperson problem, graph matching). Second, *relax* the domain of the problem from a discrete set to a larger continuous, convex set, and extend the objective function to this larger set. Third, solve the relaxed optimization problem, and use some method to translate this continuous solution back into a discrete solution for the original problem. We begin by introducing the basic setup for relaxations of combinatorial optimization problems.

Combinatorial Optimization

A typical instance of combinatorial optimization problem is comprised of an objective function f and a discrete domain of possible solutions S. The goal is to maximize $f(s)$ over elements $s \in S$ from the solution domain. We will write $\mathrm{opt}(f) \overset{\mathrm{def}}{=} \max_{s \in S} f(s)$ for this maximum. Standard examples include MAX k-SAT and MATCHING.

Example 1.1.1 (MAX k-SAT)**.** An instance Φ of MAX k-SAT is a collection of m constraints $C_1, \dots C_m$ given by boolean disjunctions of k literals (variables or negated variables) from $x \in \{0,1\}^n$. Some example constraints for the $k = 3$ case are

$$\begin{aligned} C_1 &= (x_1 \vee \neg x_2 \vee \neg x_3) \\ C_2 &= (\neg x_7 \vee x_4 \vee x_2) \\ C_3 &= (x_3 \vee x_5 \vee x_6) \\ &\vdots \\ C_n &= (x_9 \vee \neg x_1 \vee x_5 7) \end{aligned}$$

The goal of the MAX k-SAT problem is to find an assignment to the variables x that maximizes the number of constraints that are satisfied (i.e. evaluate to true when the assignment is plugged in). To put this in terms of the framework above, an instance Φ gives rise to an objective function f_Φ given by

$$f_\Phi(x) = \sum_i C_i(x)$$

where $C_i(x)$ is equal to one if x satisfies C_i and is zero otherwise. The domain of possible solutions is $S = \{0,1\}^n$.

Example 1.1.2 (MATCHING)**.** An instance of the MATCHING problem is given by a graph $G = (V, E)$. The goal is to find a collection of vertex-disjoint edges $M \subseteq E$ (called a matching) which maximizes the number of edges in M. When this maximum is $\frac{|V|}{2}$ then we say that M has a perfect matching, since the matching achieving this maximum is incident on every vertex.

For an instance G of MATCHING the corresponding objective function is f_G is very simple

$$f_G(M) = |M|.$$

The solution domain, however, is much more complicated. It consists of all subsets $M \subseteq E$ where every pair of edges $e_1, e_2 \in M$ do not share vertices. This complexity in the domain presents an important challenge for constructing linear or semidefinite programming relaxations.

Linear and Semidefinite Programming Relaxations

As mentioned before, a standard technique for solving combinatorial optimization problems is to construct a convex relaxation. We will focus on two standard and very powerful types convex relaxations: linear programs and semidefinite programs.

A linear programming relaxation for a combinatorial optimization problem is a way of encoding the objective function f as a linear function w^f defined on a polytope $\mathcal{P}$ (i.e. a convex set defined by linear inequalities). In order for this encoding to be a relaxation, we require that for every possible solution s there is a point y_s in the polytope such that $w^f(y_s) = f(s)$. One can think of this as an embedding of the possible solutions s into the polytope in a way that respects the linear objective function. In particular this implies that

$$\operatorname{opt}(f) = \max_{s \in S} f(s) = \max_{s \in S} w^f(y_s) \leq \max_{y \in \mathcal{P}} w^f(y).$$

That is, the optimal value of the linear programming relaxation is only higher than that of the original problem. A useful linear programming relaxation is one that always has optimal value close to (or ideally equal to) that of the original problem. We now give an example.

Example 1.1.3 (Linear Programming Relaxation for MAX 3-SAT)**.** Let Φ be a MAX 3-SAT instance. For every variable x_i in Φ we define a variable $y_i \in \mathbb{R}$ which is intended to indicate the assignment to x_i. For every constraint C_j in Φ we define variables $c_{j,a} \in \mathbb{R}$ for $a \in \{0,1\}^3$ which are supposed to indicate which of the possible three-bit strings is assigned to the variables of C_j. The polytope $\mathcal{P}$ is given by the linear constraints $0 \leq y_i \leq 1$, $0 \leq c_{j,a} \leq 1$, $\sum_a c_{j,a} = 1$ and consistency constraints of the form

$$\sum_{a_2,a_3} c_{j,(1,a_2,a_3)} = y_i$$

whenever x_i is the 1st variable in the constraint C_j (similar constraints are added if x_i is negated or included as the 2nd or 3rd variable).

Note that if we assign values from $\{0,1\}$ to each y_i we can then, for each j set $c_{j,a}$ to one for exactly one value of a, and zero for all the other values in a way that satisfies all the constraints. In particular, we just set $c_{j,a} = 1$ for the value of a corresponding to the assignment to the variables in C_j indicated by y_i. The linear objective function is given by

$$w(c,y) = \sum_{j,a:C_j(a)=1} c_{j,a}$$

which is intended to count the number of constraint variables that correspond to a satisfying assignment. Indeed, if for each j we set $c_{j,a}$ as described above, this function will count the number of constraints satisfied by the assignment corresponding to the $\{0,1\}$ values assigned to each y_i.

Thus, for every $x \in \{0,1\}^n$ there is a setting of the variables c and y such that $w(c,y) = f_\Phi(x)$ and all the linear constraints are satisfied. That is, there is a point (c_x, y_x) in the polytope $\mathcal{P}$ such that $w(c_x, y_x) = f_\Phi(x)$.

A semidefinite programming relaxation is exactly the same as a linear programming relaxation except that instead of the domain being a polytope $\mathcal{P}$ it is instead a spectrahedron $\mathcal{S}$ i.e. a subset of the positive semidefinite matrices (denoted $\mathbb{S}^r_+$), where the subset is defined by linear inequalities in the matrix entries. The classic example of a semidefinite programming relaxation is the Goemans-Williamson SDP, defined first in [GW95], which introduced what is still the best known approximation algorithm for MAX CUT. Indeed by Raghavendra's results in [Rag08], the algorithm based on this SDP achieves the best possible approximation ratio for MAX CUT assuming the Unique Games Conjecture.

1.2 Extended Formulation Lower Bounds

Lower bounds on the size of linear and semidefinite programming relaxations are often called, for historical reasons, extended formulation lower bounds. The first results of this form were proved by Yannakakis in [Yan91, Yan88]. He was interested in taking a given very complicated polytope $\mathcal{P}$, for example the convex hull of all TSP tours in a graph, and finding a polytope $\mathcal{Q}$, with fewer faces but living in a larger dimension, such that there exists a linear map π with $\pi(\mathcal{Q}) = \mathcal{P}$. This more complicated polytope $\mathcal{Q}$ is called an extended formulation for $\mathcal{P}$, and corresponds to a linear program with more variables but hopefully many fewer constraints than the original number required to define $\mathcal{P}$. Yannakakis was able to prove lower bounds for extended formulations of the TSP and Matching polytopes, under the assumption that these extended formulations were "symmetric" in an appropriate sense.

This requirement that the extended formulation be symmetric is not without loss of generality. In fact a series of papers [KPT10], [Goe09] and [Pas14] gave concrete examples of polytopes for which there exist asymmetric extended formulations that are provably smaller than any symmetric one. More recently, in [FKPT12, FMP$^+$12], Fiorini et. al. showed how to remove the assumption that the extended formulation be symmetric for the TSP polytope. Rothvoß extended these results further to the Matching problem in [Rot14]. Various papers were able to generalize this type of result to approximate constraint satisfaction [BFPS12, CLRS13], and further to semidefinite programs [LRS15].

Much of the more recent work, especially for approximate constraint satisfaction, slightly modifies the original framework of Yannakakis. Rather than attempt to prove that a fixed polytope has no small extended formulation, these results instead introduce a model of linear and semidefinite programming relaxations that coincides with the informal description given in the last section. They then show that any linear or semidefinite programming relaxation satisfying the requirements of this model cannot have small size for various optimization problems. The results in this thesis are also phrased in terms of this model of linear and semidefinite programming relaxations.

1.3 Results

In this thesis we will prove lower bounds on the size of extended formulations in two broad cases. First, we will prove linear programming lower bounds for refuting random constraint satisfaction problems. This result gives a simple distribution on instances which is hard to refute for any small linear program. Further, we identify a threshold for the density (i.e.

number of constraints per variable) below which no small linear program can refute the instance, and above which there exist polynomial time refutation algorithms. Second, we will prove lower bounds for symmetric semidefinite programs solving the matching problem. Because there is a polynomial time algorithm for the matching problem, this result is best understood as identifying a fundamental problem where symmetric SDPs provably do not provide the best known algorithms.

1.4 Organization

In Chapter 2 we introduce basic notation and the general setup for proving lower bounds for LP and SDP relaxations. In Chapter 3 we prove lower bounds on the size of linear programming relaxations for refuting random constraint satisfaction problems. In Chapter 4 we prove lower bounds on the size of any symmetric semidefinite programming relaxation for the matching problem.

Chapter 2

Preliminaries

2.1 Notation

Sets. We will use $[n]$ to denote the set $\{1, \ldots, n\}$. We will use S^c to denote the complement of the set S.

Probability. We will use $\mathbb{P}_{\mathcal{D}}[X]$, $\mathbb{E}_{\mathcal{D}}[X]$ and $\mathbb{V}_{\mathcal{D}}[X]$ to denote respectively the probability, variance and expectation of X over the distribution $\mathcal{D}$.

Linear Algebra. We will let $\mathbb{S}_+^r$ denote the cone of $r \times r$ real symmetric positive semidefinite (psd) matrices. We use $\mathbb{1}$ to denote the vector of all ones.

Polynomials. We write $\mathbb{R}[x]$ denote the set of polynomials in n real variables $x = (x_1, \ldots, x_n)$ with real coefficients. For a set $\mathcal{H} \subseteq \mathbb{R}[x]$ let $\langle \mathcal{H} \rangle$ denote the vector space spanned by $\mathcal{H}$ and let $\langle \mathcal{H} \rangle_I$ denote the ideal generated by $\mathcal{H}$.

Fourier Analysis. For a set Ω and distribution $\mathcal{D}$ on Ω we write $L^2(\Omega, \mathcal{D})$ to denote the space of functions f on Ω with inner product given by $\langle f, g \rangle = \mathbb{E}_{x \sim \mathcal{D}}[f(x)g(x)]$. When $\mathcal{D}$ is the uniform distribution on Ω we will denote this space simply by $L^2(\Omega)$. We will use π_p to denote the distribution on $\{-1, 1\}$ which assigns probability p to -1 and probability $q \overset{\text{def}}{=} 1 - p$ to 1. We will use $\mathbb{I}(x = a)$ to denote the function of x that is equal to one if $x = a$ and zero otherwise. For $x \in \{-1, 1\}^n$ and a subset S we let $\chi_S(x) = \prod_{i \in S} x_i$ be the Fourier character corresponding to S.

Group Actions. If a group G acts on a set X, we will denote the (left) action of $g \in G$ on $x \in X$ by $g \cdot x$.

2.2 Relaxations of Optimization Problems

We will be primarily interested in optimization problems over discrete domains. That is, we will focus on problems where the goal is to find the optimum value of some function on some finite set. We begin by introducing our model for maximization problems (minimization can be handled in a completely analogous way).

Definition 2.2.1. A *maximization problem* $\mathcal{P} = (\mathcal{S}, \mathcal{F})$ consists of:

- A finite set $\mathcal{S}$ of feasible solutions
- A finite set $\mathcal{F}$ of nonnegative objective functions.

We define $\mathrm{opt}_{\mathcal{P}}(f) \overset{\mathrm{def}}{=} \max_{s \in \mathcal{S}} f(s)$ to be the maximum value of the objective function f on the set $\mathcal{S}$.

We will also need to formalize a notion of approximation for algorithms solving maximization problems.

Definition 2.2.2. Given two functions $\tilde{C}, \tilde{S} \colon \mathcal{F} \to \mathbb{R}$ specifying approximation guarantees, an algorithm $(\tilde{C}, \tilde{S})$-approximately solves $\mathcal{P}$ if for all $f \in \mathcal{F}$ with $\max_{s \in \mathcal{S}} f(s) \leq \tilde{S}(f)$ it computes $\tilde{f} \in \mathbb{R}$ satisfying $\max_{s \in \mathcal{S}} f(s) \leq \tilde{f} \leq \tilde{C}(f)$.

2.3 Linear Programming Formulations

In this section we define a framework for linear programming formulations and show that a small linear programming formulation implies the existence of a small non-negative representation for a given problem. We begin with the definition of a linear programming formulation of a maximization problem obtained by *linearizing* the objective functions and possible solutions.

Definition 2.3.1 (LP formulation for $\mathcal{P}$)**.** Let $\mathcal{P} = (\mathcal{S}, \mathcal{F})$ be a maximization problem. A *linear programming formulation of* $\mathcal{P}$ of size R consists of a linear map $A \colon \mathbb{R}^k \to \mathbb{R}^R$ and $b \in \mathbb{R}^R$ together with

1. *Feasible solutions:* a $y^s \in \mathbb{R}^k$ with $Ay^s \leq b$ for all $s \in \mathcal{S}$, i.e., the polytope $\left\{ y \in \mathbb{R}^k \,\middle|\, Ay \leq b \right\}$ contains the points $\{y^s \mid s \in \mathcal{S}\}$,
2. *Objective functions:* a vector $w^f \in \mathbb{R}^k$ satisfying $\langle w^f, y^s \rangle = f(s)$ for all $f \in \mathcal{F}$ and all $s \in \mathcal{S}$, i.e., the linearizations are exact on solutions.

An LP formulation for a problem $\mathcal{P}$ naturally gives rise to the following linear programming relaxation for $\mathcal{P}$:

$$\mathrm{LP}(f) = \max_{y \colon Ay \leq b} \langle w^f, y \rangle.$$

By the above definition, we have for each objective function $f \in \mathcal{F}$

$$\mathrm{opt}_{\mathcal{P}}(f) = \max_{s \in \mathcal{S}} f(s)$$

$$
\begin{aligned}
&= \max_{s \in \mathcal{S}} \langle w^f, y^s \rangle \\
&\leq \max_{y\colon Ay \leq b} \langle w^f, y^s \rangle = \mathrm{LP}(f).
\end{aligned}
$$

Thus, $\mathrm{opt}_{\mathcal{P}}(f) \leq \mathrm{LP}(f)$, i.e. $\mathrm{LP}(f)$ is a relaxation of the maximization problem $\mathcal{P}$. This implies that an upper bound on the value of an LP relaxation provides an upper bound on the optimum of the original maximization problem. Such an upper bound can be thought of as a certificate that the maximum value of f is bounded. More formally, we make the following definition:

Definition 2.3.2. Let $c \geq 0$. An LP formulation of a maximization problem $\mathcal{P}$ *certifies an upper bound c on f* if $\mathrm{LP}(f) \leq c$.

Next we show how a small LP formulation for a problem $\mathcal{P}$ gives rise to a small non-negative representation.

Lemma 2.3.3 (Non-negative representation of an LP formulation)**.** *If a maximization problem $\mathcal{P} = (\mathcal{S}, \mathcal{F})$ has a linear programming formulation of size R certifying upper bound c on $f \in \mathcal{F}$, then there exist real-valued, non-negative functions $p_i(f)$, $q_i(s) \geq 0$ for $i \in \{0 \dots R\}$ such that*

$$
c - f(s) = \sum_{i=0}^{R} p_i(f) q_i(s).
$$

Proof. An LP formulation certifying an upper bound c on f implies the existence of A,b and w^f such that

$$
\max_{y\colon Ay \leq b} \langle w^f, y \rangle \leq c.
$$

In particular $c - \langle w^f, y \rangle$ is a non-negative function of y on the feasible region $\{y \mid Ay \leq b\}$. By linear programming duality, this implies that we can write $c - \langle w^f, y \rangle$ as a non-negative combination of the R constraint functions $b_i - \langle A_i, y \rangle$, where A_i is the i-th row of A. In particular, there exist non-negative numbers $p_i(f)$ such that

$$
c - \langle w^f, y \rangle = p_0(f) + \sum_{i=1}^{R} p_i(f)(b_i - \langle A_i, y \rangle).
$$

Again by the definition of LP formulation there exist y^s such that $\langle w^f, y^s \rangle = f(s)$ for all s. Plugging this into the above equation yields

$$
c - f(s) = p_0(f) + \sum_{i=1}^{R} p_i(f)(b_i - \langle A_i, y^s \rangle).
$$

Now define $q_i(s) \overset{\mathrm{def}}{=} b_i - \langle A_i, y^s \rangle$ for $i \in \{1 \dots R\}$ and $q_0(s) \equiv 1$. Then the above equation simplifies to

$$
c - f(s) = \sum_{i=0}^{R} p_i(f) q_i(s).
$$

□

2.4 Symmetric Semi-definite Programming Formulations

In this section we define a framework for symmetric semidefinite programming formulations and show that a symmetric SDP formulation implies a symmetric sum of squares representation over a small basis. Our framework extends the one in [BPZ15] with a symmetry condition; see also [LRST14].

Let G be a group with associated actions on $\mathcal{S}$ and $\mathcal{F}$. The problem $\mathcal{P}$ is *G-symmetric* if the group action satisfies the compatibility constraint $(g \cdot f)(g \cdot s) = f(s)$. For a G-symmetric problem we require G-symmetric approximation guarantees: $\tilde{C}(g \cdot f) = \tilde{C}(f)$ and $\tilde{S}(g \cdot f) = \tilde{S}(f)$ for all $f \in \mathcal{F}$ and $g \in G$.

We now define the notion of a semidefinite programming formulation of a maximization problem.

Definition 2.4.1 (SDP formulation for $\mathcal{P}$)**.** Let $\mathcal{P} = (\mathcal{S}, \mathcal{F})$ be a maximization problem with approximation guarantees $\tilde{C}, \tilde{S}$. A $(\tilde{C}, \tilde{S})$*-approximate SDP formulation of* $\mathcal{P}$ of size d consists of a linear map $\mathcal{A} \colon \mathbb{S}^d_+ \to \mathbb{R}^k$ and $b \in \mathbb{R}^k$ together with

1. *Feasible solutions:* an $X^s \in \mathbb{S}^d_+$ with $\mathcal{A}(X^s) = b$ for all $s \in \mathcal{S}$, i.e., the SDP $\left\{X \in \mathbb{S}^d_+ \,\middle|\, \mathcal{A}(X) = b\right\}$ is a relaxation of $\operatorname{conv}\{X^s \mid s \in \mathcal{S}\}$,
2. *Objective functions:* an affine function $w^f \colon \mathbb{S}^d_+ \to \mathbb{R}$ satisfying $w^f(X^s) = f(s)$ for all $f \in \mathcal{F}$ with $\max_{s \in \mathcal{S}} f(s) \leq \tilde{S}(f)$ and all $s \in \mathcal{S}$, i.e., the linearizations are exact on solutions, and
3. *Achieving guarantee:* $\max\left\{w^f(X) \mid \mathcal{A}(X) = b, X \in \mathbb{S}^d_+\right\} \leq \tilde{C}(f)$ for all $f \in \mathcal{F}$ with $\max_{s \in \mathcal{S}} f(s) \leq \tilde{S}(f)$.

If G is a group, $\mathcal{P}$ is G-symmetric, and G acts on $\mathbb{S}^d_+$, then an SDP formulation of $\mathcal{P}$ with symmetric approximation guarantees $\tilde{C}, \tilde{S}$ is *G-symmetric* if it additionally satisfies the compatibility conditions for all $g \in G$:

1. *Action on solutions:* $X^{g \cdot s} = g \cdot X^s$ for all $s \in \mathcal{S}$.
2. *Action on functions:* $w^{g \cdot f}(g \cdot X) = w^f(X)$ for all $f \in \mathcal{F}$ with $\max_{s \in \mathcal{S}} f(s) \leq \tilde{S}(f)$.
3. *Invariant affine space:* $\mathcal{A}(g \cdot X) = \mathcal{A}(X)$.

A G-symmetric SDP formulation is *G-coordinate-symmetric* if the action of G on $\mathbb{S}^d_+$ is by permutation of coordinates: that is, there is an action of G on $[d]$ with $(g \cdot X)_{ij} = X_{g^{-1} \cdot i, g^{-1} \cdot j}$ for all $X \in \mathbb{S}^d_+$, $i, j \in [d]$ and $g \in G$.

We now turn a G-coordinate-symmetric SDP formulation into a symmetric sum of squares representation over a small set of basis functions.

Lemma 2.4.2 (Sum of squares for a symmetric SDP formulation)**.** *If a G-symmetric maximization problem $\mathcal{P} = (\mathcal{S}, \mathcal{F})$ admits a G-coordinate-symmetric $(\tilde{C}, \tilde{S})$-approximate SDP formulation of size d, then there is a set $\mathcal{H}$ of at most $\binom{d+1}{2}$ functions $h \colon \mathcal{S} \to \mathbb{R}$ such that for any $f \in \mathcal{F}$ with $\max f \leq \tilde{S}(f)$ we have $\tilde{C}(f) - f = \sum_j h_j^2 + \mu_f$ for some $h_j \in \langle \mathcal{H} \rangle$ and constant $\mu_f \geq 0$. Furthermore the set $\mathcal{H}$ is invariant under the action of G given by $(g \cdot h)(s) = h(g^{-1} \cdot s)$ for $g \in G$, $h \in H$ and $s \in S$.*

Proof. For any psd matrix M let $\sqrt{M}$ denote the unique psd matrix with $\sqrt{M}^2 = M$. Note that $\sqrt{M}\sqrt{M}^\intercal = M$ also, since $\sqrt{M}$ is symmetric.

Let $\mathcal{A}$, b, $\{X^s \mid s \in \mathcal{S}\}$, $\left\{w^f \mid f \in \mathcal{F}\right\}$ comprise a G-coordinate-symmetric SDP formulation of size d. We define the set $\mathcal{H} \coloneqq \{h_{ij} \mid i, j \in [d]\}$ via $h_{ij}(s) \coloneqq \sqrt{X^s}_{ij}$. By the action of G and the uniqueness of the square root, we have $g \cdot h_{ij} = h_{g \cdot i, g \cdot j}$, so $\mathcal{H}$ is G-symmetric. As $h_{ij} = h_{ji}$, the set $\mathcal{H}$ has at most $\binom{d+1}{2}$ elements.

By standard strong duality arguments as in [BPZ15], for every $f \in \mathcal{F}$ with $\max f \leq \tilde{S}(f)$, there is a $U^f \in \mathbb{S}^d_+$ and $\mu_f \geq 0$ such that for all $s \in \mathcal{S}$,

$$\tilde{C}(f) - f(s) = \operatorname{Tr}[U^f X^s] + \mu_f.$$

Again by standard arguments the trace can be rewritten as a sum of squares:

$$\operatorname{Tr}[U^f X^s] = \operatorname{Tr}\left[\left(\sqrt{U^f}\sqrt{X^s}\right)^\intercal \left(\sqrt{U^f}\sqrt{X^s}\right)\right] = \sum_{i,j \in [d]} \left(\sum_{k \in [d]} \sqrt{U^f}_{ik} \cdot \sqrt{X^s}_{kj}\right)^2.$$

Therefore $\tilde{C}(f) - f = \sum_{i,j \in [d]} \left(\sum_{k \in [d]} \sqrt{U^f}_{ik} \cdot h_{kj}\right)^2 + \mu_f$, as claimed. □

Chapter 3

Linear Programming Lower Bounds

3.1 Introduction

In this chapter we will prove lower bounds on the size of linear programming relaxations that refute random constraint satisfaction problems. In this setting, a random instance of a CSP is chosen from some natural distribution. Examples of such distributions are: sample uniformly at random from all instances with m constraints, or sample each possible constraint independently with some fixed probability p. In either case, as the parameter (m or p) is increased above a certain threshold, a random instance of the CSP is unsatisfiable with high probability. In this regime the natural computational question is *refutation*. That is, the goal is to efficiently find a proof that the instance is unsatisfiable.

Notice that as the parameter (m or p) increases further, eventually the refutation problem becomes easy to solve as there are likely to be a constant size set of constraints which directly contradict each other. Indeed, for a broad class of random CSPs polynomial time refutation algorithms are known above a certain threshold [AOW15]. Intriguingly, below this threshold there is a class of subexponential (but super-polynomial) time refutation algorithms that smoothly trade-off between runtime and the parameter p [RRS16].

Linear programming relaxations for CSPs naturally give refutation algorithms. An LP relaxation always provides a valid upper bound on the number of constraints satisfied by any assignment. Therefore, if the value of solution from an LP relaxation is less than the number of constraints in the instance, this is a proof that the instance is unsatisfiable. However, previous work on LP lower bounds has relied on showing that every linear programming relaxation of a given size does not perform better than some standard LP hierarchy (e.g. the Sherali-Adams hierarchy) of similar size [CLRS13, KMR17].

These reductions to known lower bounds for some fixed LP hierarchy are effective for proving worst-case LP lower bounds. However, they do not work for random instances of CSPs because they output instances which are very far from random (in other words very unlikely to be sampled from one of the natural distributions). In this chapter we prove super-polynomial lower bounds on the size of linear programs for MAX k-SAT and MAX k-XOR for random instances just below the threshold where efficient refutation algorithms are known to exist.

Theorem 3.1.1. *Any linear programming relaxation that refutes random instances of* MAX k-XOR *or* MAX k-SAT *with* $n^{\frac{k}{2}-\varepsilon}$ *constraints must have size at least* $n^{\Omega\left(\varepsilon^2 \frac{\log n}{\log\log n}\right)}$.

3.2 Preliminaries

In this section we introduce basic definitions and notation that will be used for the rest of the chapter.

Random Instances of Constraint Satisfaction Problems

In this section we formally introduce the class of constraint satisfaction problems and our model for random instances. Specifically, we will consider CSPs defined by a single boolean predicate P applied to k-tuples of literals (i.e. possibly negated variables) from a set of n variables.

Definition 3.2.1. A predicate $P\colon \{-1,1\}^k \to \{0,1\}$ defines a constraint satisfaction problem CSP(P). An instance I of CSP(P) is given by a set of n variables $x_1,\ldots,x_n$ and a set of m *constraints* (S,b), where $S=(S_1,\ldots,S_k)$ is a *scope* of k distinct variable indices and $b \in \{-1,1\}^k$ is a string of k negations. The objective is: given an instance I, find an assignment of values from $\{-1,1\}^n$ to the variables $x_1,\ldots,x_n$ in order to maximize

$$f_I(x) \overset{\text{def}}{=} \sum_{(S,b)} P(b_1 x_{S_1},\ldots,b_k x_{S_k}).$$

We will use $\mathrm{opt}_P(I)$ to denote this maximum.

Note that CSP(P) is a maximization problem where the set of possible solutions is $\mathcal{S}=\{-1,1\}^n$ and the set of objective functions $\mathcal{F}$ is the set of all f_I for each instance I of CSP(P). We will also use $E_{n,k}$ to denote the set of possible scopes i.e. the set of k-tuples of distinct indices in $[n]^k$.

We say that a predicate P supports a $(t-1)$-wise uniform distribution on satisfying assignments if there is a probability distribution η_P with the following properties:

- η_P is supported on $P^{-1}(1)$.
- For every set x_S of $(t-1)$ input variables, the distribution of x_S is uniform.

Note, for example, that the uniform distribution on satisfying assignments to MAX k-XOR is $(k-1)$-wise uniform.

One natural distribution on random instances of CSPs is given by choosing Δn constraints (S,b) independently and uniformly at random, where $\Delta \geq 0$ is called the *constraint density.* For technical reasons we will use a slightly different distribution in our analysis. However, our results for this modified distribution can be translated into similar results for the original distribution with density Δ.

Definition 3.2.2. Given a parameter $p \geq 0$ a random instance I of CSP(P) on n variables is sampled as follows:

- Choose each constraint scope S independently with probability p.
- For each S chosen, sample a uniformly random string $b_S \in \{-1, 1\}^k$.
- Let I consist of all the constraints (S, b_S) chosen by this process.

We will use $\mathcal{D}(p)$ to denote this distribution.

In short, each constraint scope is included independently with probability p and each scope is assigned one uniform random string of negations. By analogy with the situation where the number of constraints is fixed to Δn, we will set $\Delta \overset{\text{def}}{=} \frac{p|E_{n,k}|}{n}$ so that the expected number of constraints in an instance sampled from $\mathcal{D}(p)$ is $p|E_{n,k}| = \Delta n$. Standard Chernoff bounds imply that the number of constraints chosen is close to Δn with high probability.

Lemma 3.2.3. *Let m be the number of constraints in an instance sampled from $\mathcal{D}(p)$.*

$$\underset{\mathcal{D}(p)}{\mathbb{P}}\left[|m - \Delta n| > \varepsilon \Delta n\right] \leq 2 \exp\left(-\frac{\varepsilon^2 \Delta n}{3}\right).$$

Proof. Every constraint scope is included independently with probability p so m is the sum of $|E_{n,k}|$ independent Bernoulli random variables. Thus

$$\underset{\mathcal{D}(p)}{\mathbb{P}}\left[|m - \Delta n| > \varepsilon \Delta n\right] \leq 2 \exp\left(-\frac{\varepsilon^2 \Delta n}{3}\right).$$

□

Refutation

Let p_c be the threshold for $\mathrm{CSP}(P)$ above which a random instance sampled from $\mathcal{D}(p)$ is unsatisfiable with high probability. If a random instance I of a CSP is sampled with parameter $p > p_c$, then with high probability I will be unsatisfiable. Thus, a natural computational problem is to efficiently find a certificate that I is unsatisfiable. This is known as *refuting* the instance I.

Definition 3.2.4. Let A be an algorithm that takes as input an instance I of $\mathrm{CSP}(P)$ and outputs a number. We say that A is a refutation algorithm for $\mathrm{CSP}(P)$ if A always outputs a valid upper bound on $\mathrm{opt}_P(I)$. If additionally A outputs $1 - \delta$ with probability s over the random choice of I sampled from a distribution μ, then we say that A is a δ-refutation algorithm for μ with success probability s.

By Definition 2.3.1, any LP formulation for $\mathrm{CSP}(P)$ immediately gives rise to a refutation algorithm. Because random CSPs with parameter $p > p_c$ are not even $1 - \delta$ satisfiable (for some constant δ) with high probability, it is natural to ask if it is possible to construct an LP formulation that δ-refutes such CSPs with any reasonable probability. Our main result shows that such an LP formulation must have super-polynomial size.

Notation

We now introduce notation that we will use from now on. Recall that $E_{n,k}$ denotes the set of possible scopes $S \in [n]^k$. Clearly $|E_{n,k}| = \frac{n!}{(n-k)!}$ is the number of possible constraint scopes.

We will encode assignments by $x \in \{-1,1\}^n$, sets of constraint scopes by $y \in \{-1,1\}^{|E_{n,k}|}$ and sets of negations by $b \in \{-1,1\}^{|E_{n,k}| \cdot k}$. We will index coordinates of x by indices $i \in [n]$, coordinates of y by scopes $S \in E_{n,k}$ and coordinates of b by pairs (S, j) of scopes $S \in E_{n,k}$ and indices $j \in [k]$.

The coordinates of x are the assignment to the n variables in the instance. The coordinate $b_{(S,j)}$ of b is the j-th entry of the string of negations b_S for the constraint scope S. The coordinate y_S is the $\{-1,1\}$ indicator of whether the constraint on scope S is included. We will also use the notation $x_S = (x_{S_1}, \ldots, x_{S_k})$ and $b_S = (b_{(S,1)}, \ldots, b_{(S,k)})$.

Note that every instance I in the support of $\mathcal{D}(p)$ can be described by a pair (y, b) where y indicates which constraint scopes S are included in I and b indicates which negations are applied on each scope, including those not included in I. We allow these coordinates of b corresponding to scopes S not included in I to be arbitrary without affecting the instance. For technical reasons, it will be useful to extend the distribution $\mathcal{D}(p)$ to a distribution on pairs (y, b), where the coordinates b_S for scopes S not included in I by y are also sampled uniformly at random.

For a subset $U \subseteq E_{n,k}$ we will write y_U to denote the subset of coordinates of y corresponding to U. By extension we will write b_U for the subset of coordinates $b_{(S,j)}$ of the negations b with $S \in U$. Finally, we will use $I_U \overset{\text{def}}{=} (y_U, b_U)$ to denote the instance I restricted to the subset U. We will refer to I_U as a restricted instance.

We will use $\mathcal{U}_P$ to denote the $(t-1)$-wise uniform distribution supported on $z \in \{-1,1\}^k$ with $P(z) = 1$. We will use η_P to denote the density of $\mathcal{U}_P$ with respect to the uniform distribution on $\{-1,1\}^k$.

Using this notation, we can rewrite the CSP(P) objective function from Definition 3.2.1 as

$$F(x, y, b) \overset{\text{def}}{=} f_{(y,b)}(x) = \sum_{S \in E_{n,k}} \mathbb{I}(y_S = -1) P(b_S \circ x_S) \tag{3.2.1}$$

Blockwise-Dense Distributions

A key element of our proof lies in being able to approximately decompose any distribution on instances into a convex combination of “simpler” distributions. The class of simple distributions we will consider are are based on the following definition.

Definition 3.2.5. Let $\mathcal{D}$ be a distribution on instances $I = (y, b)$. We say that $\mathcal{D}$ is *blockwise-dense* relative to $\mathcal{D}(p)$ with parameter δ if for all instances I' we have

$$\mathop{\mathbb{P}}_{I \sim \mathcal{D}}[I = I'] \leq \mathop{\mathbb{P}}_{I \sim \mathcal{D}(p)}[I = I']^{1-\delta}.$$

That is, a blockwise-dense distribution $\mathcal{D}$ does not assign a much higher probability to any instance than $\mathcal{D}(p)$ does. The exponent of $1 - \delta$ should be thought of as some constant with a small value of $\delta > 0$. Unfortunately, we will not be able to decompose any distribution into a convex combination of blockwise-dense distributions. In particular, a distribution which fixes a set of d scopes and their corresponding negations cannot be decomposed into blockwise-dense distributions if d is large enough. This example motivates the following definition.

Definition 3.2.6. Let $\mathcal{D}$ be a distribution on instances I. We say that $\mathcal{D}$ is *d-conjunctive blockwise-dense* (d-CBD) with parameter δ relative to $\mathcal{D}(p)$ if

1. There exists $U \subseteq E_{n,k}$ with $|U| \leq d$ and a restricted instance I_U^* such that $\mathbb{P}_{I \sim \mathcal{D}}\left[I_U = I_U^*\right] = 1$.
2. For every subset $V \subseteq E_{n,k} \setminus U$ and every instance I'_V we have

$$\underset{I \sim \mathcal{D}}{\mathbb{P}}\left[I_V = I'_V\right] \leq \underset{I \sim \mathcal{D}(p)}{\mathbb{P}}\left[I_V = I'_V\right]^{1-\delta}$$

We call the coordinates in the subset U the *fixed* coordinates of $\mathcal{D}$.

The basic intuition is that d-CBD distributions fix some small set of d coordinates and then sample all other coordinates with probability not too much higher than the probability assigned to them by $\mathcal{D}(p)$. In Section 3.5 we will show that any distribution can be decomposed into a convex combination of conjunctive blockwise-dense distributions plus an error set which has small measure under $\mathcal{D}(p)$.

Sherali-Adams Pseudo-densities

The Sherali-Adams linear programming hierarchy is a sequence of LP relaxations of increasing size. Intuitively, the d-th linear program in this sequence constructs locally defined probability distributions on d-tuples of variables which locally satisfy the constraints. This collection of local distributions gives rise to an object called a pseudo-density, because it "acts like" a true probability density on small collections of variables. We now give the formal definition for this object.

Definition 3.2.7. A degree-d Sherali-Adams pseudo-density $H : \{-1, 1\}^n \to \mathbb{R}^+$ for an instance $I = (y, b)$ is a non-negative function satisfying:

- For every non-negative d-Junta $J(x)$ (i.e. function depending on only d input variables), $\mathbb{E}_x\left[H(x)J(x)\right] \geq 0$.
- $\mathbb{E}_x\left[H(x)\right] = 1$.
- For every constraint C, $\mathbb{E}_x\left[H(x)\,\mathbb{I}(C(x) = 1)\right] = 1$.

3.3 Proof Overview

In this section we introduce our approach to proving LP lower bounds using pseudo-calibration. The main idea is that we want to use the small non-negative representation given by Lemma 2.3.3 to derive a contradiction. In particular, we want to find a function $H(x, I)$ that witnesses a violation of the equality $c - f_I(x) = \sum_i p_i(I)q_i(x)$. That is, we want both

$$\mathbb{E}[H(x, I)(c - f_I(x))] < 0 \qquad \text{and} \qquad \mathbb{E}\left[H(x, I)\sum_{i=0}^{R} p_i(I)q_i(x)\right] \geq 0$$

where the expectation is taken over random CSPs conditioned on being in the set of instances I on which the LP succeeds.

One natural way to try to construct such an H is to first choose a *planted* distribution $\mathcal{D}_*$ on pairs (x, I) of assignments and instances, so that x always completely satisfies the instance I. Next, let $\mu_*(x, I)$ be the density of $\mathcal{D}_*$ relative to the distribution where $I \sim \mathcal{D}(p)$ and $x \sim \{-1, 1\}^n$ independently. Since x always satisfies I we have

$$\mathop{\mathbb{E}}_{\substack{I\sim\mathcal{D}(p)\\ x\sim\{-1,1\}^n}}[\mu_*(x, I)(c - f_I(x))] = \mathop{\mathbb{E}}_{(x,I)\sim\mathcal{D}_*}[c - f_I(x)] < 0$$

for any c which is sufficiently less than the expected number of constraints in I. Furthermore, since each of the $q_i(x)$ and $p_i(I)$ are non-negative we have by linearity of expectation

$$\mathop{\mathbb{E}}_{\substack{I\sim\mathcal{D}(p)\\ x\sim\{-1,1\}^n}}\left[\mu_*(x, I)\sum_{i=0}^{R} q_i(x)p_i(I)\right] = \sum_{i=0}^{R}\mathop{\mathbb{E}}_{(x,I)\sim\mathcal{D}_*}[q_i(x)p_i(I)] \geq 0.$$

Of course this cannot be a correct proof of a lower bound, because we have not yet used that the LP formulation has size R at all. In particular, an LP for refuting a random CSP need only be correct on some s fraction of instances. This implies that the small non-negative representation is only guaranteed to exist for an s fraction of the I. Thus, μ_* may only be supported on those I in the $1 - s$ fraction of instances where no non-negative representation is guaranteed to exist, making it totally useless for proving lower bounds.

The solution to this problem is a technique called pseudo-calibration, which allows us to construct a function $H(x, I)$ with similar properties to $\mu_*(x, I)$, but which is much more spread out over the space of possible instances. Briefly, H is constructed by simply dropping the high-degree terms from the Fourier expansion of μ_* over an appropriately chosen Fourier basis.

3.4 Pseudo-Calibration

In this section we formally define the pseudo-calibration of the planted density and prove bounds on its Fourier coefficients.

The Planted Density

The first step to formally define pseudo-calibration is to choose a planted distribution $\mathcal{D}_*$ as described in Section 3.3.

Definition 3.4.1. An assignment and instance pair (x, I) is sampled from the planted distribution $\mathcal{D}_*$ as follows:

- Sample x uniformly at random from $\{-1,1\}^n$.
- Choose to include each constraint scope S independently with probability p.
- For each scope S chosen, sample $z_S \sim \mathcal{U}_P$ and set $b_S = z_S \circ x_S$.

As with the distribution $\mathcal{D}(p)$, we can extend $\mathcal{D}_*$ to a distribution on triples (x, y, b) by sampling the coordinates b_S for scopes S not included in y uniformly at random, as these do not affect the satisfiability of the instance. Let $\mu_*(x, y, b)$ be the density of $\mathcal{D}_*$ relative to the distribution $\{-1,1\}^n \times \mathcal{D}(p)$.

Now we claim that

$$\mu_*(x,y,b) \propto \prod_{S \in E_{n,k}} \mathbb{I}(y_S = -1)\eta_P(b_S \circ x_S) + \mathbb{I}(y_S = 1). \tag{3.4.1}$$

Indeed for any fixed x, y, b we have

$$\begin{aligned}
\mathop{\mathbb{P}}_{\substack{(y',b')\sim\mathcal{D}(p)\\ x'\sim\{-1,1\}^n}}[(x',y',b') = (x,y,b)]\mu_*(x,y,b) &= \left(2^{-(n+k|E_{n,k}|)} \prod_{S:y_S=-1} p \prod_{S:y_S=1} (1-p)\right) \cdot \mu_*(x,y,b) \\
&\propto 2^{-n} \prod_{S:y_S=-1} 2^{-k}\eta_P(b_S \circ x_S)p \prod_{S:y_S=1} 2^{-k}(1-p) \\
&= 2^{-n} \prod_{S:y_S=-1} \mathop{\mathbb{P}}_{z\sim\mathcal{U}_P}[z = b_S \circ x_S]p \prod_{S:y_S=1} 2^{-k}(1-p) \\
&= \mathop{\mathbb{P}}_{(x',y',b')\sim\mathcal{D}_*}[(x',y',b') = (x,y,b)]
\end{aligned}$$

Next we describe the appropriate Fourier basis for expanding μ_*. Let χ_α be the Fourier basis for $L^2(\{-1,1\}^n)$. That is, for $\alpha \subseteq [n]$

$$\chi_\alpha(x) = \prod_{i\in\alpha} x_i.$$

Let ϕ_β be the p-biased Fourier basis for $L^2(\{-1,1\}^{|E_{n,k}|}, \pi_p^{\otimes|E_{n,k}|})$. That is, for $\beta \subseteq E_{n,k}$

$$\phi_\beta(y) = \prod_{S\in\beta} \phi(y_S), \qquad \phi(-1) = -\sqrt{\frac{q}{p}}, \qquad \phi(1) = \sqrt{\frac{p}{q}}$$

where $q = 1 - p$. Finally, let ψ_γ be the Fourier basis for $L^2(\{-1,1\}^{|E_{n,k}|\cdot k})$. That is, for $\gamma \subseteq E_{n,k} \times [k]$

$$\psi_\gamma(b) = \prod_{(S,j)\in\gamma} b_{(S,j)}.$$

We can now write $\mu_*(x, y, b)$ as a function in the tensor product of the spaces spanned by the χ_α, ϕ_β and ψ_γ. That is

$$\mu_*(x, y, b) = \sum_{\alpha,\beta,\gamma} \widehat{\mu_*}(\alpha, \beta, \gamma)\chi_\alpha(x)\phi_\beta(y)\psi_\gamma(b)$$

where the Fourier coefficients above are given by the inversion formula

$$\widehat{\mu_*}(\alpha, \beta, \gamma) = \mathop{\mathbb{E}}_{(x,y,b)\sim\mathcal{D}(p)} [\mu_*(x, y, b)\chi_\alpha(x)\phi_\beta(y)\psi_\gamma(b)] = \mathop{\mathbb{E}}_{(x,y,b)\sim\mathcal{D}_*} [\chi_\alpha(x)\phi_\beta(y)\psi_\gamma(b)] .$$

Finally, we are ready to define the pseudo-calibrated density. For a pair $d = (d_x, d_y) \in \mathbb{N}^2$ let $I(d) \stackrel{\text{def}}{=} \{(\alpha, \beta, \gamma) \mid |\alpha| \leq d_x, |\beta| \leq |\bar{\gamma}| \leq d_y\}$. That is, $I(d)$ corresponds to the set of Fourier coefficients where x has degree d_x, each scope is included only if some of its negations are included, and the number of scopes whose negations are included is at most d_y. We will use L_d to denote the linear projection operator onto the span of $\{\chi_\alpha\phi_\beta\psi_\gamma\}_{(\alpha,\beta,\gamma)\in I(d)}$.

Definition 3.4.2. For $d = (d_x, d_y) \in \mathbb{N}^2$ we define the d-pseudo-calibration $\bar{\mu}_* \stackrel{\text{def}}{=} L_d\mu_*$. Equivalently

$$\bar{\mu}_*(x, y, b) = \sum_{(\alpha,\beta,\gamma)\in I(d)} \widehat{\mu_*}(\alpha, \beta, \gamma)\chi_\alpha(x)\phi_\beta(y)\psi_\gamma(b)$$

The Fourier Coefficients of the Planted Density

In this section we will compute bounds on the Fourier coefficients of the planted density $\mu_*(x, y, b)$. A key fact that we will exploit is that μ_* does not directly depend on x, but rather only on tuples of the form $b_S \circ x_S = (b_{(S,1)}x_{S_1}, \ldots, b_{(S,k)}x_{S_k})$. In order to formalize this statement we will need the following definition:

Definition 3.4.3. Let $\gamma \subseteq E_{n,k} \times [k]$ and let $\alpha \subset [n]$. Let $c_i = |\{(S, j) \in \gamma \mid S_j = i\}|$ be the number of appearances of the coordinate i as S_j for some $(S, j) \in \gamma$. Then $\gamma \vdash \alpha$ (in words, γ derives α) if c_i is odd for every $i \in \alpha$, and c_i is even for every $i \notin \alpha$.

Recalling that for a pair $(S, j) \in \gamma$ the value S_j is simply an index in $[n]$, the above definition says that every index $i \in \alpha$ must appear an odd number of times as some S_j, and every $i \notin \alpha$ must appear an even number of times. To see why this definition is useful consider the following example. Suppose $S_1 = 1$, $S_2 = 2$, $T_1 = 2$ and $T_2 = 3$. Then

$$\begin{aligned} b_{(S,1)}x_{S_1}b_{(S,2)}x_{S_2} \cdot b_{(T,1)}x_{T_1}b_{(T,2)}x_{T_2} &= b_{(S,1)}x_1b_{(S,2)}x_2 \cdot b_{(T,1)}x_2b_{(T,2)}x_3 \\ &= b_{(S,1)}b_{(S,2)}b_{(T,1)}b_{(T,2)}x_1x_3 \\ &\stackrel{\text{def}}{=} \psi_\gamma(b)\chi_\alpha(x) \end{aligned}$$

where the second to last equality used the fact that $x_2^2 = 1$. Thus, when multiplying these two monomials, corresponding to scope S and T respectively, the resulting product has the form $\psi_\gamma(b)\chi_\alpha(x)$ where $\gamma \vdash \alpha$. This is due to the fact that x_2 appeared an even number of times and so was eliminated from the product.

We now proceed with the computation of the Fourier coefficient bounds for μ_*.

Lemma 3.4.4. *If $\widehat{\mu_*}(\alpha, \beta, \gamma) \neq 0$ then $\gamma \vdash \alpha$.*

Proof. Let η_P be the density of the $(t-1)$-wise independent distribution supported on $P^{-1}(1)$, and let $\eta_P(z) = \sum_T \widehat{\eta_P}(T) \prod_{j\in T} z_j$ be its Fourier expansion. Recall from (3.4.1)

$$\mu_*(x, y, b) \propto \prod_{S\in E_{n,k}} \mathbb{I}(y_S = -1)\eta_P(b_S \circ x_S) + \mathbb{I}(y_S = 1).$$

The only terms in the above product depending on either b or x are of the form

$$\eta_P(b_S \circ x_S) = \sum_{T\subseteq k} \widehat{\eta_P}(T) \prod_{j\in T} b_{(S,j)} x_{S_j}.$$

We can use this formula to expand the expression for μ_* as a sum of products of the variables b and x and the indicators $\mathbb{I}(y_S = \pm 1)$. Any term depending on b or x will then be a product, over *distinct* S, of terms of the form

$$\prod_S \mathbb{I}(y_S = \pm 1)\widehat{\eta_P}(T_S) \prod_{j\in T_S} b_{(S,j)} x_{S_j}$$

where the $T_S \subseteq [k]$ are a sequence of subsets depending on S. In any such product over distinct S all the variables $b_{(S,j)}$ for $j \in T_S$ are distinct. So letting $\gamma = \cup_S\{(S, j) \mid j \in T_S\}$, we have

$$\prod_S \mathbb{I}(y_S = \pm 1)\widehat{\eta_P}(T_S) \prod_{j\in T_S} b_{(S,j)} x_{S_j} = \psi_\gamma(b) \prod_S \mathbb{I}(y_S = \pm 1)\widehat{\eta_P}(T_S) \prod_{j\in T_S} x_{S_j}.$$

Furthermore, every x_i appears exactly $c_i = |\{(S, j) \in \gamma \mid S_j = i\}|$ times in the product. If c_i is even $x_i^{c_i} = 1$, and if c_i is odd $x_i^{c_i} = x_i$. Thus,

$$\psi_\gamma(b) \prod_S \mathbb{I}(y_S = \pm 1)\widehat{\eta_P}(T_S) \prod_{j\in T_S} x_{S_j} = \psi_\gamma(b)\chi_\alpha(x) \prod_S \widehat{\eta_P}(T_S)\, \mathbb{I}(y_S = \pm 1)$$

where $\gamma \vdash \alpha$. Now if we expand the indicator functions $\mathbb{I}(y_S = \pm 1)$ in the ϕ_β basis, we conclude that $\mu_*(x, y, b)$ can be written as a linear combination of terms $\chi_\alpha(x)\phi_\beta(y)\psi_\gamma(b)$ where all such terms appearing with a non-zero coefficient satisfy $\gamma \vdash \alpha$. There may be additional cancellations between terms in the final Fourier expansion of μ_* as some of the basis functions with non-zero coefficients could appear multiple times. However, we can still conclude that any non-zero coefficients $\widehat{\mu_*}(\alpha, \beta, \gamma)$ must have $\gamma \vdash \alpha$.

□

Two additional definitions are required in order to state the formula for the Fourier coefficients. First, we need notation for the set of scopes contained in $\gamma \subseteq E_{n,k} \times [k]$.

Definition 3.4.5. Let $\gamma \subseteq E_{n,k} \times [k]$. Then $\bar{\gamma} \overset{\text{def}}{=} \{S \mid (S, j) \in \gamma \text{ for some } j\}$ is the set of scopes S present in γ.

Second, we need notation for the minimum number of coordinates contained in any scope S present in γ.

Definition 3.4.6. Let $\gamma \subseteq E_{n,k} \times [k]$. Then $r(\gamma) \overset{\text{def}}{=} \min_{S \in \bar{\gamma}} |\{j \mid (S, j) \in \gamma\}|$ is the *minimum arity* of γ.

Now using Lemma 3.4.4 we can compute bounds on the Fourier coefficients of μ_*.

Lemma 3.4.7. *If $\gamma \vdash \alpha$, $r(\gamma) \geq t$ and $\beta \subseteq \bar{\gamma}$ then*

$$|\widehat{\mu_*}(\alpha, \beta, \gamma)| \leq \sqrt{pq}^{|\bar{\gamma} \cap \beta|} p^{|\bar{\gamma} \setminus \beta|}$$

otherwise $\widehat{\mu_*}(\alpha, \beta, \gamma) = 0$.

Proof. The Fourier coefficients are given by

$$\widehat{\mu_*}(\alpha, \beta, \gamma) = \mathop{\mathbb{E}}_{(x,y,b) \sim \mathcal{D}(p)} \left[\mu_*(x, y, b) \chi_\alpha(x) \phi_\beta(y) \psi_\gamma(b)\right] = \mathop{\mathbb{E}}_{(x,y,b) \sim \mathcal{D}_*} \left[\chi_\alpha(x) \phi_\beta(y) \psi_\gamma(b)\right].$$

By Lemma 3.4.4 any non-zero Fourier coefficients must have $\gamma \vdash \alpha$. In this case, letting $c_i = |\{(S, j) \in \gamma \mid S_j = i\}|$ as in Definition 3.4.3, we have

$$\begin{aligned}
\chi_\alpha(x) \phi_\beta(y) \psi_\gamma(b) &= \phi_\beta(y) \prod_{(S,j) \in \gamma} b_{(S,j)} \prod_{i \in \alpha} x_i \\
&= \phi_\beta(y) \prod_{(S,j) \in \gamma} b_{(S,j)} \prod_{i \in [n]} x_i^{c_i} \\
&= \phi_\beta(y) \prod_{(S,j) \in \gamma} b_{(S,j)} x_{S_j}
\end{aligned}$$

where the second equality used the fact that $\gamma \vdash \alpha$ implies that $x_i^{c_i} = x_i$ for $i \in \alpha$ and $x_i^{c_i} = 1$ for $i \notin \alpha$. Expanding ϕ_β as a product over scopes $S \in E_{n,k}$ and grouping terms yields

$$\chi_\alpha(x) \phi_\beta(y) \psi_\gamma(b) = \left(\prod_{S \in \beta \setminus \bar{\gamma}} \phi(y_S) \right) \left(\prod_{S \in \beta \cap \bar{\gamma}} \phi(y_S) \prod_{j : (S,j) \in \gamma} b_{(S,j)} x_{S_j} \right) \left(\prod_{S \in \bar{\gamma} \setminus \beta} \prod_{j : (S,j) \in \gamma} b_{(S,j)} x_{S_j} \right).$$

Now we claim that, for each $S \in E_{n,k}$, the term corresponding to S in the above product is independent of all the other terms under the distribution $\mathcal{D}_*$. This is clearly true for the $S \in \beta \setminus \bar{\gamma}$ terms, since under $\mathcal{D}_*$ the variables y_S depend only on the variables x_S and negations b_S on scope S, but none of those variables appear in the above product. For the remaining terms, note that the distribution of $b_S \circ x_S$ is independent of all other variables and constraints except y_S. This is because under $\mathcal{D}_*$ both x and y are sampled independently, and then each b_S is sampled independently so that if $y_S = -1$ the distribution of $b_S \circ x_S$ is equal to the $(t-1)$-wise uniform distribution supported on satisfying assignments to P, and if $y_S = 1$ then $b_S \circ x_S$ is uniform and independent of everything else. Therefore, taking the expectation of $\chi_\alpha(x)\phi_\beta(y)\psi_\gamma(b)$ over $\mathcal{D}_*$ yields the product of expectations

$$\begin{aligned}&\mathop{\mathbb{E}}_{(x,y,b)\sim\mathcal{D}_*}\left[\chi_\alpha(x)\phi_\beta(y)\psi_\gamma(b)\right]\\ &= \left(\prod_{S\in\beta\setminus\bar{\gamma}} \mathbb{E}\left[\phi(y_S)\right]\right)\left(\prod_{S\in\beta\cap\bar{\gamma}} \mathbb{E}\left[\phi(y_S)\prod_{j:(S,j)\in\gamma} b_{(S,j)}x_{S_j}\right]\right)\left(\prod_{S\in\bar{\gamma}\setminus\beta}\mathbb{E}\left[\prod_{j:(S,j)\in\gamma} b_{(S,j)}x_{S_j}\right]\right).\end{aligned}$$

First, note that $\mathbb{E}_{\mathcal{D}_*}\left[\phi(y_S)\right] = \mathbb{E}_{\mathcal{D}(p)}\left[\phi(y_S)\right] = 0$, and so the above product is zero whenever $\beta \setminus \bar{\gamma} \neq \emptyset$. Thus all non-zero Fourier coefficients $\widehat{\mu_*}(\alpha, \beta, \gamma)$ must have $\beta \subseteq \bar{\gamma}$.

Second, suppose there exists some $S^* \in \bar{\gamma}$ such that $|\{j \mid (S^*, j) \in \gamma\}| < t$. Then the distribution of the coordinates $\{b_{(S^*,j)}x_{S^*_j}\}_{(S^*,j)\in\gamma}$ is uniform and independent of everything else. This is because $b_S \circ x_S$ is sampled independently from a $(t-1)$-wise uniform distribution if $y_S = -1$ and from the uniform distribution if $y_S = 1$. Thus, in both cases the aforementioned coordinates are uniform and independent of all other variables. This implies that

$$\mathbb{E}\left[\prod_{j:(S^*,j)\in\gamma} b_{(S^*,j)}x_{S^*_j}\right] = 0 = \mathbb{E}\left[\phi(y_{S^*})\prod_{j:(S^*,j)\in\gamma} b_{(S^*,j)}x_{S^*_j}\right].$$

So we conclude that whenever such an S^* exists, the Fourier coefficient is zero. Taking the contrapositive, we have that all nonzero Fourier coefficients must have $r(\gamma) \geq t$.

Next, for $S \in \beta \cap \bar{\gamma}$

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{\mathcal{D}_*}\left[\phi(y_S)\prod_{j:(S,j)\in\gamma} b_{(S,j)}x_{S_j}\right]\right| &= \left|p\cdot\phi(-1)\,\mathbb{E}\left[\prod_{j:(S,j)\in\gamma} b_{(S,j)}x_{S_j} \mid y_S = -1\right] + q\cdot\phi(1)\cdot 0\right|\\ &\leq |p\cdot\phi(-1)| = \sqrt{pq}\end{aligned}$$

where in the first equality we have used the fact that $b_S \circ x_S$ is uniformly random conditioned on $y_S = 1$, and the inequality follows from the fact that the variables b and x are bounded by one. Finally, for $S \in \bar{\gamma} \setminus \beta$, we have by the same argument

$$\left|\mathop{\mathbb{E}}_{\mathcal{D}_*}\left[\prod_{j:(S,j)\in\gamma} b_{(S,j)}x_{S_j}\right]\right| = \left|p\cdot\mathbb{E}\left[\prod_{j:(S,j)\in\gamma} b_{(S,j)}x_{S_j} \mid y_S = -1\right] + q\cdot 0\right| \leq p.$$

Plugging these bounds into the original product of expectations completes the proof. □

The Planted Density for max k-xor and max k-sat

At this point it is instructive to see what the planted density looks like for a concrete example. For the MAX k-XOR problem the predicate is simply the sum of the input bits modulo two. Thus, $P(x) = \frac{1}{2} - \frac{1}{2}\prod_i x_i$. The uniform distribution on all satisfying assignments to P is $k-1$-wise independent. Also note that the planted density for MAX k-SAT is exactly the same as that for MAX k-XOR. This is because the $k-1$-wise independent distribution supported on satisfying assignments to the MAX k-SAT predicate can be taken to be the uniform distribution on assignments satisfying the MAX k-XOR predicate on the same bits. Combining these facts, and using the above analysis we can precisely compute the Fourier expansion of $\bar{\mu}_*$ for MAX k-XOR (and MAX k-SAT).

$$\mu_*(x,y,b) = \sum_{\alpha} \chi_\alpha(x) \sum_{\substack{\gamma \vdash \alpha \\ r(\gamma) \geq k}} \chi_\gamma(b) \sum_{\beta \subseteq \bar{\gamma}} (-\sqrt{pq})^{\beta \cap \bar{\gamma}} p^{\bar{\gamma} \setminus \beta} \phi_\beta(S)$$

Since each scope has size k, the fact that $r(\gamma) \geq k$ implies that for each scope S, either the full product of the k bits b_S appears in χ_γ or none of them do. Thus, all that matters in the above Fourier expansion is if the parity of b_S is odd or even. This makes sense, because for the MAX k-XOR predicate all that is relevant about the negations is their parity. Let us use c_S to denote this parity and let $c_\gamma \overset{\text{def}}{=} \prod_{S \in \bar{\gamma}} c_S$. Now, observing that $\mathbb{I}(y_S = -1) = p - \sqrt{pq}\phi(y_S)$ we can rewrite the above expression as

$$\mu_*(x,y,b) = \sum_{\alpha} \chi_\alpha(x) \sum_{\substack{\gamma \vdash \alpha \\ r(\gamma) = k}} \mathbb{I}(y_{\bar{\gamma}} = -\mathbb{1}) c_\gamma.$$

Now imagine fixing y and b to some value i.e. fixing some instance of MAX k-XOR. In this case, the above expansion of μ_* simplifies to a function of x. For each parity $\chi_\alpha(x)$ we have a sum over derivations of α by the constraints included in the instance y. The coefficients of the sum are ± 1 depending on whether the negations on a given constraint have odd or even parity. Here the derivations of α by some set of constraints correspond exactly to linear combinations over $\mathbb{F}_2$ of the set of constraints when thought of as vectors in $\mathbb{F}_2$.

Further the pseudo-calibrated density $\bar{\mu}_*$ is obtained by restricting $|\alpha| \leq d_x$ and $|\bar{\gamma}| \leq d_y$. This corresponds above to only considering length d_y derivations of parities of size d_x. The reader familiar with the Grigoriev/Schoenebeck gap instances for the Sum-of-Squares sdp[Gri01, Sch08] might recognize the similarity with this pseudo-calibrated density. Indeed the only difference is that in the above SoS gap instances, the sum is over small-width derivations as opposed to small length derivations of a given parity. However, for the parameter regime we are interested in, small-width and small-length derivations coincide with high probability over y and b. This is follows from vertex expansion of the constraint graph, see for example [OW14].

Thus, by previous results we have the following lemma:

Lemma 3.4.8. *Let $\bar{\mu}_*(x,y,b)$ be the pseudo-calibrated density for* MAX k-XOR *and* MAX k-SAT *and for $I = (y,b)$ let $H_I(x) \overset{\text{def}}{=} \bar{\mu}_*(x,y,b)$. Then with high probability over I, we have that $H_I(x)$ is a degree-d_x Sherali-Adams pseudo-density.*

Norm Bounds for the Pseudo-calibrated Density

Now that we have computed bounds for the Fourier coefficients of the planted density μ_* we can use them to obtain bounds on the L^2 norm of the pseudo-calibrated density $\bar{\mu}_*$ when it is averaged over a conjunctive blockwise-dense distribution. The first step is to estimate the number of nonzero Fourier coefficients $\widehat{\mu_*}(\alpha, \beta, \gamma)$ for each fixed α.

Lemma 3.4.9. *Fix $\alpha \subseteq [n]$ and let $l \leq \frac{cn}{k}$ for a sufficiently small constant $c > 0$. The number of Fourier coefficients with $\widehat{\mu_*}(\alpha, \beta, \gamma) \neq 0$ and $|\bar{\gamma}| = l$ is at most*

$$C^l n^{kl - \frac{tl+|\alpha|}{2}} l^{\frac{tl+|\alpha|}{2} - l}$$

where C is a constant depending only on k and t.

Proof. By Lemma 3.4.7 all nonzero Fourier coefficients must satisfy $\gamma \vdash \alpha$, $r(\gamma) \geq t$ and $\beta \subseteq \bar{\gamma}$. For fixed α we first count the number of γ with $|\bar{\gamma}| = l$, $\gamma \vdash \alpha$, and $r(\gamma) \geq t$. To do so we think of choosing γ by first choosing numbers $s_1, \dots, s_l$ and then selecting l subsets $S_1, \dots, S_l \subseteq [n]$ with $|S_i| = s_i$. Further, we think of selecting each subset S_i by filling in s_i slots with indices from $[n]$. Since $r(\gamma) \geq t$ we have that $t \leq s_i \leq k$ for all i. Let $s = \sum_i s_i$.

Now we count the number of ways to fill in the s slots so that $\gamma \vdash \alpha$. We will slightly over-count, by allowing indices to appear more than once in each S_i and allowing pairs S_i and S_j to be identical. For $\gamma \vdash \alpha$ to hold, some subset T of the s slots must be assigned the indices from α since each such index appears an odd number of times in γ. There are $\binom{s}{|\alpha|}$ ways to select the subset T and $|\alpha|!$ ways to assign the indices from α to T.

Further, the slots outside of T must be assigned indices that come in matching pairs, since each such index must appear an even number of times. In this case the total number d of distinct indices appearing in slots outside of T is at most $\frac{s-|\alpha|}{2}$. This follows from the fact that there are $s - |\alpha|$ slots outside of T and each index must appear at least twice. There are $\binom{n}{d}$ choices for the d indices and at most $d^{s-|\alpha|}$ ways to assign these indices to the remaining slots. Putting this all together we have that the total number of ways to assign indices to the s slots is

$$\begin{aligned}
\binom{s}{|\alpha|}|\alpha|! \sum_{d=1}^{\frac{s-|\alpha|}{2}} \binom{n}{d} d^{s-|\alpha|} &\leq \binom{s}{|\alpha|}|\alpha|! \left(\frac{s-|\alpha|}{2}\right) \binom{n}{\frac{s-|\alpha|}{2}} \left(\frac{s-|\alpha|}{2}\right)^{s-|\alpha|} \\
&\leq (es)^{|\alpha|} \frac{s}{2} \left(\frac{en}{\frac{(s-|\alpha|)}{2}}\right)^{\frac{s-|\alpha|}{2}} \left(\frac{s-|\alpha|}{2}\right)^{s-|\alpha|} \\
&\leq \frac{s}{2} (en)^{\frac{s-|\alpha|}{2}} (es)^{\frac{s+|\alpha|}{2}}
\end{aligned}$$

where the second to last inequality uses the bounds $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ and $n! \leq n^n$.

Now note that if we sum the above bound over all sequences $s_1, \dots s_l$ each γ will be counted $l!$ times, once for each ordering of the subsets $S_1, \dots, S_l$ comprising γ. Thus the total number of γ where $\gamma \vdash \alpha$ and $r(\gamma) \geq t$ is at most

$$\begin{aligned}
|\{\gamma \mid \gamma \vdash \alpha, r(\gamma) \geq t\}| &\leq \frac{1}{l!} \sum_{s_1,\ldots,s_l} \frac{s}{2}(en)^{\frac{s-|\alpha|}{2}}(es)^{\frac{s+|\alpha|}{2}} \\
&= \frac{1}{l!} \sum_{s_1,\ldots,s_l} \frac{s}{2}(en)^{\frac{s-|\alpha|}{2}}(es)^{\frac{s-tl}{2}}(es)^{\frac{tl+|\alpha|}{2}}.
\end{aligned}$$

Since $r(\gamma) \geq t$ we have $s \geq tl$. Further $s \leq kl = cn$, so $(es)^{\frac{s-tl}{2}} \leq (en)^{\frac{s-tl}{2}}$. Plugging this in to the above inequality we have

$$\begin{aligned}
|\{\gamma \mid \gamma \vdash \alpha, r(\gamma) \geq t\}| &\leq \frac{1}{l!} \sum_{s_1,\ldots,s_l} \frac{s}{2}(en)^{\frac{s-|\alpha|}{2}}(en)^{\frac{s-tl}{2}}(es)^{\frac{tl+|\alpha|}{2}} \\
&\leq \frac{k^l}{l!}\frac{kl}{2}(en)^{kl-\frac{tl+|\alpha|}{2}}(ekl)^{\frac{tl+|\alpha|}{2}} \\
&\leq l^{-l}2^l(ek)^l\frac{k}{2}(en)^{kl-\frac{tl+|\alpha|}{2}}(ekl)^{\frac{tl+|\alpha|}{2}} \\
&\leq C^l n^{kl-\frac{tl+|\alpha|}{2}} l^{\frac{tl+|\alpha|}{2}-l}
\end{aligned}$$

where in the last inequality C is a constant depending only on k and t. Finally, since $\beta \subseteq \bar{\gamma}$ there are 2^l possible values of β for each γ satisfying $\gamma \vdash \alpha$ and $r(\gamma) \geq t$. Combining this with the above bound completes the proof. □

Now we are ready to compute the L^2-norm of the degree-s part of $\bar{\mu}_*$ after averaging over a CBD distribution.

Lemma 3.4.10. *Let $d_y \leq \frac{\varepsilon}{10}\frac{\log n}{\log\log n}$, let $l \leq d_y$ and set $\delta \leq \frac{\varepsilon}{2k}$. Let $\mathcal{D}$ be an l-CBD distribution with parameter δ and fixed block U. For each fixed assignment x'_U to the variables of the scopes contained in U define*

$$H_{x'_U}(x) = \mathop{\mathbb{E}}_{(y,b)\sim\mathcal{D}}[\bar{\mu}_*((x, x'_U), y, b)]$$

and set $p = \frac{\Delta n}{|E_{n,k}|}$ where $\Delta n = n^{\frac{t}{2}-\varepsilon}$. Then

$$\sum_{|\alpha|=s} \widehat{H_{x'_U}}(\alpha)^2 \leq O(n^{-\frac{\varepsilon}{k}s})$$

Proof. First fix $\alpha \subseteq [n] \setminus U$ with $|\alpha| = s$. We will compute a bound on $\widehat{H_{x'_U}}^2(\alpha)$. Let U be the fixed block of the CBD distribution $\mathcal{D}$. By fixing $y_U = y^*_U, b_U = b^*_U$ and $x_U = x'_U$ in $\bar{\mu}_*(x, y, b)$ some of the distinct Fourier basis functions in the expansion of $\bar{\mu}_*$ will restrict to the same function, and the corresponding Fourier coefficients will add. In particular, let $N(U)$ be the set of coordinates of negations for scopes $S \in U$ and $V(U)$ the set of coordinates for variables in the scopes in U. The Fourier coefficients will, after the restriction on U, be the result of adding up terms as follows

$$\widehat{\mu_*}|_U(\alpha,\beta,\gamma) = \sum_{\substack{A\subseteq V(U)\\ B\subseteq U\\ C\subseteq N(U)}} \widehat{\mu_*}(\alpha\cup A, \beta\cup B, \gamma\cup C)\chi_A(x'_U)\phi_B(y^*_U)\psi_C(b^*_U)$$

By Lemma 3.4.7 the magnitude of every non-zero Fourier coefficient of μ_* is bounded by

$$|\widehat{\mu_*}(\alpha,\beta,\gamma)| \leq \sqrt{pq}^{|\bar{\gamma}\cap\beta|}p^{|\bar{\gamma}\setminus\beta|}.$$

Using the fact that $|\phi(y_S)| \leq \sqrt{\frac{q}{p}}$ and $|\psi(b)| = |\chi(x)| = 1$ we have

$$\begin{aligned}|\widehat{\mu_*}|_U(\alpha,\beta,\gamma)| &\leq \sum_{\substack{A\subseteq V(U)\\ B\subseteq U\\ C\subseteq N(U)}} \sqrt{pq}^{|\bar{\gamma}\cap\beta|}p^{|\bar{\gamma}\cup B\setminus\beta|}\\ &\leq 2^{(k+1)d}\sqrt{pq}^{|\bar{\gamma}\cap\beta|}p^{|\bar{\gamma}\setminus\beta|}.\end{aligned}$$

Observe that

$$H_{x'_U}(x) = \sum_{\alpha}\chi_\alpha(x)\sum_{\beta,\gamma:\bar{\gamma}\leq d_y}\widehat{\mu_*}|_U(\alpha,\beta,\gamma)\mathop{\mathbb{E}}_{y,b}\left[\phi_\beta(y)\psi_\gamma(b)\right]$$

Thus we have

$$\begin{aligned}|\widehat{H_{x'_U}}(\alpha)| &\leq \sum_{\beta,\gamma:\bar{\gamma}\leq d_y}|\widehat{\mu_*}|_U(\alpha,\beta,\gamma)||\mathop{\mathbb{E}}_{y,b}\left[\phi_\beta(y)\psi_\gamma(b)\right]|\\ &\leq \sum_{\beta,\gamma:\bar{\gamma}\leq d_y} 2^{(k+1)d}p^{(1-\delta)|\bar{\gamma}\cap\beta|}p^{|\bar{\gamma}\setminus\beta|}\,\mathbb{I}(\widehat{\mu_*}|_U(\alpha,\beta,\gamma)\neq 0)\\ &\leq \sum_{\beta,\gamma:\bar{\gamma}\leq d_y} 2^{(k+1)d}p^{(1-\delta)|\bar{\gamma}|}\,\mathbb{I}(\widehat{\mu_*}|_U(\alpha,\beta,\gamma)\neq 0)\end{aligned}$$

Further, since we have fixed all variables x_U, y_U, b_U in μ_* corresponding to the set of scopes U, we still have that $\widehat{\mu_*}|_U(\alpha,\beta,\gamma)\neq 0$ only when $\gamma\vdash\alpha$, $r(\gamma)\geq t$ and $\beta\subseteq\bar{\gamma}$. Combining this with Lemma 3.4.9 we have

$$\begin{aligned}\widehat{H_{x'_U}}^2(\alpha) &\leq 2^{2(k+1)d}\left(\sum_{r=\frac{|\alpha|}{k}}^{d_y} p^{(1-\delta)r}\cdot C^r n^{kr-\frac{tr+|\alpha|}{2}}r^{\frac{tr+|\alpha|}{2}-r}\right)^2\\ &\leq 2C^{2(k+1)d}n^{-|\alpha|}\left(\sum_{r=\frac{|\alpha|}{k}}^{d_y} n^{\delta kr-\frac{\delta}{2}tr-(1-\delta)\varepsilon r}r^{\frac{tr+|\alpha|}{2}-r}\right)^2\end{aligned}$$

$$\leq 2C^{2(k+1)d}n^{-|\alpha|}\left(\sum_{r=\frac{|\alpha|}{k}}^{d_y} n^{-\frac{\varepsilon}{2}r}r^{\frac{tr+|\alpha|}{2}-r}\right)^2$$

Using that $d_y \leq \frac{\varepsilon}{10}\frac{\log n}{\log\log n}$ we conclude that the term with r raised to a power is negligible compared to the term with n and so

$$\widehat{H_{x'_U}}^2(\alpha) \leq O(n^{-(1+\frac{\varepsilon}{k})|\alpha|})$$

Summing over all $\binom{n}{|\alpha|}$ choices for α completes the proof. □

3.5 Conjunctive Blockwise-Dense Decompositions

In this section we show that any distribution can be decomposed into a convex combination of distributions, each of which is conjunctive blockwise dense, along with an error set which is small in an appropriate sense. The proof of this fact is inspired by that in [KMR17]. The main difference is that the error set in their decomposition has small measure under the distribution which was decomposed, whereas in our case the error set only has small measure under the background distribution $\mathcal{D}(p)$. This means our error set may actually contain all the probability mass of the decomposed measure. However, this allows us to decompose *any* distribution in this way, and is actually necessary for our application.

Lemma 3.5.1. *Let $\mathcal{D}$ be a probability distribution supported on instances $I = (y, b) \in E_{n,k} \times (E_{n,k} \times \{-1,1\}^k)$. Then there is a partition of $E_{n,k} \times (E_{n,k} \times \{-1,1\}^k)$ into subsets $A_1, \cdots A_l, B, C$ such that*

1. *For each i the distribution $\mathcal{D}|_{A_i}$ is $\frac{2}{\delta}t$-CBD with parameter δ.*

2. *$\mathbb{P}_{\mathcal{D}(p)}[B] \leq n^{k+1}\left(\frac{p}{2^k}\right)^t$.*

3. *$\mathbb{P}_{\mathcal{D}}[C] \leq O(\exp(-n))$.*

Proof. The proof follows from the analysis of a greedy algorithm which constructs the desired partition along with the two error sets.

Algorithm 3.5.2. Let $\mathcal{D}$ be as in the statement of the lemma. We define a recursive function which takes as input a set of instances A and builds up a partition as described in the lemma. Initially let $B = \emptyset$, let $C = \{I \mid |I| = (1 \pm \varepsilon)\Delta n\}$ and let A be the set of all instances excluding C.

Decompose(A).

1. If $\mathcal{D}|_A$ is blockwise-dense, then add A to the partition and end the recursion.
2. If $\mathbb{P}_{\mathcal{D}(p)}[A] \leq 2^{-kt}p^t$, then assign $B \leftarrow A \cup B$, and end the recursion.
3. Set $(A', B') \leftarrow$**Truncate(A)**, then assign $B \leftarrow B' \cup B$. If $\mathbb{P}_{\mathcal{D}}(A') \leq \exp(-n)$ set $C \leftarrow A' \cup C$ and end the recursion.
4. Choose $U \subseteq E_{n,k}$ to be a maximal set such that there exist I_U^* with
$$\mathop{\mathbb{P}}_{\mathcal{D}|_{A'}}[I_U = I_U^*] > \mathop{\mathbb{P}}_{\mathcal{D}(p)}[I_U = I_U^*]^{1-\delta}$$
5. Assign $A_0 \leftarrow A' \cap \{I \mid I_U = I_U^*\}$.
6. Assign $A_1 \leftarrow A' \cap \{I \mid I_U \neq I_U^*\}$.
7. Add A_0 to the partition and call **Decompose(A_1)**.

The algorithm calls the following truncation subroutine, which iteratively removes instances from a set A so as to truncate those which appear with much higher probability under $\mathcal{D}$ than under $\mathcal{D}(p)$.

Algorithm 3.5.3. This subroutine takes as input a set of instances A and constructs a partition of A into two sets A', B'.

Truncate(A).

1. If $\mathbb{P}_{\mathcal{D}}[A] \leq \exp(-n)$ set $A' \leftarrow A$ and terminate.
2. If for all $I' \in A$ we have $\mathbb{P}_{\mathcal{D}|_A}[I = I'] < 2^{kt}p^{-t}\,\mathbb{P}_{\mathcal{D}(p)|_A}[I = I']$, set $A' \leftarrow A$ and terminate.
3. Else choose I^* which maximizes $\frac{\mathbb{P}_{\mathcal{D}|_A}[I=I^*]}{\mathbb{P}_{\mathcal{D}(p)|_A}[I=I^*]}$ and set $B' \leftarrow B' \cup \{I^*\}$.
4. Call **Truncate($A \setminus I^*$)**.

We prove the lemma through a series of claims.

Claim 3.5.4. In any execution of **Decompose** that makes it to the end of the function we must have $\mathcal{D}|_{A_0}$ is conjunctive blockwise-dense with fixed block U.

Proof. To see why let $V \subseteq E_{n,k} \setminus U$ and suppose that there exists I'_V such that $\mathbb{P}_{\mathcal{D}|_{A_0}}[I_V = I'_V] > \mathbb{P}_{\mathcal{D}(p)}[I_V = I'_V]^{1-\delta}$. Then

$$\begin{aligned}\mathop{\mathbb{P}}_{\mathcal{D}|_A}[I_{U\cup V} = (I_U^*, I'_U)] &\geq \mathop{\mathbb{P}}_{\mathcal{D}(p)}[I_U = I_U^*]^{1-\delta} \cdot \mathop{\mathbb{P}}_{\mathcal{D}(p)}[I_V = I'_V]^{1-\delta} \\ &= \mathop{\mathbb{P}}_{\mathcal{D}(p)}[I_{U\cup V} = (I_U^*, I'_U)]^{1-\delta}\end{aligned}$$

which contradicts the maximality of U. □

Next we show that **Truncate** does not remove too much probability mass from $\mathcal{D}(p)$.

Claim 3.5.5. If $(A', B') = \textbf{Truncate}(A)$, then $\mathbb{P}_{\mathcal{D}(p)}[A'] \geq \left(1 - n\left(\frac{p}{2^k}\right)^t\right)\mathbb{P}_{\mathcal{D}(p)}[A]$.

Proof. Let A_i denote the input to the i-th recursive call to **Truncate**. Now note that in each recursive call if $\mathbb{P}_{\mathcal{D}(p)}[A_i]$ decreases by a factor of $(1-\eta)$ after removing $I*$, then $\mathbb{P}_{\mathcal{D}}[A_i]$ decreases by a factor of $(1 - \eta\left(\frac{p}{2^k}\right)^{-t})$. Thus after $n\eta^{-1}\left(\frac{p}{2^k}\right)$ recursive calls,

$$\mathop{\mathbb{P}}_{\mathcal{D}(p)}[A_i] = \exp\left(-n\left(\frac{p}{2^k}\right)^t\right)\mathop{\mathbb{P}}_{\mathcal{D}(p)}[A_0] \geq \left(1 - n\left(\frac{p}{2^k}\right)^t\right)\mathop{\mathbb{P}}_{\mathcal{D}(p)}[A_0]$$

while on the other hand

$$\mathop{\mathbb{P}}_{\mathcal{D}}[A_i] = \exp(-n)\mathop{\mathbb{P}}_{\mathcal{D}}[A_0] \leq \exp(-n).$$

Since the subroutine terminates once $\mathbb{P}_{\mathcal{D}}[A_i] \leq \exp(-n)$, the claim is proven. □

Next we show that the fixed block U is not too large.

Claim 3.5.6. $|U| \leq \frac{2}{\delta}t$.

Proof. If the call to **Truncate** does not return a set A' with $\mathbb{P}_{\mathcal{D}}[A'] \leq \exp(-n)$, then we must have for all I'

$$\mathop{\mathbb{P}}_{\mathcal{D}|_{A'}}[I = I'] < 2^{kt}p^{-t}\mathop{\mathbb{P}}_{\mathcal{D}(p|_{A'})}[I = I'].$$

Thus, for any restricted instance I'_U after the truncation step

$$\mathop{\mathbb{P}}_{\mathcal{D}|_{A'}}[I_U = I'_U] \leq 2^{kt}p^{-t}\mathop{\mathbb{P}}_{\mathcal{D}(p)|_{A'}}[I_U = I'_U].$$

Thus, we have that for the instance I^*_U,

$$\mathop{\mathbb{P}}_{\mathcal{D}(p)}[I_U = I^*_U]^{1-\delta} \leq 2^{kt}p^{-t}\mathop{\mathbb{P}}_{\mathcal{D}(p)|_{A'}}[I_U = I^*_U] = 2^{kt}p^{-t}\frac{\mathbb{P}_{\mathcal{D}(p)|_{A'}}[I_U = I^*_U]}{\mathbb{P}_{\mathcal{D}(p)}[A']}.$$

Since we have passed the second step in the algorithm, $\mathbb{P}_{\mathcal{D}(p)}[A'] \geq 2^{-kt}p^t$. Using this fact and rearranging terms gives

$$2^{2kt}p^{-2t} \geq \mathop{\mathbb{P}}_{\mathcal{D}(p)}[I_U = I^*_U]^{-\delta} \geq 2^{\delta k|U|}p^{-\delta|U|},$$

where the last inequality comes from the definition of $\mathcal{D}(p)$. Thus $|U| \leq \frac{2}{\delta}t$. □

To wrap up the proof of the lemma, note that **Decompose** is called recursively at most n^k times, because we fix at least one constraint scope in every call that does not terminate. Thus, by 3.5.5 the total probability mass added to B is at most $n^{k+1}\left(\frac{p}{2^k}\right)^t$. □

3.6 Proof of LP Lower Bounds

In this section we prove lower bounds for LP formulations of CSPs. We proceed with the approach based on pseudo-calibration as introduced in Section 3.3. In particular, assume there exists an LP formulation for a CSP which certifies the upper bound c on a subset of instances $A \subset \{-1,1\}^{|E_{n,k}|} \times \{-1,1\}^{|E_{n,k}|k}$ of measure s under $\mathcal{D}(p)$. We will identify a subset $B \subseteq A$ of instances and some $\lambda > 0$ such that both

$$\mathop{\mathbb{E}}_{\{-1,1\}^n \times \mathcal{D}(p)} \left[\mathbb{I}((y,b) \in B)\bar{\mu}_*(x,y,b)(c - F(x,y,b))\right] < -\lambda \tag{3.6.1}$$

and

$$\mathop{\mathbb{E}}_{\{-1,1\}^n \times \mathcal{D}(p)} \left[\mathbb{I}((y,b) \in B)\bar{\mu}_*(x,y,b) \sum_{i=0}^{R} p_i(y,b)q_i(x)\right] \geq -\lambda. \tag{3.6.2}$$

This will then contradict the fact that the LP formulation certifies the upper bound c on the instances in A.

Setting Parameters. For the rest of this section let $d = (d_x, d_y)$ where $d_x = d_y = \varepsilon\frac{\log n}{10 \log\log n}$, and let $\bar{\mu}_*$ be the d-pseudo-calibration of μ_*. Additionally, set $p = \frac{\Delta n}{|E_{n,k}|}$ where $\Delta n = n^{\frac{t}{2}-\varepsilon}$.

The first step will be to prove that $\mathbb{E}_x\left[\bar{\mu}_*(x,y,b)(c - F(x,y,b))\right]$ is negative with high probability over instances $I = (y,b)$. This follows almost immediately whenever $\bar{\mu}_*(x,y,b)$ is a Sherali-Adams pseudo-density with high probability over instances.

Lemma 3.6.1. *Let $c = (1-\eta)\Delta n$ and assume $\bar{\mu}_*(x,y,b)$ is a Sherali-Adams pseudo-density with probability $1 - o(1)$ over instances (y,b). Let $m = m(y,b)$ be the number of constraints in the instance (y,b) and let E be the event that both:*

1. $|m(y,b) - \Delta n| \leq \frac{\eta}{2}\Delta n$
2. $\mathbb{E}_x\left[\bar{\mu}_*(x,y,b)(c - F(x,y,b))\right] \leq -\frac{\eta}{2}\Delta n.$

Then

$$\mathop{\mathbb{P}}_{(y,b)\sim\mathcal{D}(p)}[E] = 1 - o(1).$$

Proof. For all (y,b) such that $\bar{\mu}_*$ is a Sherali-Adams pseudo-density and $m \geq (1-\frac{\eta}{2})\Delta n$

$$\mathop{\mathbb{E}}_{x}\left[\bar{\mu}_*(c - F)\right] = c - m \geq \frac{\eta}{2}\Delta n$$

By Lemma 3.2.3 $|m(y,b) - \Delta n| \leq \frac{\eta}{2}\Delta n$ with probability at least $1 - 2\exp^{\Omega(\eta^2\Delta n)} = 1 - o(1)$. Since $\bar{\mu}_*$ is a Sherali-Adams pseudo-density with probability $1 - o(1)$ the desired result follows by the union bound. □

The next lemma will rely on the following concentration bound based on hypercontractivity.

Theorem 3.6.2 ([O'D14] Chapter 9)**.** *Let $f : \{-1,1\} \to \mathbb{R}$ have degree at most k. Then for any $t \geq \sqrt{2e}^k$ we have*

$$\mathop{\mathbb{P}}_{x\sim\{-1,1\}^n}\left[|f(x)| \geq t\|f\|_2\right] \leq \exp\left(-\frac{k}{2e}t^{\frac{2}{k}}\right).$$

Lemma 3.6.3. *Let $d \leq \frac{d_x}{k}$. Let $\mathcal{D}$ be a d-CBD distribution with parameter $\delta \leq \frac{\varepsilon}{2k}$ and fixed block U. Suppose further that $\mathcal{D}$ is supported on instances (y,b) where $\bar{\mu}_*(x,y,b)$ is a Sherali-Adams pseudo-density. For each fixed let $H(x) = \mathbb{E}_{\mathcal{D}}\left[\bar{\mu}_*\right]$. Then*

$$\mathop{\mathbb{P}}_{x}\left[H(x) \leq -O\left(n^{-\frac{\varepsilon}{4k}}\right)\right] \leq O\left(\exp\left(-\frac{1}{2e}n^{\frac{\varepsilon}{2k}}\right)\right).$$

Proof. We begin by breaking up the probability by conditioning on each possible setting of the variables x_U.

$$\mathop{\mathbb{P}}_{x}\left[H(x) \leq -O\left(n^{-\frac{\varepsilon}{4k}}\right)\right] = \frac{1}{2^{k|U|}}\sum_{x'_U}\mathop{\mathbb{P}}_{x_{U^c}}\left[H(x'_U, x_{U^c}) \leq -O\left(n^{-\frac{\varepsilon}{4k}}\right)\right] \tag{3.6.3}$$

Next let $H_{x'_U}(x) = H(x'_U, x)$. Since $\mathcal{D}$ is supported only on instances where $\bar{\mu}_*(x,y,b)$ is a Sherali-Adams pseudo-density we have that

$$\begin{aligned}\mathop{\mathbb{E}}_{x}\left[H_{x'_U}(x)\right] &= \mathop{\mathbb{E}}_{x}\left[H(x)\,\mathbb{I}(x_U = x'_U)\right]\\ &= \mathop{\mathbb{E}}_{\mathcal{D}}\left[\mathop{\mathbb{E}}_{x}\left[\bar{\mu}_*(x,y,b)\,\mathbb{I}(x_U = x'_U)\right]\right]\\ &\geq 0\end{aligned}$$

because the inner expectation is the average of a Sherali-Adams pseudo-density times a non-negative junta on $kd \leq d_x$ variables.

Let $H^{=s}_{x'_U}$ denote the degree s part of $H_{x'_u}$. We have by Lemma 3.4.10 that

$$\left\|H^{=s}_{x'_U}\right\|_2^2 \leq O\left(n^{-\frac{\varepsilon}{k}s}\right).$$

Therefore by Theorem 3.6.2 we have

$$\mathop{\mathbb{P}}_{x}\left[|H^{=s}_{x'_U}(x)| \geq tO\left(n^{-\frac{\varepsilon}{2k}s}\right)\right] \leq \exp\left(-\frac{s}{2e}t^{\frac{2}{s}}\right).$$

Setting $t = n^{\frac{\varepsilon s}{4k}}$ and summing over $s \geq 1$ yields

$$\mathop{\mathbb{P}}_{x}\left[\left|\sum_{s=1}^{d_x} H^{=s}_{x'_U}(x)\right| \geq O\left(n^{-\frac{\varepsilon}{4k}}\right)\right] \leq O\left(\exp\left(-\frac{1}{2e}n^{\frac{\varepsilon}{2k}}\right)\right).$$

Since $H_{x'_U}(x) = \mathbb{E}\left[H_{x'_U}\right] + \sum_{s=1}^{d_x} H^{=s}_{x'_U}(x)$ and we already have that $\mathbb{E}\left[H_{x'_U}\right] \geq 0$, we conclude that

$$\mathop{\mathbb{P}}_{x}\left[\left|H_{x'_U}(x)\right| \geq -O\left(n^{-\frac{\varepsilon}{4k}}\right)\right] \leq O\left(\exp\left(-\frac{1}{2e}n^{\frac{\varepsilon}{2k}}\right)\right).$$

Plugging this in to 3.6.3 completes the proof. □

We are now ready to prove our main result.

Theorem 3.6.4. *Let P be a predicate supporting a $(t-1)$-wise uniform distribution on satisfying assignments. Let $\Delta n = n^{\frac{t}{2}-\varepsilon}$ and set $p = \frac{\Delta n}{|E_{n,k}|}$. Assume $\bar{\mu}_*(x, y, b)$ is a Sherali-Adams pseudo-density with probability $1 - o(1)$ over $\mathrm{CSP}(P)$ instances $(y, b) \sim \mathcal{D}(p)$. Then for any constant $\eta > 0$, any linear programming formulation that η-refutes random instances of $\mathrm{CSP}(P)$ with constant probability must have size at least $n^{\Omega\left(\varepsilon^2 \frac{\log n}{\log \log n}\right)}$.*

Proof. Let $c > 0$ be a sufficiently small constant to be set later. Suppose that there is a linear programming formulation of size $R \le n^{c\varepsilon^2 \frac{\log n}{\log \log n}}$ for η-refuting random instances of $\mathrm{CSP}(P)$. By Lemma 2.3.3 this means there exists a set A of constant measure under $D(p)$, and non-negative functions $p_i(y, b), q_i(x)$ such that, for $c = (1-\eta)\Delta n$:

$$c - F(x, y, b) = \sum_{i=0}^{R} p_i(y, b) q_i(x) \tag{3.6.4}$$

whenever $(y, b) \in A$. We will derive a contradiction by multiplying each side of the above equation by $\bar{\mu}_*$ and then averaging over (x, y, b).

Let $m(y, b)$ be the number of constraints in (y, b) and let E be the event defined in Lemma 3.6.1. By the lemma, $\mathbb{P}_{\mathcal{D}(p)}[E] = 1 - o(1)$. Thus, letting $A' = A \cap E$ we have $\mathbb{P}_{\mathcal{D}(p)}[A']$ is constant, because A has constant measure. We now restrict (3.6.4) to instances in A' and normalize the equation by dividing both sides by Δn. Note that by the definition of E we have $|\frac{1}{\Delta n}(c - F(x, y, b))| \le 1 + \frac{\eta}{2}$ for all x, and for all $(y, b) \in E$.

Thus, we must have the same bound for the right hand side of (3.6.4)

$$\frac{1}{\Delta n} \sum_{i=0}^{R} p_i(y, b) q_i(x) \le 1 + \frac{\eta}{2} \tag{3.6.5}$$

Now, since the p_i, are non-negative, we can re-normalize each p_i to be a density relative to $\mathcal{D}(p)$ by simply rescaling p_i by $\mathbb{E}[p_i]^{-1}$ and q_i by $\mathbb{E}[p_i]$. After this rescaling, averaging (3.6.5) over (y, b) implies that $\sum_i q_i(x) \le 1 + \frac{\eta}{2}$ for all x.

Let $\mathcal{D}_i$ be the probability distribution given by the density p_i, let $\delta \le \frac{\varepsilon}{2k}$ and let $r = \varepsilon^2 \frac{\log n}{100k^2 \log \log n}$. By Lemma 3.5.1 we can partition A' into sets $A_1, \ldots A_N, B_i, C_i$ such that each A_j is a $\frac{2}{\delta}r$-CBD distribution with parameter δ, $\mathbb{P}_{\mathcal{D}(p)}[B_i] \le n^{k+1}\left(\frac{p}{2^k}\right)^r$, and $\mathbb{P}_{\mathcal{D}_i}[C_i] \le O(\exp(-n))$. Note that with these parameter settings we have $\frac{2}{\delta}r \le \varepsilon \frac{\log n}{10 \log n}$ i.e. each density is d-CBD for $d \le \frac{d_x}{k}$. Letting $B_i' \overset{\text{def}}{=} A' \cap B_i$ and $H_{i,j}(x) \overset{\text{def}}{=} \mathbb{E}_{\mathcal{D}_i | A_j}[\bar{\mu}_*]$ yields

$$\begin{aligned} \mathbb{E}\left[\mathbb{I}((y, b) \notin B_i')\bar{\mu}_*(x, y, b) p_i(y, b)\right] &= \sum_j \mathbb{P}_{\mathcal{D}_i}[A_j] \mathop{\mathbb{E}}_{\mathcal{D}_i | A_j}[\bar{\mu}_*] + \mathbb{P}_{\mathcal{D}_i}[C_i] \mathop{\mathbb{E}}_{\mathcal{D}_i | C_i}[\bar{\mu}_*] \\ &= \sum_j \mathbb{P}_{\mathcal{D}_i}[A_j] H_{i,j}(x) + \mathbb{P}_{\mathcal{D}_i}[C_i] \mathop{\mathbb{E}}_{\mathcal{D}_i | C_i}[\bar{\mu}_*] \end{aligned}$$

Since $\mathbb{P}_{\mathcal{D}_i}[C_i] \leq O(\exp(-n))$ the second term above is bounded in magnitude by

$$|\mathbb{P}_{\mathcal{D}_i}[C_i] \mathop{\mathbb{E}}_{\mathcal{D}_i|C_i}[\bar{\mu}_*]| \leq O(\exp(-n))\|\bar{\mu}_*\|_\infty \leq O(\exp(-n)) \cdot n^{d_x+2dy}. \tag{3.6.6}$$

Further, since each distribution $\mathcal{D}_i$ is d-CBD for $d \leq \frac{d_x}{k}$ we have by Lemma 3.6.3 that for each i, j

$$\mathop{\mathbb{P}}_x\left[H_{i,j}(x) \leq -O\left(n^{-\frac{\varepsilon}{4k}}\right)\right] \leq O\left(\exp\left(-\frac{1}{2e}n^{\frac{\varepsilon}{2k}}\right)\right).$$

Since the $\sum_i q_i(x) \leq 1 - \frac{\eta}{2}$ we conclude that

$$\sum_{i,j} \mathop{\mathbb{P}}_{\mathcal{D}_i}[A_j] \mathop{\mathbb{E}}_x[q_i(x)H_{i,j}(x)] \geq \sum_j \mathop{\mathbb{P}}_{\mathcal{D}_i}[A_j] \cdot \left(-O\left(n^{-\frac{\varepsilon}{4k}}\right)\right) \geq -O\left(n^{-\frac{\varepsilon}{4k}}\right)$$

Let $B = \cup_i B_i'$. Then the above inequality combined with (3.6.6) implies that

$$\sum_i \mathbb{E}\left[\mathbb{I}((y,b) \notin B)\bar{\mu}_*(x,y,b)p_i(y,b)q_i(x)\right] \geq -O\left(n^{-\frac{\varepsilon}{4k}}\right).$$

To summarize, averaging the right hand side of (the normalized version of) (3.6.4) multiplied by $\bar{\mu}_*$ and restricted to instances not in B has expectation that is at least a small negative number. Now observe that by our choice of r and R we have

$$\mathop{\mathbb{P}}_{\mathcal{D}(p)}[B] \leq Rn^{k+1}\left(\frac{p}{2k}\right)^r \leq n^{-\Omega\left(\varepsilon^2 \frac{\log n}{\log\log n}\right)}.$$

Recall on the left hand side of (3.6.4) we have that

$$\frac{1}{\Delta n}\mathop{\mathbb{E}}_x\left[\bar{\mu}_*(x,y,b)(c - F(x,y,b))\right] \leq -\frac{\eta}{2}$$

whenever $(y,b) \in A'$. Further, $B \subseteq A'$ and A' has constant measure under $\mathcal{D}(p)$, so B cannot be all of A'. We conclude that

$$\frac{1}{\Delta n}\mathop{\mathbb{E}}_{x,y,b}\left[\mathbb{I}((y,b) \notin B)\bar{\mu}_*(x,y,b)(c - F(x,y,b))\right] \leq -\frac{\eta}{2}$$

For $\eta \geq O\left(n^{-\frac{\varepsilon}{4k}}\right)$ this yields the desired contradiction. □

Finally, Lemma 3.4.8 implies that $\bar{\mu}_*$ is a Sherali-Adams pseudo-density with high probability for both MAX k-SAT and MAX k-XOR. This immediately proves Theorem 3.1.1.

Chapter 4

Semidefinite Programming Lower Bounds

4.1 Introduction to SDP Lower Bounds

In his seminal work, [Yan91, Yan88] showed that any symmetric linear program for the matching problem has exponential size. [Rot14] recently showed that one can drop the symmetry requirement: any linear program for the matching problem has exponential size. Since it is possible to optimize over matchings in polynomial time, it follows that there is a gap between problems that have small linear formulations and problems that allow efficient optimization.

In light of this gap, it is reasonable to ask whether semidefinite programming (SDP) can characterize all problems that allow efficient optimization. Semidefinite programs generalize linear programs and can be solved efficiently both in theory and practice (see [VB96]). SDPs are the basis of some of the best algorithms currently known, for example the approximation of [GW95] for MAX CUT .

Following prior work (see for example [GPT11]) we define the size of an SDP formulation as the dimension of the psd cone from which the polytope can be obtained as an affine slice. Some recent work has shown limits to the power of small SDPs. [BDP13, BDP15] nonconstructively give an exponential lower bound on the size of SDP formulations for most 0/1 polytopes. [LRS15] give an exponential lower bound for solving the traveling salesperson problem (TSP) and approximating MAX 3-SAT. However the question of whether the matching problem has a small SDP remains open. We give a partial negative answer to this question by proving the analog of Yannakakis's result for semidefinite programs:

Theorem. *Any symmetric SDP for the matching problem has exponential size.*

We also show that for the asymmetric metric traveling salesperson problem the optimal symmetric semidefinite formulation of a given size is essentially achieved by the respective level of the Lasserre hierarchy.

Main Result

The question of whether the matching problem admits a small SDP relaxation remains open. The main result of this chapter is an analogue of the theorem of [Yan91, Yan88] for SDP relaxations of the matching problem. Specifically, we show the following.

Theorem 4.1.1. *There exists an absolute constant $\alpha > 0$ such that for every $\varepsilon \in [0, 1)$, every* symmetric *SDP relaxation approximating the perfect matching problem within a factor $1 - \frac{\varepsilon}{n-1}$ has size at least $2^{\alpha n}$.*

Analogous to the work of [LRST14] on MaxCSPs, we will show that among all symmetric SDP relaxations for the matching problem, the Lasserre SDP hierarchy is optimal. We will then appeal to a result by [Gri01] that shows that $\Omega(n)$-rounds of the Lasserre SDP hierarchy cannot refute the existence of a perfect matching in an odd clique of size n.

The key technical obstacle in going from MaxCSPs to the matching problem is the non-trivial algebraic structure of the underlying solution space, namely the space of all perfect matchings. Specifically, given a multilinear polynomial F, testing whether the polynomial F is identically zero over all perfect matchings is non-trivial in itself. In contrast, a multilinear polynomial is nonzero over the solution space of a MaxCSP, namely the set $\{0, 1\}^n$, if and only if all the coefficients of the polynomial are zero. Roughly speaking, for the Lasserre SDP relaxation to be optimal for the matching problem, it must at least be powerful enough to detect whether a given polynomial is identically zero over matchings. We show that every multilinear polynomial F that is identically zero over all perfect matchings can be certified as such via a degree $2\deg(F) - 1$ derivation, starting from the linear and quadratic constraints that define the space of perfect matchings.

Our second result shows the optimality of Lasserre SDP relaxations among all symmetric SDP relaxations for approximating the asymmetric metric traveling salesperson problem. The formal statement of the result is as follows.

Theorem. *For every constant $\rho > 0$, if there exists a symmetric SDP relaxation of size $r < \sqrt{\binom{2n}{k}} - 1$ which achieves a ρ-approximation for asymmetric metric TSP instances on $2n$ vertices, then the $(2k-1)$-round Lasserre relaxation achieves a ρ-approximation for asymmetric metric TSP instances on n vertices.*

4.2 The perfect matching problem

We now present the *perfect matching problem* $\mathrm{PM}(n)$ as a maximization problem in the framework of Section 2.4 and show that any symmetric SDP formulation has exponential size.

Let n be an even positive integer, and let K_n denote the complete graph on n vertices. The feasible solutions of $\mathrm{PM}(n)$ are all the perfect matchings M on K_n. The objective functions f_F are indexed by the edge sets F of K_n and are defined as $f_F(M) := |M \cap F|$. For approximation guarantees we use $\tilde{S}(f) := \max f$ and $\tilde{C}(f) := \max f + \varepsilon/2$ for some fixed $0 \leq \varepsilon < 1$ as in [BP15]. Since $\tilde{S}(f) = \max f \leq (n-1)/2$ when f is associated with an odd set, we have $(1 - \varepsilon/(n-1))\tilde{C}(f) \geq \tilde{S}(f)$, which establishes an inapproximability ratio of $1 - \varepsilon/(n-1)$.

The alternating group A_n acts naturally on $\mathrm{PM}(n)$ via permutation of vertices, and the guarantees $\tilde{C}, \tilde{S}$ are A_n-symmetric. Our main theorem is an exponential lower bound on the size of any A_n-coordinate-symmetric SDP extension of $\mathrm{PM}(n)$.

Theorem 4.2.1. *There exists an absolute constant $\alpha > 0$ such that for all even n and every $0 \leq \varepsilon < 1$, every A_n-coordinate-symmetric $(\tilde{C}, \tilde{S})$-approximate SDP extended formulation for the perfect matching problem* $\mathrm{PM}(n)$ *has size at least $2^{\alpha n}$. In particular, every A_n-coordinate-symmetric SDP extended formulation approximating the perfect matching problem* $\mathrm{PM}(n)$ *within a factor of* $1 - \varepsilon/(n-1)$ *has size at least* $2^{\alpha n}$.

Lower bounds on matching

A key step in proving our lower bound is obtaining low-degree derivations of approximation guarantees for objective functions of $\mathrm{PM}(n)$. Therefore we start with a standard representation of functions as polynomials. We define the *matching constraint polynomials $\mathcal{P}_n$* as:

$$\begin{aligned} \mathcal{P}_n \coloneqq & \{x_{uv}x_{uw} \mid u, v, w \in [n] \text{ distinct}\} \\ & \cup \left\{ \sum_{u \in [n], u \neq v} x_{uv} - 1 \;\middle|\; v \in [n] \right\} \\ & \cup \left\{ x_{uv}^2 - x_{uv} \mid u, v \in [n] \text{ distinct} \right\}. \end{aligned} \tag{4.2.1}$$

Intuitively, the first set of polynomials ensures that no vertex is matched more than once, the second set ensures that each vertex is matched, and the third set ensures that each coordinate is 0-1 valued. We observe that the ring of real valued functions on perfect matchings is isomorphic to $\mathbb{R}[\{x_{uv}\}_{\{u,v\} \in \binom{[n]}{2}}]/\langle \mathcal{P}_n \rangle_I$ with x_{uv} representing the indicator function of the edge uv being contained in a perfect matching.

Now we formulate low-degree derivations. Let $\mathcal{P}$ denote a set of polynomials in $\mathbb{R}[x]$. For polynomials F and G, we write $F \simeq_{(\mathcal{P},d)} G$, or *$F$ is congruent to G from $\mathcal{P}$ in degree d*, if and only if there exist polynomials $\{q(p) : p \in \mathcal{P}\}$ such that

$$F + \sum_{p \in \mathcal{P}} q(p) \cdot p = G$$

and $\max_p \deg(q(p) \cdot p) \leq d$. We often drop the dependence on $\mathcal{P}$ when it is clear from context. We shall write $F \equiv G$ for two polynomials F and G defining the same function on perfect matchings, i.e., $F - G \in \langle \mathcal{P}_n \rangle_I$.

A crucial part of our argument is the result that any $F \in \langle \mathcal{P}_n \rangle_I$ can be generated by low-degree coefficients from $\mathcal{P}_n$:

Theorem 4.2.2. *For every polynomial $F \in \mathbb{R}[\{x_{uv}\}_{\{u,v\} \in \binom{n}{2}}]$, if $F \in \langle \mathcal{P}_n \rangle_I$ then $F \simeq_{(\mathcal{P}_n, 2\deg F - 1)} 0$.*

The proof is presented in Section 4.2. We will also make use of the following proposition, whose proof appears in Section 4.2:

Proposition 4.2.3. *Let $n \geq 10$, let $k < n/2$ and let $\mathcal{H}$ be an A_n-symmetric set of functions on the set of perfect matchings of K_n of size less than $\binom{n}{k}$. Then for every $h \in \mathcal{H}$ there is a vertex set $W \subseteq [n]$ of size less than k such that h depends only on the (at most $\binom{k-1}{2}$) edges in W.*

We now have all the ingredients to present the proof of our main theorem.

Proof of Theorem 4.2.1. Fix an even integer $n \geq 10$ and let $k = \lceil \beta n \rceil$ for some small enough constant $0 < \beta < 1/2$ chosen later. Suppose for a contradiction that $\mathrm{PM}(n)$ admits a symmetric SDP extended formulation of size $d < \sqrt{\binom{n}{k}} - 1$.

Let m equal $n/2$ or $n/2 - 1$, whichever is odd. Let $S = [m]$ and let $T = \{m+1, \dots, 2m\}$. If $m = n/2$ then let $U = \{2m+1, 2m+2\}$, otherwise let $U = \varnothing$. Note that $S \cup T \cup U = [n]$ and $|S| = |T| = m = \Theta(n)$. Consider the objective function for the set of edges $E[S]$ on S. Since $|S|$ is odd we have $\max f_{E[S]} = (|S| - 1)/2$, from which we derive:

$$f(x) \overset{\text{def}}{=} \tilde{C}(f_{E[S]}) - f_{E[S]}(x) = \frac{|S| - 1}{2} + \frac{\varepsilon}{2} - \sum_{u,v \in S} x_{uv} \equiv \frac{1}{2} \sum_{u \in S, v \in T \cup U} x_{uv} - \frac{1 - \varepsilon}{2}. \tag{4.2.2}$$

By Lemma 2.4.2, as $\binom{d+1}{2} < \binom{n}{k}$, there is a constant $\mu_f \geq 0$ and an A_n-symmetric set $\mathcal{H}$ of functions of size at most $\binom{n}{k}$ on the set of perfect matchings with

$$f \equiv \sum_g g^2 + \mu_f \qquad \text{with each } g \in \langle \mathcal{H} \rangle.$$

By Proposition 4.2.3, every $h \in \mathcal{H}$ depends only on the edges within a vertex set of size less than k, and hence can be represented by a polynomial of degree less than $k/2$ over perfect matchings. As the g are linear combinations of the $h \in \mathcal{H}$, they can also be represented by polynomials of degree less than $k/2$, which we assume for the rest of the proof.

Applying Theorem 4.2.2 with (4.2.2), we conclude

$$\frac{1}{2} \sum_{u \in S, v \in T \cup U} x_{uv} - \frac{1 - \varepsilon}{2} \simeq_{(\mathcal{P}_n, 2k-1)} \sum_g g^2 + \mu_f.$$

We now apply the following substitution: set $x_{2m+1,2m+2} := 1$ if U is not empty, set $x_{u+m,v+m} := x_{uv}$ for each $uv \in E[S]$, and set $x_{uv} := 0$ otherwise. Intuitively, the substitution ensures that U is matched, ensures the matching on T is identical to the matching on S, and ensures every edge is entirely within S, T, or U. The main point is that the substitution maps every polynomial in $\mathcal{P}_n$ either to 0 or into $\mathcal{P}_m$.

Applying this substitution we obtain a new polynomial identity on the variables $\{x_{uv}\}_{\{u,v\} \in \binom{S}{2}}$:

$$-\frac{1 - \varepsilon}{2} \simeq_{(\mathcal{P}_m, 2k-1)} \sum_g g^2 + \mu_f.$$

This equation is a sum of squares refutation of the existence of a perfect matching in a clique of size m, i.e. an odd clique. By [Gri01, Corollary 2] (see also [BGIP99]), it follows that $2k - 1 = \Omega(m) = \Omega(n)$, a contradiction when β is chosen small enough. □

Proof of Proposition 4.2.3

Here we show that functions on perfect matchings with high symmetry are actually *juntas*: they depend only on the edges of a small vertex set. The key is the following lemma stating that perfect matchings coinciding on a vertex set belong to the same orbit of the pointwise stabilizer of the vertex set. For any set $W \subseteq [n]$ let $E[W]$ denote the edges of K_n with both endpoints in W.

Lemma 4.2.4. *Let $S \subseteq [n]$ with $|S| < n/2$ and let M_1 and M_2 be perfect matchings in K_n. If $M_1 \cap E[S] = M_2 \cap E[S]$ then there exists $\sigma \in A([n] \setminus S)$ such that $\sigma \cdot M_1 = M_2$.*

Proof. Let $\delta(S)$ denote the edges with exactly one endpoint in S. There are three kinds of edges: those in $E[S]$, those in $\delta(S)$, and those disjoint from S. We construct σ to handle each type of edge, then fix σ to be even.

To handle the edges in $E[S]$ we set σ to the identity on S, since $M_1 \cap E[S] = M_2 \cap E[S]$.

To handle the edges in $\delta(S)$ we note that $V(M_1 \cap \delta(S))$ equals $V(M_2 \cap \delta(S))$ when both are restricted to S, since M_1 and M_2 are perfect matchings. Therefore for each edge $(s, v) \in M_1$ with $s \in S$ and $v \notin S$ there is a unique edge $(s, w) \in M_2$ with $w \notin S$; we extend σ to map v to w for each such s.

To handle the edges disjoint from S, we again use the fact that M_1 and M_2 are perfect matchings, so the number of edges in each that are disjoint from S is the same. We extend σ to be an arbitrary bijection on those edges.

We now show that we can choose σ to be even. Since $|S| < n/2$ there is an edge $(u, v) \in M_2$ disjoint from S. Let $\tau_{u,v}$ denote the transposition of u and v and let $\sigma' := \tau_{u,v} \circ \sigma$. We have $\sigma' \cdot M_1 = \sigma \cdot M_1 = M_2$, and either σ or σ' is even. □

We also need the following lemma, which has been used extensively for symmetric linear extended formulations. See references [Yan88, Yan91, KPT10, BP11, LRST14] for examples.

Lemma 4.2.5 ([DM96, Theorems 5.2A and 5.2B])**.** *Let $n \geq 10$ and let $G \leq A_n$ be a group. If $|A_n : G| < \binom{n}{k}$ for some $k < n/2$, then there is a subset $W \subseteq [n]$ such that $|W| < k$, W is G-invariant, and $A([n] \setminus W)$ is a subgroup of G.*

We combine the previous two lemmas to prove Proposition 4.2.3.

Proof. Applying Lemma 4.2.5 to the stabilizer of h, we obtain a subset $W \subseteq [n]$ of size less than k such that h is stabilized by $A([n] \setminus W)$, i.e.,

$$h(M) = (g \cdot h)(M) = h(g^{-1} \cdot M)$$

for all $g \in A([n] \setminus W)$.

Therefore for every perfect matching M the function h is constant on the $A([n] \setminus W)$-orbit of M. As the orbit is determined by $M \cap E[W]$ by Lemma 4.2.4, so is the function value $h(M)$. Therefore h depends only on the edges in $E[W]$. □

Low-degree certificates for matching ideal membership

In this section we prove Theorem 4.2.2 showing that every degree d polynomial identically zero over perfect matchings is congruent to 0 within degree $O(d)$.

For a partial matching M, let $x_M \coloneqq \prod_{e \in M} x_e$ denote the product of edge variables for the edges in M.

We rely on the following lemma, whose proof appears in Section 4.2.

Lemma 4.2.6. *For any polynomial F, there is a constant c_F with $\sum_{\sigma \in S_n} \sigma F \simeq_{(\mathcal{P}_n, \deg F)} c_F$.*

The next lemma will allow us to apply induction:

Lemma 4.2.7. *If L is a polynomial with $L \simeq_{(\mathcal{P}_{n-2}, d)} 0$ for some d, and a, b are the two additional vertices in K_n, then $Lx_{ab} \simeq_{(\mathcal{P}_n, d+1)} 0$.*

Proof. It is enough to prove the claim for $L \in \mathcal{P}_{n-2}$. For $L = x_e^2 - x_e$ and $L = x_{uv}x_{uw}$ the claim is trivial since $L \in \mathcal{P}_n$ also. The remaining case is $L = \sum_{u \in K_{n-2}} x_{uv} - 1$ for some $v \in K_{n-2}$. Then

$$Lx_{ab} = \left(\sum_{u \in K_n} x_{uv} - 1 \right) x_{ab} - x_{av}x_{ab} - x_{bv}x_{ab} \simeq_{d+1} 0.$$

□

We are now ready to prove Theorem 4.2.2.

Proof of Theorem 4.2.2. We use induction on the degree d of F. If $d = 0$ then $F = 0$ and the statement holds trivially. (Note that $\simeq_{-1}$ is just equality.) The case $d = 1$ rephrased means that the affine space spanned by the characteristic vectors of all perfect matchings is defined by the $\sum_v x_{uv} - 1$ for all vertices u. This follows from Edmonds's description of the perfect matching polytope by linear inequalities in [Edm65].

For the case $d \geq 2$ we first prove the following claim:

Claim. If $F \in \langle \mathcal{P}_n \rangle_I$ is a degree d polynomial and $\sigma \in S_n$ is a permutation of vertices, then

$$F \simeq_{(\mathcal{P}_n, 2d-1)} \sigma F.$$

We use induction on the degree. If $d = 0$ or $d = 1$ the claim follows from the corresponding cases $d = 0$ and $d = 1$ of the theorem. For $d \geq 2$ it is enough to prove the claim when σ is a transposition of two vertices a and u. Note that in $F - \sigma F$ all monomials which are independent of both a and u cancel:

$$F - \sigma F = \sum_{e\colon a \in e \text{ or } u \in e} L_e x_e \tag{4.2.3}$$

where each L_e has degree at most $d - 1$. We now show that every summand is congruent to a sum of monomials containing edges incident to both a and u. For example, for $e = \{a, b\}$ in (4.2.3) we apply the generator $\sum_v x_{uv} - 1$ to find:

$$L_{ab}x_{ab} \simeq_{d+1} L_{ab}x_{ab} \sum_v x_{uv} \simeq_{d+1} \sum_v L_{ab}x_{ab}x_{uv}.$$

Therefore

$$F - \sigma F \simeq_{d+1} \sum_{bv} L'_{bv} x_{ab} x_{uv}$$

for some polynomials L'_{bv} of degree at most $d-1$. We may assume that L'_{bv} does not contain variables x_e with e incident to a, b, u, v, as these can be removed using generators like $x_{ab}x_{ac}$ or $x_{ab}^2 - x_{ab}$. Moreover, it can be checked that L'_{bv} is zero on all perfect matchings containing $\{a, b\}$ and $\{u, v\}$. By induction, $L'_{bv} \simeq_{(\mathcal{P}_{n-4}, 2d-3)} 0$ (identifying K_{n-4} with the graph $K_n \setminus \{a, b, u, v\}$), from which $L'_{bv} \simeq_{(\mathcal{P}_n, 2d-1)} 0$ follows by two applications of Lemma 4.2.7. (The special case $a = v, b = u$ is also handled by induction and one application of Lemma 4.2.7.) This concludes the proof of the claim.

We now apply the claim followed by Lemma 4.2.6:

$$F \simeq_{2d-1} \frac{1}{n!} \sum_{\sigma \in S_n} \sigma F \simeq_d \frac{c_F}{n!}$$

for a constant c_F. As $F \in \langle \mathcal{P}_n \rangle_I$, it must be that $c_F = 0$, and therefore $F \simeq_{2d-1} 0$. □

Deriving that symmetrized polynomials are constant

In this section we prove Lemma 4.2.6. The first step is to reduce every polynomial to a linear combination of the x_M.

Lemma 4.2.8. *For every polynomial F there is a polynomial F' with $\deg F' \leq \deg F$ and $F \simeq_{(\mathcal{P}_n, \deg F)} F'$, where all monomials of F' have the form x_M for some partial matching M.*

Proof. It suffices to prove the lemma when F is a monomial. Let $F = \prod_{e \in A} x_e^{k_e}$ for a set A of edges with multiplicities $k_e \geq 1$. From $x_e^2 \simeq_2 x_e$ it follows that $x_e^k \simeq_k x_e$ for all $k \geq 1$, hence $F \simeq_{\deg F} \prod_{e \in A} x_e$. If A is a partial matching we are done, otherwise there are distinct $e, f \in A$ with a common vertex, hence $x_e x_f \simeq_2 0$ and $F \simeq_{\deg F} 0$. □

Lemma 4.2.9. *For any partial matching M on $2d$ vertices and a vertex a not covered by M, we have*

$$x_M \simeq_{(\mathcal{P}_n, d+1)} \sum_{\substack{M_1 = M \cup \{a,u\} \\ u \in K_n \setminus (M \cup \{a\})}} x_{M_1}. \tag{4.2.4}$$

Proof. We use the generators $\sum_u x_{au} - 1$ to add variables corresponding to edges at a, and then use $x_{au}x_{uv}$ to remove monomials not corresponding to a partial matching:

$$x_M \simeq_{(\mathcal{P}_n, d+1)} x_M \sum_{u \in K_n} x_{au} \simeq_{(\mathcal{P}_n, d+1)} \sum_{\substack{M_1 = M \cup \{a,u\} \\ u \in K_n \setminus (M \cup \{a\})}} x_{M_1}.$$

□

This leads to a similar congruence using all containing matchings of a larger size:

Lemma 4.2.10. *For any partial matching M of $2d$ vertices and $d \leq k \leq n/2$, we have*

$$x_M \simeq_{(\mathcal{P}_n,k)} \frac{1}{\binom{n/2-d}{k-d}} \sum_{\substack{M' \supset M \\ |M'|=k}} x_{M'} \tag{4.2.5}$$

Proof. We use induction on $k-d$. The start of the induction is with $k=d$, when the sides of (4.2.5) are actually equal. If $k>d$, let a be a fixed vertex not covered by M. Applying Lemma 4.2.9 to M and a followed by the inductive hypothesis gives:

$$x_M \simeq_{(\mathcal{P}_n,d+1)} \sum_{\substack{M_1 = M \cup \{a,u\} \\ u \in K_n \setminus (M \cup \{a\})}} x_{M_1} \simeq_{(\mathcal{P}_n,k)} \frac{1}{\binom{n/2-d-1}{k-d-1}} \sum_{\substack{M' \supset M_1 \\ |M'|=k \\ M_1 = M \cup \{a,u\} \\ u \in K_n \setminus (M \cup \{a\})}} x_{M'}.$$

Averaging over all vertices a not covered by M, we obtain:

$$\begin{aligned} x_M &\simeq_{(\mathcal{P}_n,k)} \frac{1}{n-2d} \frac{1}{\binom{n/2-d-1}{k-d-1}} \sum_{\substack{M' \supset M_1 \\ |M'|=k \\ M_1 = M \cup \{a,u\} \\ a,u \in K_n \setminus M}} x_{M'} \\ &= \frac{1}{n-2d} \frac{1}{\binom{n/2-d-1}{k-d-1}} 2(k-d) \sum_{\substack{M' \supset M \\ |M'|=k}} x_{M'} \\ &= \frac{1}{\binom{n/2-d}{k-d}} \sum_{\substack{M' \supset M \\ |M'|=k}} x_{M'}. \end{aligned}$$

where in the second step the factor $2(k-d)$ accounts for the number of ways to choose a and u. □

We are now ready to prove the main lemma.

Proof of Lemma 4.2.6. Given Lemma 4.2.8, it suffices to prove the claim for $F = x_M$ for some partial matching M. Note that if $|M| = k$ the size of the stabilizer of M is $2^k k!(n-2k)!$, then apply Lemma 4.2.10 with $d=0$:

$$\sum_{\sigma \in S_n} \sigma x_M = 2^k k!(n-2k)! \sum_{M' \colon |M'|=k} x_{M'} \simeq_k 2^k k!(n-2k)! \binom{n/2}{k}.$$

□

4.3 The Metric Traveling Salesperson Problem (TSP) revisited

In this section, we prove that a particular Lasserre SDP is optimal among all symmetric SDP relaxations for the asymmetric metric traveling salesperson problem on K_n. The *feasible solutions* of the problem are all permutations $\sigma \in S_n$. A permutation σ corresponds to the tour in K_n in which vertex i is the $\sigma(i)$-th vertex visited. An *instance* $\mathcal{I}$ of TSP is a set of non-negative distances $d_{\mathcal{I}}(i, j)$ for each edge $(i, j) \in K_n$, obeying the triangle inequality. The value of a tour σ is just the sum of the distances of edges traversed $\operatorname{val}_{\mathcal{I}}(\sigma) = \sum_i d_{\mathcal{I}}(\sigma^{-1}(i), \sigma^{-1}(i + 1))$. The *objective functions* are all the $\operatorname{val}_{\mathcal{I}}$. Note that TSP is a *minimization* problem rather than a maximization problem, but the framework presented in Section 2.4 generalizes naturally to minimization problems by just flipping the inequalities. We shall use approximation guarantees $\tilde{S}(f) = \min f$ and $\tilde{C}(f) = \min f/\rho$ for a factor $\rho \geq 1$, and for clarity, instead of $(\tilde{C}, \tilde{S})$-approximate formulation we shall use formulation within a factor ρ.

The natural action of A_n on TSP is by permutation of vertices, which means here that A_n acts on S_n by composition from the left: $(\sigma_1 \cdot \sigma_2)(i) = \sigma_1(\sigma_2(i))$. Obviously, the problem TSP is A_n-symmetric.

The ring of real-valued functions on the set S_n of feasible solutions is isomorphic to $\mathbb{R}[\{x_{ij}\}_{\{i,j\}\in[n]}]/\langle \mathcal{Q}_n \rangle_I$, with x_{ij} being the indicator of $\sigma(i) = j$, and $\mathcal{Q}_n$ is the set of *TSP constraints*:

$$\begin{aligned}\mathcal{Q}_n = &\left\{ \sum_{i\in[n]} x_{ij} - 1 \,\middle|\, j \in [n] \right\} \cup \left\{ \sum_{j\in[n]} x_{ij} - 1 \,\middle|\, i \in [n] \right\} \\ &\cup \{x_{ij}x_{ik} \mid i, j, k \in [n]\} \cup \{x_{ij}x_{kj} \mid i, j, k \in [n]\} \\ &\cup \left\{x_{ij}^2 - x_{ij} \,\middle|\, i, j \in [n]\right\}.\end{aligned}$$

We emphasize that the variables x_{ij} do not directly encode the edges of a Hamiltonian cycle but instead specify a permutation of n vertices, encoded as perfect bipartite matching on $K_{n,n}$.

The Lasserre Hierarchy for TSP is defined as follows. The (dual of) the k-th level Lasserre SDP relaxation for a TSP instance $\mathcal{I}$ is given by

$$\begin{aligned}&\text{Maximize } C \\ &\text{subject to } \operatorname{val}_{\mathcal{I}} - C \simeq_{(\mathcal{Q}_n,k)} \sum_p p^2 \qquad \text{for some polynomials } p.\end{aligned}$$

We now state our main theorem regarding optimal SDP relaxations for TSP.

Theorem 4.3.1. *Suppose that there is some coordinate A_{2n}-symmetric SDP relaxation of size $r < \sqrt{\binom{n}{k}} - 1$ approximating TSP within some factor $\rho \geq 1$ for instances on $2n$ vertices. Then the $(2k - 1)$-level Lasserre relaxation approximates TSP within the factor of ρ on instances on n vertices.*

To prove Theorem 4.3.1 there is an equivalent of Proposition 4.2.3 we need for TSP tours, so that a small set of invariant functions depends only on the positions of a small number of indices. We start with the following proposition.

Proposition 4.3.2. *Let $\mathcal{H}$ be an A_n-symmetric set of functions of size $\binom{n}{k}$ on the set of TSP tours $\sigma \in S_n$. Then for every $h \in \mathcal{H}$ there is a set $W \subseteq [n]$ of size less than k, such that $h(\sigma)$ depends only on the positions of the vertices in W in the tour σ, and the sign of σ as a permutation.*

Proof. For every $h \in \mathcal{H}$ we can apply Lemma 4.2.5 to the stabilizer of h to obtain a subset $W \subseteq [n]$ of size at most k such that h is stabilized by $A([n] \setminus W)$. Thus for every tour σ, h is constant on the $A([n] \setminus W)$-orbit of σ. This orbit is clearly determined by the positions of the vertices in W and, since $A([n] \setminus W)$ preserves signs, the sign of the permutation σ. □

Next we give a reduction which allows us to eliminate the dependence of the functions $h \in \mathcal{H}$ on the sign of the permutation σ. In particular we encode every TSP tour σ on an n vertex graph as some new tour $\Phi(\sigma)$ in a $2n$ vertex graph, such that $\Phi(\sigma)$ is always an even permutation in S_{2n}.

Lemma 4.3.3. *Let $\mathcal{I}$ be an instance of TSP on K_n. Then there exists an instance $\mathcal{I}'$ of TSP on K_{2n} and an injective map $\Phi : S_n \to S_{2n}$ such that*

1. $\mathrm{val}_{\mathcal{I}}(\sigma) = \mathrm{val}_{\mathcal{I}'}(\Phi(\sigma))$ *for all* $\sigma \in S_n$.
2. *For every tour* $\tau \in S_{2n}$ *there exists* $\sigma \in S_n$ *such that* $\mathrm{val}_{\mathcal{I}'}(\Phi(\sigma)) \leq \mathrm{val}_{\mathcal{I}'}(\tau)$
3. *For all* $\sigma \in S_n$ *the permutation* $\Phi(\sigma)$ *is even.*

Proof. Given a TSP instance $\mathcal{I}$ on K_n we construct a new instance $\mathcal{I}'$ on K_{2n} as follows:

- For every vertex $i \in \mathcal{I}$ add a pair of vertices i and i' to $\mathcal{I}'$.
- For every distance $d(i,j)$ in $\mathcal{I}$ add 4 edges all with the same distance $d(i,j) = d(i',j) = d(i,j') = d(i',j')$ to $\mathcal{I}'$.
- For every pair of vertices $i, i' \in \mathcal{I}'$ add an edge of distance zero, i.e. set $d(i,i') = 0$.

We will call a tour $\tau \in S_{2n}$ *canonical* if it visits i' immediately after i, i.e. $\sigma(i') = \sigma(i) + 1$. We will write T for the set of canonical tours in S_{2n}. It is easy to check using the triangle inequality that for every tour τ there is a canonical tour with no larger value. For every tour σ in $\mathcal{I}$ define $\Phi(\sigma)$ to be the corresponding canonical tour in $\mathcal{I}'$. That is $\Phi(\sigma)(i) = 2\sigma(i) - 1$ and $\Phi(\sigma)(i') = 2\sigma(i)$. Note that $\Phi : S_n \to S_{2n}$ is an injective map whose image is all of T. By construction we have:

$$\mathrm{val}_{\mathcal{I}}(\sigma) \equiv \mathrm{val}_{\mathcal{I}'}(\Phi(\sigma))$$

which proves property (1). Property (2) follows from the fact that every tour $\tau \in S_{2n}$ has a canonical tour with no larger value, and that T is the image of Φ.

For property (3), note that every canonical tour is an even permutation. To see why suppose $\sigma \in S_n$ is given by $\sigma = (i_1, j_1)(i_2, j_2), \ldots, (i_m, j_m)$ where (i,j) denotes the permutation that swaps i and j. Then $\Phi(\sigma) = (i_1, j_1)(i'_1, j'_1), \ldots, (i_m, j_m)(i'_m, j'_m)$ is comprised of $2m$ swap permutations, and is therefore even. □

The last ingredient we need is a version of Theorem 4.2.2 for the TSP.

Theorem 4.3.4. *If F is a multilinear polynomial whose monomials are partial matchings on $K_{n,n}$ and $F \in \langle \mathcal{Q}_n \rangle_I$, then $F \simeq_{(\mathcal{Q}_n, 2\deg F - 1)} 0$.*

Because $\mathcal{Q}_n$ is so similar to $\mathcal{P}_n$, it should come as no surprise that the proof of the above theorem is extremely similar to the proof of Theorem 4.2.2. We include the full proof for completeness, but defer it to Section 4.3. We now have all the tools necessary to prove Theorem 4.3.1.

Proof of Theorem 4.3.1. First let $\mathcal{I}$ be an instance of TSP on K_n. Use Lemma 4.3.3 to construct a TSP instance $\mathcal{I}'$ on K_{2n} and the corresponding map Φ. Now assume we have an arbitrary A_{2n}-symmetric SDP relaxation of size $d < \sqrt{\binom{2n}{k}} - 1$ for TSP on K_{2n}. By Lemma 2.4.2 there is a corresponding A_{2n}-symmetric family of functions $\mathcal{H}'$ of size $\binom{d+1}{2}$ such that whenever $\min_\tau \mathrm{val}_{\mathcal{I}'}(\tau) \geq \tilde{S}(\mathrm{val}_{\mathcal{I}'})$ we have:

$$\mathrm{val}_{\mathcal{I}'}(\tau) - \tilde{C}(\mathrm{val}_{\mathcal{I}'}) \equiv \sum_j h_j(\tau)^2 + \mu_{\mathcal{I}'} \qquad \text{where } h_j \in \langle \mathcal{H}' \rangle \text{ and } \mu_{\mathcal{I}'} \geq 0\,.$$

Let $h' \in \mathcal{H}'$. By Proposition 4.3.2 $h'(\tau)$ depends only on some subset W' of size at most k, and possibly on the sign of τ.

Now we restrict the above relaxation to the image of Φ. By Lemma 4.3.3 this does not change the optimum. Using the fact that $\mathrm{val}_{\mathcal{I}}(\sigma) \equiv \mathrm{val}_{\mathcal{I}'}(\Phi(\sigma))$ and setting $\mu_{\mathcal{I}} = \mu_{\mathcal{I}'}$ then gives rise to a new relaxation where whenever $\min_\sigma \mathrm{val}_{\mathcal{I}}(\sigma) \geq \tilde{S}(\mathrm{val}_{\mathcal{I}})$ we have:

$$\mathrm{val}_{\mathcal{I}}(\sigma) - \tilde{C}(\mathrm{val}_{\mathcal{I}}) \equiv \sum_j h_j(\Phi(\sigma))^2 + \mu_{\mathcal{I}} \qquad \text{where } h_j \in \langle \mathcal{H}' \rangle \text{ and } \mu_{\mathcal{I}} \geq 0$$

as $\tilde{S}(\mathrm{val}_{\mathcal{I}}) = \tilde{S}(\mathrm{val}_{\mathcal{I}'})$ and $\tilde{C}(\mathrm{val}_{\mathcal{I}}) = \tilde{C}(\mathrm{val}_{\mathcal{I}'})$ by Lemma 4.3.3. Next for each $h' \in \mathcal{H}'$ define $h : S_n \to \mathbb{R}$ by $h(\sigma) = h'(\Phi(\sigma))$. Since $\Phi(\sigma)$ is even, we then have that each h depends only on the position of some subset $W \subseteq [n]$ of size at most k. Such a function can be written as a degree k polynomial p in the variables x_{ij} so that $p(x^\sigma) \equiv f(\sigma)$ on the vertices of $P_{TSP}(n)$. Now by Theorem 4.3.4 we have that $p \simeq_{(Q_n, 2k-1)} h$. Since $\mu_{\mathcal{I}} \geq 0$ it is clearly the square of a (constant) polynomial, and we conclude that whenever $\min_\sigma \mathrm{val}_{\mathcal{I}}(\sigma) \leq \tilde{S}(\mathrm{val}_{\mathcal{I}})$ we have:

$$f_{\mathcal{I}}(x) - \min f_{\mathcal{I}}/\rho \simeq_{(\mathcal{Q}_n, 2k-1)} \sum_p p(x)^2$$

which is precisely the statement that the $(2k-1)$-level Lasserre relaxation for $P_{TSP}(n)$ is a ρ-approximation. □

Low-degree certificates for tour ideal membership

In this section we prove Theorem 4.3.4 showing that every degree d polynomial identically zero over *TSP tours* is congruent to 0 within degree $O(d)$.

Note that any partial tour τ can be thought of as a partial matching M in $K_{n,n}$, namely if $\tau(i) = j$, then M includes the edge (i, j). Because of this, it will come as no surprise that the proof proceeds in a very similar manner to Section 4.2, and hereafter we shall always refer to partial matchings on $K_{n,n}$ rather than on K_n.

For a partial matching M, let $x_M \coloneqq \prod_{e \in M} x_e$ denote the product of edge variables for the edges in M. The first step is to reduce every polynomial to a linear combination of the x_M.

Lemma 4.3.5. *For every polynomial F there is a polynomial F' with $\deg F' \leq \deg F$ and $F \simeq_{(\mathcal{Q}_n, \deg F)} F'$, where all monomials of F have the form x_M for some partial matching M.*

Proof. It is enough to prove the lemma when F is a monomial: $F = \prod_{e \in A} x_e^{k_e}$ for a set $A \subseteq E[K_{n,n}]$ of edges with multiplicities $k_e \geq 1$. From $x_e^2 \simeq_2 x_e$ it follows that $x_e^k \simeq_k x_e$ for all $k \geq 1$, hence $F \simeq_{\deg F} \prod_{e \in A} x_e$, proving the claim if A is a partial matching. If A is not a partial matching, then there are distinct $e, f \in A$ with a common vertex, hence $x_e x_f \simeq_2 0$ and $F \simeq_{\deg F} 0$. □

The rest of the proof proceeds identically to Theorem 4.2.2, but we let the symmetric group act on polynomials slightly differently. If $K_{n,n} = U_n \cup V_n$ is the bipartite decomposition of $K_{n,n}$, then we only let the permutation group act on the labels of vertices of U_n, i.e. $\sigma x_{(a,b)} = x_{(\sigma(a),b)}$. We show that under this action, symmetrized polynomials are congruent to a constant, which can again be seen in the same sequence of lemmas:

Lemma 4.3.6. *For any partial matching M on $2d$ vertices and a vertex $a \in U_n$ not covered by M, we have*

$$x_M \simeq_{(\mathcal{Q}_n, d+1)} \sum_{\substack{M_1 = M \cup \{a,u\} \\ v \in V_n \setminus (M \cap V_n)}} x_{M_1}. \tag{4.3.1}$$

Proof. We use the generators $\sum_v x_{av} - 1$ to add variables corresponding to edges at a, and then use $x_{av} x_{bv}$ to remove monomials not corresponding to a partial matching:

$$x_M \simeq_{(\mathcal{Q}_n, d+1)} x_M \sum_{v \in V_n} x_{av} \simeq_{(\mathcal{Q}_n, d+1)} \sum_{\substack{M_1 = M \cup \{a,v\} \\ v \in V_n \setminus (M \cap V_n)}} x_{M_1}.$$

□

This leads to a similar congruence using all containing matchings of a larger size:

Lemma 4.3.7. *For any partial matching M of $2d$ vertices and $d \leq k \leq n$, we have*

$$x_M \simeq_{(\mathcal{Q}_n, k)} \frac{1}{\binom{n-d}{k-d}} \sum_{\substack{M' \supset M \\ |M'| = k}} x_{M'} \tag{4.3.2}$$

Proof. We use induction on $k-d$. The start of the induction is when $k=d$, when the sides of Equation (4.3.2) are equal.

If $k>d$, let $a \in U_n$ be a fixed vertex not covered by M. Applying Lemma 4.3.6 to M and a followed by the inductive hypothesis gives:

$$x_M \simeq_{(\mathcal{Q}_n,d+1)} \sum_{\substack{M_1=M\cup\{a,u\}\\ u\in V_n\setminus(M\cap V_n)}} x_{M_1} \simeq_{(\mathcal{Q}_n,k)} \frac{1}{\binom{n-d-1}{k-d-1}} \sum_{\substack{M'\supset M_1\\ |M'|=k\\ M_1=M\cup\{a,u\}\\ u\in V_n\setminus(M\cap V_n)}} x_{M'}.$$

Averaging over all vertices $a \in U_n$ not covered by M, we obtain

$$\begin{aligned} x_M &\simeq_{(\mathcal{Q}_n,k)} \frac{1}{n-d}\frac{1}{\binom{n-d-1}{k-d-1}} \sum_{\substack{M'\supset M_1\\ |M'|=k\\ M_1=M\cup\{a,u\}\\ a\in U_n\setminus(M\cap U_n)\\ u\in V_n\setminus(M\cap V_n)}} x_{M'} \\ &= \frac{1}{n-d}\frac{1}{\binom{n-d-1}{k-d-1}}(k-d) \sum_{\substack{M'\supset M\\ |M'|=k}} x_{M'} \\ &= \frac{1}{\binom{n-d}{k-d}} \sum_{\substack{M'\supset M\\ |M'|=k}} x_{M'}. \end{aligned}$$

□

Corollary 4.3.8. *For any polynomial F, there is a constant c_F with $\sum_{\sigma\in S_n} \sigma F \simeq_{(\mathcal{Q}_n,\deg F)} c_F$.*

Proof. In view of Lemma 4.3.5, it is enough to prove the claim for $F = x_M$ for some partial matching M on $2k$ vertices, which is an easy application of Lemma 4.3.7 with $d=0$:

$$\sum_{\sigma\in S_n} \sigma x_M = (n-k)! \sum_{M'\colon |M'|=k} x_{M'} \simeq_k (n-k)!\binom{n}{k}.$$

□

The next lemma will allow us to apply induction:

Lemma 4.3.9. *If L is a polynomial with $L \simeq_{(\mathcal{Q}_{n-2},d)} 0$ and a,b are the additional vertices in $\mathcal{Q}_n$ then $Lx_{ab}x_{ba} \simeq_{(\mathcal{Q}_n,d+2)} 0$.*

Proof. It is enough to prove the claim when L is from $\mathcal{Q}_{n-2}$. For $L = x_e^2 - x_e$, $L = x_{uv}x_{uw}$, and $L = x_{uv}x_{wv}$ the claim is trivial, as then $L \in \mathcal{Q}_n$. The remaining cases are

1. $L = \sum_{u\in U_{n-2}} x_{uv} - 1$ for some $v \in V_{n-2}$

2. $L = \sum_{v \in V_{n-2}} x_{uv} - 1$ for some $u \in U_{n-2}$

. We only deal with the first case, as the second one is analogous. Then

$$Lx_{ab}x_{ba} = \left(\sum_{u \in U_n} x_{uv} - 1 \right) x_{ab}x_{ba} - x_{av}x_{ab}x_{ba} - x_{bv}x_{ab}x_{ba} \simeq_{(\mathcal{Q}_n, d+1)} 0.$$

□

We are now ready to prove Theorem 4.3.4.

Proof of Theorem 4.3.4. We use induction on the degree d of F. The case $d = 0$ is obvious, as then clearly $F = 0$. (Note that $\simeq_{-1}$ is just equality.) The case $d = 1$ rephrased means that the affine space spanned by the characteristic vectors of all perfect matchings is defined by the $\sum_v x_{uv} - 1$ for all vertices u. This follows again from Edmonds's description of the perfect matching polytope by linear inequalities in [Edm65] (valid for any graph in addition to K_{2n} and $K_{n,n}$).

For the case $d \geq 2$ we first prove the following claim:

Claim. If $F \in \langle \mathcal{Q}_n \rangle_I$ is a degree d polynomial and $\sigma \in S_n$ is a permutation of vertices, then

$$F \simeq_{(\mathcal{Q}_n, 2d-1)} \sigma F.$$

We use induction on the degree. If $d = 0$ or $d = 1$ the claim follows from the corresponding cases $d = 0$ and $d = 1$ of the theorem. For $d \geq 2$ it is enough to prove the claim when σ is a transposition of two vertices a and u. Note that in $F - \sigma F$ all monomials which do not contain an x_e with e incident to a or u on the left cancel:

$$F - \sigma F = \sum_{e\colon e=(a,r) \textbf{ or } e=(u,r)} L_e x_e \tag{4.3.3}$$

where each L_e has degree at most $d - 1$. We now show that every summand is congruent to a sum of monomials containing edges incident to both a and u on the left. For example, for $e = \{a, b\}$ in (4.3.3), we apply the generator $\sum_v x_{uv} - 1$ to find:

$$L_{ab}x_{ab} \simeq_{d+1} L_{ab}x_{ab} \sum_v x_{uv} \simeq_{d+1} \sum_v L_{ab}x_{ab}x_{uv}.$$

Therefore

$$F - \sigma F \simeq_{d+1} \sum_{bv} L'_{bv} x_{ab} x_{uv}$$

for some polynomials L'_{bv} of degree at most $d-1$. We may assume that L'_{bv} does not contain variables x_e with e incident to a, u on the left or b, v on the right, as these can be removed using generators like $x_{ab}x_{ac}$ or $x_{ab}^2 - x_{ab}$. Moreover, since F is zero on all perfect matchings, it can be checked that L'_{bv} is zero on all perfect matchings containing $\{a, b\}$ and $\{u, v\}$. By induction, $L'_{bv} \simeq_{(\mathcal{Q}_{n-4}, 2d-3)} 0$ (identifying K_{n-4} with the graph $K_n \setminus \{a, b, u, v\}$), from which $L'_{bv} \simeq_{(\mathcal{Q}_n, 2d-1)} 0$ follows by two applications of Lemma 4.3.9. (The special case $a = v, b = u$

is also handled by induction and one application of Lemma 4.3.9.) This concludes the proof of the claim.

We now apply the claim followed by Corollary 4.3.8:

$$F \simeq_{2d-1} \frac{1}{n!} \sum_{\sigma \in S_n} \sigma F \simeq_d \frac{c_F}{n!}$$

for a constant c_F. As $F \in \langle \mathcal{Q}_n \rangle_I$, it must be that $c_F = 0$, and therefore $F \simeq_{2d-1} 0$. □

Bibliography

[AOW15] Sarah R. Allen, Ryan O'Donnell, and David Witmer, *How to refute a random CSP*, CoRR **abs/1505.04383** (2015).

[BDP13] Jop Briët, Daniel Dadush, and Sebastian Pokutta, *On the existence of 0/1 polytopes with high semidefinite extension complexity*, Proceedings of ESA / arXiv:1305.3268 (2013), 217–228.

[BDP15] ______, *On the existence of 0/1 polytopes with high semidefinite extension complexity*, To appear in Mathematical Programming B (2015).

[BFPS12] Gábor Braun, Samuel Fiorini, Sebastian Pokutta, and David Steurer, *Approximation limits of linear programs (beyond hierarchies)*, FOCS, 2012, pp. 480–489.

[BGIP99] Sam Buss, Dima Grigoriev, Russell Impagliazzo, and Toniann Pitassi, *Linear gaps between degrees for the polynomial calculus modulo distinct primes*, Proceedings of the thirty-first annual ACM symposium on Theory of computing, ACM, 1999, pp. 547–556.

[BP11] Gábor Braun and Sebastian Pokutta, *An algebraic take on symmetric extended formulations*, Manuscript, 2011.

[BP15] Gábor Braun and Sebastian Pokutta, *The matching polytope does not admit fully-polynomial size relaxation schemes*, Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015 (Piotr Indyk, ed.), SIAM, 2015, pp. 837–846.

[BPZ15] Gábor Braun, Sebastian Pokutta, and Daniel Zink, *Inapproximability of combinatorial problems via small LPs and SDPs*, 2015.

[CLRS13] Siu On Chan, James R. Lee, Prasad Raghavendra, and David Steurer, *Approximate constraint satisfaction requires large LP relaxations*, Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on, IEEE, 2013, pp. 350–359.

[DM96] John D. Dixon and Brian Mortimer, *Permutation groups*, Springer Verlag, 1996.

[Edm65] Jack Edmonds, *Maximum matching and a polyhedron with* $0,1$*-vertices*, J. Res. Nat. Bur. Standards Sect. B **69B** (1965), 125–130. MR 0183532 (32 #1012)

[FKPT12] S. Fiorini, V. Kaibel, K. Pashkovich, and D. O. Theis, *Combinatorial bounds on nonnegative rank and extended formulations*, 2012, arXiv:1111.0444v2.

[FMP$^+$12] Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf, *Linear vs. semidefinite extended formulations: exponential separation and strong lower bounds*, STOC, 2012, pp. 95–106.

[Goe09] Michel X. Goemans, *Smallest compact formulation for the permutahedron*, Manuscript, 2009.

[GPT11] João Gouveia, Parrilo A. Parrilo, and Rekha R. Thomas, *Lifts of convex sets and cone factorizations*, Math. Oper. Res. **38** (2011), no. 2, 248–264.

[Gri01] Dima Grigoriev, *Linear lower bound on degrees of positivstellensatz calculus proofs for the parity*, Theoretical Computer Science **259** (2001), no. 1, 613–622.

[GW95] Michel X. Goemans and David P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. Assoc. Comput. Mach. **42** (1995), 1115–1145.

[KMR17] Pravesh K. Kothari, Raghu Meka, and Prasad Raghavendra, *Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of csps*, Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017, 2017, pp. 590–603.

[KPT10] Volker Kaibel, Kanstantsin Pashkovich, and Dirk Oliver Theis, *Symmetry matters for the sizes of extended formulations*, Proc. IPCO 2010, 2010, pp. 135–148.

[LRS15] James R. Lee, Prasad Raghavendra, and David Steurer, *Lower bounds on the size of semidefinite programming relaxations*, Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015 (Rocco A. Servedio and Ronitt Rubinfeld, eds.), ACM, 2015, pp. 567–576.

[LRST14] James R. Lee, Prasad Raghavendra, David Steurer, and Ning Tan, *On the power of symmetric LP and SDP relaxations*, Proceedings of the 2014 IEEE 29th Conference on Computational Complexity, IEEE Computer Society, 2014, pp. 13–21.

[O'D14] Ryan O'Donnell, *Analysis of boolean functions*, Cambridge University Press, 2014.

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

[Pas14] Kanstantsin Pashkovich, *Tight lower bounds on the sizes of symmetric extensions of permutahedra and similar results*, Math. Oper. Res. **39** (2014), no. 4, 1330–1339.

[Rag08] Prasad Raghavendra, *Optimal algorithms and inapproximability results for every CSP? [extended abstract]*, STOC'08, ACM, New York, 2008, pp. 245–254. MR 2582901

[Rot14] Thomas Rothvoß, *The matching polytope has exponential extension complexity*, Proceedings of STOC (2014), 263–272.

[RRS16] Prasad Raghavendra, Satish Rao, and Tselil Schramm, *Strongly refuting random csps below the spectral threshold*, CoRR **abs/1605.00058** (2016).

[Sch08] G. Schoenebeck, *Linear level Lasserre lower bounds for certain k-CSPs*, Proc. FOCS, IEEE, 2008, pp. 593–602.

[VB96] Lieven Vandenberghe and Stephen Boyd, *Semidefinite programming*, SIAM Rev. **38** (1996), 49–95.

[Yan88] Mihalis Yannakakis, *Expressing combinatorial optimization problems by linear programs (extended abstract)*, Proc. STOC, 1988, pp. 223–228.

[Yan91] ______, *Expressing combinatorial optimization problems by linear programs*, J. Comput. System Sci. **43** (1991), no. 3, 441–466. MR 1135472 (93a:90054)