ON HIERARCHIES OF STRONG SDP RELAXATIONS FOR COMBINATORIAL OPTIMIZATION PROBLEMS

by

Siavosh Benabbas

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Computer Science
University of Toronto

© Copyright 2012 by Siavosh Benabbas
Abstract

On hierarchies of strong SDP relaxations for combinatorial optimization problems

Siavosh Benabbas
Doctor of Philosophy
Graduate Department of Computer Science
University of Toronto
2012

Studying the approximation threshold of NP-hard optimization problems, i.e. the ratio of the objective value achievable by a polynomial time algorithm to that of the optimal solution is an important field in theoretical computer science. In the past two decades there has been significant development both in terms of finding good approximation algorithms, e.g. through the use of semidefinite programming and hardness of approximation, e.g. the development of Probabilistically Checkable Proofs and Unique Games Conjecture.

Trying to prove lower bounds for the approximation threshold of an optimization problem, one could take one of two approaches. In the first approach, one proves such lower bounds under a complexity assumption like \( P \neq NP \) or Unique Games Conjecture. In the second approach, one studies the behaviour of prominent algorithmic schemes, such as Linear Programming (LP) and Semidefinite Programming (SDP) relaxations for the problem of interest. There, the measure of efficiency is the integrality gap which sets the approximation limitation of the algorithms based on these relaxations.

In this work we study the integrality gap of families of strong LP and SDP relaxations for a number of combinatorial optimization problems. The relaxations come from the context of Lift-and-Project systems. There one starts from a standard (weak) relaxation from a problem and iteratively adds more constraints to make the relaxation stronger, i.e. closer to an exact formulation of the problem. We mostly study the performance of the Sherali-Adams SDP relaxations. Specifically, the main contributions of this thesis are as follows.

- We show optimal integrality gaps for the level-\( \Theta(n) \) Sherali-Adams SDP relaxation of MAX \( \kappa \)-CSP\( _q(P) \) for any predicate \( P : [q]^k \to \{0,1\} \) satisfying a technical condition, which we call being promising. Our result show that for such predicates MAX \( \kappa \)-CSP\( _q(P) \)
cannot be approximated (beyond a trivial approximation) by the Sherali-Adams SDP relaxations of even very high level. Austrin and Håstad (SIAM J. Comput., 2011) show that a random predicate is almost surely promising.

- We complement the above result by showing that for some class of predicates which are not promising MAX \( k \)-CSP \( q(P) \) can be approximated (beyond the trivial approximation) by its canonical SDP relaxation.

- We show optimal integrality gap lower bounds for level-poly(\( n \)) Sherali-Adams SDP relaxations of Quadratic Programming. We also present superconstant integrality gap lower bounds for superconstant levels of the same hierarchy for MAX CUT GAIN.

- We show optimal integrality gap lower bounds for the level-5 Sherali-Adams SDP relaxation of Vertex Cover. We also conjecture a positivity condition on the Taylor expansion of a certain function which, if proved, shows optimal integrality gaps for any constant level of the Sherali-Adams SDP hierarchy for Vertex Cover.

- We revisit the connection between integrality gap lower bounds and the Frankl-Rödl theorem (Trans. of the AMS, 1987). We prove a new density version of that theorem which can be interpreted as a new isoperimetric inequality of the Hamming cube. Using this inequality we prove integrality gap lower bounds for the Lovász-Schrijver SDP (resp. Sherali-Adams LP) relaxation of Vertex Cover (resp. Independent Set) in degree bounded graphs.
Dedication

Dedicated to the memory of Avner Magen.

Acknowledgements

I am grateful for the academic guidance received first from Avner Magen and then Toni Pitassi; for many interesting discussions with Per Austrin, Yuval Filmus, Mohammad Taghi Hajiaghayi, Gil Kalai, Nati Linial and all my coauthors; for the many helpful comments of members of my PhD committee and anonymous reviewers of my papers; for the gift of great friendships with Costis Georgiou, Paul Madvedev, Mohammad Moharrami, Joel Oren, Giovanna Thron, and Yevgeniy Vahlis among others; and for the moral support I received from my parents Akbar and Nayerh, and sister Roshanak. But above all else I am grateful to Ioana.
## Contents

1 Introduction ........................................ 1
   1.1 Contributions .................................... 3

2 Preliminaries and Notation ........................... 6
   2.1 Combinatorial Optimization Problems ........... 7
       2.1.1 Vertex Cover and Independent Set .......... 8
       2.1.2 MAX CUT, MAX CUT GAIN, and QUADRATIC PROGRAMMING ... 8
       2.1.3 Constraint Satisfaction Problems .......... 9
       2.1.4 Pairwise Independence and Approximation Resistant Predicates .. 10
   2.2 Linear and Semidefinite Relaxations .................. 11
   2.3 Lift and Project Systems .......................... 15
       2.3.1 A formal definition of the Sherali-Adams hierarchy .... 19
       2.3.2 An alternative definition of the Sherali-Adams Hierarchy ... 21
       2.3.3 Previous work on Lift-and-Project hierarchies ........ 22
   2.4 The Hamming cube and Fourier Representation ......... 23
   2.5 Borsuk and Frankl-Rödl Graphs ..................... 24

3 SDP Gaps from Pairwise Independence ................. 26
   3.1 Expanding CSP Instances .......................... 28
   3.2 Towards Defining Consistent Distributions .......... 29
       3.2.1 Finding Advice-Sets ........................ 30
       3.2.2 Defining the Distributions $\mathcal{P}_\mu(S)$ .......... 31
   3.3 Constructing the Integrality Gap ................... 35
   3.4 Proof of Lemma 3.7 ................................ 37

4 On Quadratic Threshold CSPs .......................... 40
   4.1 $(\epsilon, \eta)$-hyperplane rounding .................. 43
   4.2 Signs of Symmetric Quadratic Polynomials .......... 49
   4.3 Monarchy ....................................... 51
   4.4 Proof of Lemma 4.7 ................................ 58
List of Figures

2.1 Formulations of VERTEX COVER ............................................ 12
2.2 Canonical relaxations of VERTEX COVER .............................. 13
2.3 Standard SDP relaxations of MAX k-CSP(P) and MAX k-CSP_q(P) .. 16

5.1 Canonical formulation and relaxation of QUADRATIC PROGRAMMING ...... 62

7.1 Comparison of $\chi_U \cdot S_d \chi_U$ and $\chi_U \cdot T_\rho \chi_U$ for $n = 100$, $\rho = 1 - 2d/n$, $d = 6, 7$ for sets, $U$, of size 0 to $n$. The horizontal axis is $|U|$ and the vertical is $\chi_U \cdot S_d \chi_U$ and $\chi_U \cdot T_\rho \chi_U$. ................................................................. 103
# List of Algorithms

1. A 2-approximation algorithm for VERTEX COVER .............................................. 13
2. Algorithm Advice ........................................................................................................ 30
3. $(\epsilon, \eta)$-Hyperplane Rounding ........................................................................ 44
4. Rounding SDP solutions for Monarchy ..................................................................... 52
5. Local-Global Experiment ............................................................................................ 77
Chapter 1

Introduction

Consider an NP-hard optimization problem like Vertex Cover. In this problem one is given a graph $G$ and the goal is to find the smallest subset of the vertices that touches all the edges of $G$, i.e. a vertex cover. The size of the smallest vertex cover is usually denoted by $VC(G)$. It is well known ([Kar72]) that given a graph $G$ finding the smallest vertex cover of $G$, or even the value $VC(G)$, is NP-hard; thus assuming that $P \neq NP$ no polynomial time algorithm can find the smallest vertex cover in a given graph $G$. It is then natural, to ask what is the minimum size of a vertex cover a polynomial time algorithm can find compared to $VC(G)$. In particular, the approximation ratio of an algorithm for Vertex Cover is defined as the (maximum over all instances $G$ of the) size of the vertex cover found by algorithm divided by $VC(G)$ and the approximability threshold of Vertex Cover is defined as the infimum of the approximation ratio of all polynomial time algorithms for vertex cover. Finding the approximability threshold of various combinatorial optimization problems has been a very active area of theoretical computer science; see e.g. [H˚ as99, H˚ as01, ARV09, DS05, H˚ as05b, Chl07, Rag08].

In the case of Vertex Cover it is not hard to see that if $M$ is a matching in $G$ any vertex cover of $G$ must include at least one end point of every edge in $M$, so $VC(G) \geq |M|$. On the other hand if $M$ is a maximal matching the set of vertices incidents to $M$ makes a vertex cover of size $2|M|$. So finding a maximal matching in $G$ and taking both endpoints of its edges results is a polynomial time algorithm for Vertex Cover with approximation ratio 2, i.e. a 2-approximation algorithm. As a consequence the approximability threshold of vertex cover is at most 2. On the other hand, using the machinery of Probabilistically Checkable Proofs [FGL+96, AS98, ALM+98], Dinur and Safra [DS05] show that unless $P = NP$ any polynomial time algorithm for Vertex Cover has approximation ratio at least 1.36, i.e. the approximability threshold of Vertex Cover is at least 1.36.

Unfortunately, for Vertex Cover (and indeed many other combinatorial optimization problems) the gap between the lower bound and the upper bound on the approximability threshold is not closed. The best negative result in this case is still the 1.36 bound of [DS05] while the best algorithm is that of Karakostas [Kar05] which achieves an approximation ratio of
Chapter 1. Introduction

2 − O(√1/\log n) where \( n \) is the number of vertices in the graph. The correct approximability threshold for Vertex Cover is believed to be 2 − o(1).

Trying to resolve the approximability of Vertex Cover and in fact any optimization problem, one could take one of two routes. First, one could prove that the approximability threshold is 2 − o(1) assuming a conjecture stronger than \( \text{P} \neq \text{NP} \). In particular, Khot and Regev [KR08] show that assuming Khot’s Unique Game Conjecture (UGC) [Kho02] the approximability threshold for Vertex Cover is 2 − o(1). In fact if one is willing to assume UGC the problem of determining the precise approximability threshold of many combinatorial optimization problems would be resolved; see e.g. [KKMO07, KR08, Aus07, GHM+11]. But the validity of UGC is the subject of an active area of research; see e.g. [ABS10, Kho10, BRS11].

The second approach is to study the behaviour of prominent algorithmic schemes, Linear Programming (LP) and Semidefinite Programming (SDP) in particular, which have yielded state-of-the-art algorithms for many combinatorial optimization problems. There, the measure of efficiency is the integrality gap which sets the approximation limitation of the algorithms based on these relaxations.

There seems to be a connection between integrality gaps for strong LP/SDP relaxations of a problem and its hardness of approximation. In one direction the reductions used to establish hardness of approximation for many problems have been used to prove integrality gaps for them, e.g. [KV05, CMM09, RS09b, Tul09]. In the other direction for the Vertex Cover problem, which is the subject of Chapter 6, Vishwanathan [Vis09] shows that any hard instance of the problem should have subgraphs that look like the so called “Borsuk graphs”. Interestingly a specific subfamily of Borsuk graphs were previously used in many integrality gap instances for Vertex Cover, e.g. [GK98, Cha02, HMM07, GMT08, GMT09a].

Tight integrality gaps of the standard LP and SDP relaxations for many combinatorial optimization problems have long been resolved but celebrated algorithms for a number of combinatorial problems require strengthening the standard relaxation, i.e. adding extra constraints to it to drop its integrality gap. In fact a number of systematic procedures, known as Lift-and-Project systems have been proposed to systematically improve the integrality gap of a relaxation. These systems build strong hierarchies of either LP relaxations (e.g. theLovász-Schrijver and the Sherali-Adams systems) or SDP relaxations (e.g. the Lovász-Schrijver SDP, the Sherali-Adams SDP and the Lasserre systems.) Lift-and-Project systems, can be thought of as being applied in rounds (also called levels.) The bigger the number of rounds used, the more accurate the obtained relaxation is. In fact, for high enough rounds of any of these systems, the final relaxation is exact and no integrality gap exists. On the other hand the size of the derived relaxation grows exponentially with the number of rounds, which implies that the time one needs to solve it also grows. It is then natural to ask whether looking at a modest number of rounds (say \( O(1) \) or \( \log n \) ) will (significantly) improve the approximation factor. A positive answer to such a question then results in an algorithm with a reasonable (polynomial or quasi-polynomial) runtime that beats known approximation algorithms for the problem. A negative
answer however rules out a rich family strengthened (or non-standard) relaxations of the problem at hand as useful for devising (better than state of the art) approximation algorithms for the problem.

Identifying the limitations of relaxations derived by Lift-and-Project system has attracted much attention and showing integrality gaps for the Sherali-Adams SDP and the Lasserre systems stand as the most attractive subjects in this area of research due to a number of reasons. Firstly, the best algorithms known for many combinatorial optimization problems (including the ones studied in Chapters 3, 5, 4 and 6) are based on relaxations weaker than those derived by a constant (say four) rounds of the Sherali-Adams SDP system, see e.g. [GW95, KZ97, ARV09, Kar05]. Lift-and-Project hierarchies have been also used in designing approximation algorithms with a runtime-approximation ratio trade off, e.g. [dlVKM07, MM09, CS08, BCG09, KMN11]. Finally, for constraint satisfaction problems\footnote{Here by constraint satisfaction problems we mean the MAX $\kappa$-CSP$_q(P)$ problems defined in Subsection 2.1.3}, and modulo the Unique Games Conjecture, no approximation algorithm can perform better than the one obtained by Sherali-Adams SDP of a constant number of rounds; see [Rag08]. One can then think of algorithms based on the Sherali-Adams SDP as an interesting model of computation.

\section{1.1 Contributions}

In this work we study the strength of some of these LP and SDP relaxations for various problems. Most of our results are about the Sherali-Adams SDP hierarchy. The performance of this hierarchy of SDPs has been studied for other combinatorial problems; see e.g. [RS09b]. For some problems, we show that these strong relaxations have “large” integrality gaps, often as bad as (best) known approximation algorithms; in other words we show that they cannot be used to improve upon already known results. Such an integrality gap rules out a rich and important family of approximation algorithms for the problem at hand. For others we show that the relaxations have “small” integrality gaps and, in fact, present approximation algorithms based on them that outperform previously known algorithms.

Specifically, the main contributions of this thesis are as follows. We refer the reader to Chapter 2 for the precise definition of the hierarchies and the optimization problems studied.

- In Chapter 3 (joint work with K. Georgiou, A. Magen and M. Tulsiani [BGMT12]) we show that for any predicate $P$ satisfying a certain technical condition, the Sherali-Adams SDP hierarchy of level $\Omega(n)$ for the constraint satisfaction problem MAX $\kappa$-CSP$_q(P)$ has very large integrality gap. In particular, the relaxations derived from this hierarchy cannot even improve on the approximation algorithm which assigns an independent, uniformly random value to each of the variables of the constraint satisfaction problem. See Theorem 3.1.

Austrin and Håstad [AH11] show that almost all predicates satisfy this condition. For the same constraints Austrin and Mossel [AM09] show that assuming UGC no polynomial
time algorithm can improve on the one assigning random values to all the variables. Also, as discussed in Remark 3.2, weaker integrality gaps for the same problem follow from [RS09b] and [CMM09].

- In Chapter 4 (joint work with P. Austrin and A. Magen [ABM10]) we complement the above results by showing that for certain predicates $P$ that do not satisfy the aforementioned technical condition the Sherali-Adams SDP hierarchy of level $\Theta(1)$ can be used to approximate MAX CSP($P$) with approximation factor strictly better than what a random assignment achieves. The predicates studied are “symmetric quadratic thresholds” and the “Monarchy” predicate. See Theorems 4.2 and 4.4.

We also conjecture (see Conjecture 4.3) that the same holds for any (possibly non-symmetric) “linear thresholds”. This conjecture, if true, would be a weak converse to the results of Chapter 3 and [AM09].

- In Chapter 5 (joint work with A. Magen [BM10]) we show that under certain conditions one can extend integrality gap constructions for the standard SDP relaxation of a problem to integrality gaps for level-$k$ Sherali-Adams SDP system. In particular, we show (asymptotically) optimal integrality gaps for level-poly$(n)$ Sherali-Adams SDP relaxation of the QUADRATIC PROGRAMMING problem and super constant integrality gaps for super constant levels of the same hierarchy for the MAX CUT GAIN problem. See Theorems 5.1 and 5.2.

Our results should be compared to those of the chain of papers [AMMN06, ABK+05, KO06]. In particular our results are based on those of Khot and O’Donnell who prove the same integrality gaps but for the much weaker standard SDP relaxation of these problems.

- In Chapter 6 (joint work with S. O. Chan, K. Georgiou and A. Magen [BCGM11]) we show that the level-5 Sherali-Adams SDP relaxation for VERTEX COVER has integrality gap $2 - o(1)$; see Theorem 6.1. We also show that under a conjecture about the Taylor expansion of a certain simple function, the result extends to level-$l$ of the same hierarchy for any constant $l$; see Conjecture 6.18 and Theorem 6.2.

There is a wealth of previous integrality gap results for VERTEX COVER for various hierarchies and relaxations. We give a full list of these results in Chapter 6 on page 73.

Our results are the first non-trivial integrality gap constructions for the Sherali-Adams SDP relaxation for VERTEX COVER.

We also complement a result of Vishwanathan [Vis09] by showing that any so called “Borsuk graph” (i.e. any graph with very low vector chromatic number) is a good integrality gap instance for the level-$\Omega(\sqrt{\log n / \log \log n})$ Sherali-Adams LP system for VERTEX COVER.\footnote{To be precise, we show that the level-$\Omega(\sqrt{\log n / \log \log n})$ Sherali-Adams LP relaxation for VERTEX COVER for such graphs has very low objective value. The graph is only a good integrality gap instance if it also happens} We note that our results are not the first integrality gap constructions for the
Sherali-Adams LP system for Vertex Cover (see [CMM09]) but our proof is much more intuitive than what was known before.

- In Chapter 7 (joint work with H. Hatami and A. Magen [BHM11]) we study a certain density variant of a theorem of Frankl and Rödl [FR87]; see Theorem 7.1. We note that [FR87] and [MOR+06] each present theorems similar (but incomparable) to Theorem 7.1. We refer the reader to Chapter 7 on page 97 for a full comparison of these results to ours.

The Frankl-Rödl theorem is used in the proof of many integrality gap constructions including those in Chapter 6 and our density variant comes up naturally when one tries to extend known integrality gap constructions to the Vertex Cover problem in bounded degree graphs. In fact, we use our theorem to show lower bounds for the integrality gap of the level-$l$ Lovász-Schrijver SDP relaxation of Vertex Cover in bounded degree graphs for any constant $l$. We also show lower bounds for the integrality gap of the level-$l$ Sherali-Adams LP relaxation of Independent Set in bounded degree graphs for any constant $l$; see Theorems 7.3 and 7.4.

Although some integrality gap for Vertex Cover in degree bounded graphs is implicit in [CMM09], the trade-off between the integrality gap, the level of the hierarchy it applies to and the maximum degree are not explored and thus their result is incomparable to ours. To the best of our knowledge there are no other integrality gap lower bounds for Vertex Cover and Independent Set in degree bounded graphs. However, if one assumes UGC, Austrin et al. [AKS11] show that Vertex Cover and Independent Set cannot be approximated better than certain thresholds depending on the maximum degree. Their threshold for Vertex Cover is close to our integrality gap lower bound while for Independent Set it is worse by a factor of $\log d$ where $d$ is the maximum degree.

to only have very big vertex covers.
Chapter 2

Preliminaries and Notation

In what follows \( E \) stands for expectation. For any positive integer \( n \) we use the notation \([n]\) for the set \( \{1, \ldots, n\} \). For a set \( A \), often a subset of \([n]\), we will use \( \varphi(A), \binom{A}{k} \) and \( [n] \leq k \) for the set of all subsets of \( A \), those of size exactly \( k \) and those of size at most \( k \) respectively. We use \((n)_k = n!/(n-k)!\) as a shorthand for the \( k \)th falling power of \( n \).

When \( \mu \) is a distribution on some finite set \( D \) we sometimes abuse notation by writing \( \mu \) for the probability mass function, i.e.,

\[
\mu(x) \overset{\text{def}}{=} \Pr_{y \sim \mu}[y = x].
\]

Note that this allows us to think of distributions as real functions with domain \( D \). For a finite set \( S \) (often a subset of \([n]\)) we use the notation \( \{-1,1\}^S \) for the set of all \(-1,1\) vectors indexed by elements of \( S \). When \( x \in \{-1,1\}^S \) and \( y \in \{-1,1\}^{S'} \) are two vectors indexed by disjoint sets, i.e. \( S \cap S' = \emptyset \), we use \( x \circ y \in \{-1,1\}^{S \cup S'} \) to denote their natural concatenation. Conversely, for \( x \in \{-1,1\}^S \), we denote its projection to \( S' \subseteq S \) as \( x(S') \). We define \( [q]^S \) similarly to \( \{-1,1\}^S \) and use the same notation for members of that set. In context of variables with domain \([q]\) members of \( [q]^S \) can be understood as partial assignments to a given subset of variables; in this setting we often use greek letters for members of \([q]^S\).

For a distribution \( \mu \) on \( \{-1,1\}^S \) and for \( S' \subseteq S \) we let \( \text{Mar}_{S'} \mu \) be the marginal distribution of \( \mu \) on the set \( S' \). In particular, for \( y \in \{-1,1\}^{S'} \) we have

\[
(\text{Mar}_{S'} \mu)(x) = \Pr_{y \sim \mu}[\forall i \in S' \ x_i = y_i] = \sum_{x \in \{-1,1\}^{S\setminus S'}} \mu(x \circ z). \tag{2.1}
\]

The second equation lets us extend the definition to any function \( f : \{-1,1\}^S \to \mathbb{R} \).

We use \( \varphi \) and \( \Phi \) for the probability density function and the cumulative distribution function of a standard normal random variable, respectively, i.e.

\[
\varphi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \quad \Phi(t) = \int_{x=-\infty}^t \varphi(x)dx \tag{2.2}
\]
We use the notation $S^n$ for the $n$-sphere of radius 1, i.e. the set of unit vectors in $\mathbb{R}^{n+1}$. We will use bold letters for real vectors, and $\mathbf{v} \cdot \mathbf{u}$ for the inner product of two vectors $\mathbf{v}, \mathbf{u} \in \mathbb{R}^d$.

**Definition 2.1.** For $n \geq 1$, $\mathbf{v} \in S^n$ and $0 \leq h \leq 2$ the spherical cap of height $h$ centered at $\mathbf{v}$ is defined as the following subset of $S^n$,

$$\{ \mathbf{w} \in S^n : \mathbf{v} \cdot \mathbf{w} \geq 1 - h \},$$

or alternatively,

$$\{ \mathbf{w} \in S^n : \| \mathbf{v} - \mathbf{w} \|_2 = \sqrt{2 - 2 \mathbf{v} \cdot \mathbf{w}} \leq \sqrt{2h} \}.$$

A random spherical cap of height $h$ is a spherical cap of height $h$ centered around a uniformly random point in $S^n$.

When $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}$ are $n$ vectors, their *Gram matrix* is the $n \times n$ matrix $A = [a_{ij}]$ defined as, $a_{ij} = \mathbf{v}_i \cdot \mathbf{v}_j$. For a symmetric matrix $A_{n \times n}$ we use $A \succeq 0$ to indicate that $A$ is *positive semidefinite*, i.e. has non-negative eigenvalues or equivalently is the gram-matrix of some set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

We use $\text{sign}(x)$ for the sign function defined as,

$$\text{sign}(x) = \begin{cases} 
1 & x > 0, \\
-1 & x \leq 0.
\end{cases} \quad (2.3)$$

Note that $\text{sign}(0) = -1$.

When $G = (V, E)$ is a graph we use $\text{IS}(G)$ to denote the size of the biggest independent set of $G$, i.e. the biggest subset $S \subseteq V$ such that $i, j \in S$ implies $ij \notin E$. Similarly, we define $\text{VC}(G)$ to denote the size of the smallest vertex cover of $G$, i.e. $\text{VC}(G) = |V(G)| - \text{IS}(G)$.

## 2.1 Combinatorial Optimization Problems

In an *optimization problem* one is asked to assign values to a set of variables so as to satisfy a set of conditions and minimize (or maximize) an objective function. When the domain of the variables is discrete the problem is a *combinatorial optimization problem*. The problem is called a *minimization* (respectively *maximization*) if the goal is to minimize (resp. maximize) the objective function. All the problems discussed in this thesis are NP-hard to solve exactly.

The *approximation ratio* of an algorithm for a minimization problem is the supremum (over the possible inputs) of the value of the objective function achieved by the algorithm divided by the minimum possible value for the objective function. Likewise, for a maximization problem the approximation ratio is the supremum (over the possible inputs) of the optimal value of the objective function divided by the value achieved by the algorithm. So the approximation ratio is always at least 1. An algorithm with approximation ratio $\alpha$ is sometimes called an $\alpha$-approximation algorithm. We sometimes abuse this notation slightly by allowing $\alpha$ to be a
function of the size of the instance, e.g. a $2 - O(\sqrt{1/\log n})$-approximation algorithm for VERTEX COVER or a $O(n/\log^2 n)$-approximation algorithm for INDEPENDENT SET.

The approximability threshold of a combinatorial optimization problem is defined as the infimum of the approximation ratio of all polynomial time algorithms for the problem.

In the rest of this section we define the various combinatorial optimization problems studied in this thesis.

### 2.1.1 Vertex Cover and Independent Set

Given a graph $G$ the VERTEX COVER problem asks for the (size of the) minimum vertex cover of $G$, i.e. $VC(G)$ and the INDEPENDENT SET problem asks for the (size of the) maximum independent set of $G$, i.e. $IS(G)$. Notice that a subset of the vertices of $G$ is an independent set if and only if its complement is a vertex cover, so an exact algorithm that solves one of these two problems also solves the other. However, an $\alpha$-approximation algorithm for one is not an $\alpha$-approximation for the other.

As mentioned before the best approximation algorithm known for VERTEX COVER is a $2 - \Omega(1/\sqrt{\log n})$-approximation algorithm due to Karakostas [Kar05]. On the hardness side Dinur and Safra [DS05] prove that unless $P = NP$ VERTEX COVER has no 1.36-approximation algorithm and Khot and Regev [KR08] show that if one assumes UGC it does not have a $(2 - \epsilon)$-approximation algorithm (for any constant $\epsilon > 0$) either. We compliment these results with integrality gaps for the Sherali-Adams SDP hierarchy in Chapter 6.

For INDEPENDENT SET the best approximation algorithm, due to Feige [Fei04], has approximation factor $O(n\log^2 n/\log^3 n)$. On the other hand, Håstad [Hås99] shows that unless $NP \subseteq ZPP$ INDEPENDENT SET has no $n^{1-\epsilon}$-approximation algorithm (for any constant $\epsilon > 0$) and Khot and Ponnuswami [KP06] rule out a $n/2^{(\log n)^{0.75+\epsilon}}$-approximation algorithm under a slightly stronger assumption.

We also study the VERTEX COVER and INDEPENDENT SET problems in degree bounded graphs, i.e. when the input graph is promised to have maximum degree at most $d$. In this context $d$ is assumed to be a large constant, i.e. the approximation factor of algorithms are studied asymptotically as a function of $d$.

### 2.1.2 Max Cut, Max Cut Gain, and Quadratic Programming

In the MAX CUT problem one is given a graph $G = (V, E)$ and is asked to find a partition of the vertices, $V = S \cup \overline{S}$, such that the number of edges going between $S$ and $\overline{S}$ is maximized. In other words, define $E(S, \overline{S})$ as the set of edges going across the cut, i.e.,

$$E(S, \overline{S}) = \{ \{u, v\} \in E : u \in S, v \in \overline{S} \}.$$  \hspace{1cm} (2.4)

Then the value of a cut is $|E(S, \overline{S})|$ and the objective is to find a cut of maximum value.
The Max Cut Gain problem is a variant of Max Cut in which the objective is to maximize the number of edges cut minus the number of edge not cut. In other words, one wants to find a cut \((S, \overline{S})\) that maximizes \(2|E(S, \overline{S})| - |E|\). Notice that an algorithm optimally solves Max Cut Gain if and only if it optimally solves Max Cut but in terms of hardness of approximation the two problems are different. While a uniformly random cut of \(G\) achieves expected value \(1/2\) for Max Cut, its expected objective value for Max Cut Gain is precisely 0. In fact, while the celebrated algorithm of Goemans and Williamson [GW95] has approximation ratio 1.38 for Max Cut, no constant factor approximation algorithm for Max Cut Gain is known; in fact Khot and O’Donnell [KO06] show that (assuming UGC) such an algorithm does not exist.

We note that the weighted version of Max Cut Gain (and Max Cut) is not harder than the unweighted version in terms of hardness of approximation, up to any polynomially small additive factor; see Crescenzi, Silvestri, and Trevisan [CST01].

The Quadratic Programming problem is a generalization of the Max Cut Gain problem introduced by Charikar and Wirth [CW04]. In Quadratic Programming one is given a real matrix \(A_{n \times n}\) and the goal is to find an \(x \in \mathbb{R}^n\) that maximizes \(\sum_{i \neq j} A_{ij} x_i x_j\). One can assume that the matrix \(A\) is symmetric and has zero diagonal. It is easy to see that Max Cut Gain is the special case of Quadratic Programming when the elements of \(A\) are either 0 or \(-1\), or alternatively (given the above mentioned result of [CST01]) non-positive.

### 2.1.3 Constraint Satisfaction Problems

An instance \(\Psi\) of Max \(k\)-CSP\(_q\) is a set of variables \(\{x_1, \ldots, x_n\}\) with domain \([q]\) and the constraints \(C_1, \ldots, C_m\). Each constraint is a function of the form \(C_i: [q]^{T_i} \to \{0, 1\}\) depending only on the values of the variables in the ordered tuple \(T_i\) with \(|T_i| = k\). A solution to \(\Psi\) is a function assigning values from the domain to each variables, i.e. \(\psi: \{x_1, \ldots, x_n\} \to [q]\). The objective value of the solution is the number of \(C_i\)’s that evaluate to 1, i.e.

\[
\sum_{i=1}^m C_i(\psi(T_{i,1}), \psi(T_{i,2}), \ldots, \psi(T_{i,k})).
\]

We will study constraint satisfaction problems where every constraint is specified by the same Boolean predicate \(P: [q]^k \to \{0, 1\}\). We denote the set of assignments for which the predicate evaluates to 1 by \(P^{-1}(1)\). A CSP instance for such a problem is a collection of constraints of the form \(P\) applied to \(k\)-tuples of literals. To define a literal we think of \([q]\) as the \(q\) element cyclic group with operator +; a literal is then defined to be \(x_i + a\) for a variable \(x_i\) and \(a \in [q]\).\(^1\) More formally,

\(^1\)Note that when \(q = 2\) we usually say that each literal is either \(x_i\) or \(\neg x_i\). When \(q \geq 2\), \(x_i + a\) corresponds to a generalization of the concept of negation. Another possible generalization would be \(\pi(x_i)\) where \(\pi\) is allowed to be any permutation of \([q]\). Note that our notion of literals is more restrictive and hence our lower bound result is stronger. For a similar discussion see, e.g. [AM09, AH11].
Definition 2.2. For a given $P : [q]^k \rightarrow \{0, 1\}$, an instance $\Psi$ of $\text{MAX } \kappa\text{-CSP}_q(P)$ is a set of constraints $C_1, \ldots, C_m$ where each constraint $C_i$ is over a $k$-tuple of variables $T_i = (x_{i1}, \ldots, x_{ik})$ and is of the form $P(x_{i1} + a_{i1}, \ldots, x_{ik} + a_{ik})$ for some $a_{i1}, \ldots, a_{ik} \in [q]$. We denote the maximum number of constraints that can be simultaneously satisfied by $\text{OPT}(\Psi)$. The $C_i$ are sometimes called the clauses of $\Psi$.

Definition 2.3. For a given $k \geq 1, q \geq 2$, an instance $\Psi$ of $\text{MAX } \kappa\text{-CSP}_q$ is a set of constraints $C_1, \ldots, C_m$ where each constraint $C_i$ is over a set of $k$ variables $T_i \subseteq \{x_1, \ldots, x_n\}$, $|T_i| = k$ and is of the form $C_i : [q]^{T_i} \rightarrow \{0, 1\}$. We denote the maximum number of constraints that can be simultaneously satisfied by $\text{OPT}(\Psi)$.

When $q = 2$ we use $\text{MAX } \kappa\text{-CSP}_q$ (resp. $\text{MAX } \kappa\text{-CSP}(P)$) for $\text{MAX } \kappa\text{-CSP}_2$ (resp. $\text{MAX } \kappa\text{-CSP}_2(P)$.) When it is more convenient we think of the domain of the variables as $\{-1, 1\}$ instead of $\{0, 1\}$. For a fixed $P : [q]^k \rightarrow \{0, 1\}$ notice that $k$ is just the arity of the predicate and is implicit in $P$, so we sometimes use $\text{MAX } \kappa\text{-CSP}_q(P)$ (or $\text{MAX } \kappa\text{-CSP}(P)$ when $q = 2$) for $\text{MAX } \kappa\text{-CSP}_q(P)$.

Many well known optimization problems can be viewed as $\text{MAX } \kappa\text{-CSP}_q(P)$ for a specific predicate $P$. Examples include $\text{MAX } \kappa\text{-SAT}$, $\text{MAX } \kappa\text{-XOR}$ and $\text{MAX } \kappa\text{-LIN}_q$, which is itself a generalization of $\text{MAX } \text{Cut}$ for $k = q = 2$. Unfortunately, $\text{MAX } \kappa\text{-CSP}_q(P)$ is NP-hard for essentially all choices of $P$. In fact it follows from the dichotomy theorem of Creignou [Cre95] that in the case of $q = 2$, $\text{MAX } \kappa\text{-CSP}(P)$ cannot be approximated within a factor better than $1 + \epsilon$ for some $\epsilon > 0$ unless the predicate $P$ is a function of just one variable.

### 2.1.4 Pairwise Independence and Approximation Resistant Predicates

The arguably simplest approximation algorithm for $\text{MAX } \kappa\text{-CSP}_q(P)$ is to pick a uniformly random assignment to the variables. As this algorithm satisfies each constraint with probability $\frac{[P^{-1}(1)]}{q^k}$, it follows that it has approximation ratio $\frac{q^k}{[P^{-1}(1)]}$. A predicate $P$ is called approximation resistant if this algorithm is the best polynomial time algorithm for $\text{MAX } \kappa\text{-CSP}_q(P)$, i.e. the approximability threshold of $\text{MAX } \kappa\text{-CSP}_q(P)$ is $\frac{q^k}{[P^{-1}(1)]}$; $P$ is called approximable otherwise.

An important line of research is to characterize the approximation resistant predicates. In their classic paper Goemans and Williamson [GW95] used semidefinite programming to obtain improved approximation algorithms for predicates on two variables. For instance, for $\text{MAX } 2\text{-LIN}(2)$ they gave an algorithm with approximation ratio $1/\alpha_{GW} \approx 1.138$. Following [GW95], many new approximation algorithms were found showing various specific predicates are approximable. However, for some cases, perhaps most prominently the $\text{MAX } 3\text{-SAT}$ problem, no such progress was made. Then, in another classic paper Håstad [Hås01] proved that $\text{MAX } 3\text{-SAT}$ is in fact NP-hard to approximate within $8/7 - \epsilon$, i.e. the predicate $P(x_1, x_2, x_3) = x_1 \lor x_2 \lor x_3$ is approximation resistant. For predicates on three variables, the work of Håstad together with work of Zwick [Zwi98] shows that a predicate is approximation resistant if and only if it is implied by an XOR of the three variables, or the negation thereof, where a predicate $P$
is said to imply a predicate $P'$ if for all $x$, $P(x) = 1 \Rightarrow P'(x) = 1$. For predicates over four variables, Hast [Has05a] made an extensive classification leaving open the status of 46 out of 400 different predicates. There have been several papers [ST00, EH05, ST06], mainly motivated by the soundness-query trade-off for PCPs, giving increasingly general conditions under which a predicate is approximation resistant. In a recent paper Austrin and Mossel [AM09] proved a general condition under which a predicate is approximation resistant. We need the following definition to state their result.

We say that a distribution $\mu$ over variables $x_1, \ldots, x_k$, is a balanced pairwise independent distribution over $[q]^k$, if we have

$$\forall j \in [q], \forall i. \Pr_{\mu}[x_i = j] = \frac{1}{q} \quad \text{and} \quad \forall j_1, j_2 \in [q], \forall i_1 \neq i_2. \Pr_{\mu}[(x_{i_1} = j_1) \land (x_{i_2} = j_2)] = \frac{1}{q^2}.$$ 

**Definition 2.4.** A predicate $P : [q]^k \rightarrow \{0, 1\}$ is called promising, if there exist a distribution (on $[q]^k$) supported over a subset of $P^{-1}(1)$ that is pairwise independent and balanced. If $\mu$ is such a distribution we say that $P$ is promising supported by $\mu$.

Assuming the Unique Games Conjecture, Austrin and Mossel [AM09] show that all promising predicates are approximation resistant.

This condition is very general and turned out to give many new cases of approximation resistant predicates; see Austrin and Håstad [AH11]. In Chapter 3 we show that a strong family of algorithms (some of which do not run in polynomial time) does not achieve approximation factor better than $\frac{q^k}{|P^{-1}(1)|}$ for MAX $k$-CSP$_q(P)$ if $P$ is promising. Furthermore, in Chapter 4 we show that a subclass of the predicates which are not promising are not approximation resistant, i.e. we show a weak converse to the results of [AM09].

### 2.2 Linear and Semidefinite Relaxations

It is often easy to reformulate a combinatorial optimization problem as an **integer linear program** (sometimes integer program or IP) or **quadratic program**. An integer program (resp. quadratic program) has three parts:

- a set of variables $x_1, \ldots, x_n$ each with domain the integers (resp. the reals);
- a set of constraints on the variables where each constraint is of the form $p(x_1, x_2, \ldots, x_n) \geq 0$ where $p$ is a linear (resp. quadratic) polynomial;
- and a linear (resp. quadratic) objective function $\sum_i c_i x_i$ (resp. $\sum_{i,j} c_{i,j} x_i x_j$) to be either minimized or maximized over all the possible values of $x_1, \ldots, x_n$ that satisfy all the constraints.

Such programs are often used to formulate optimization problems. For example, Figure 2.1 shows formulations of the VERTEX COVER problem as integer and quadratic programs.
Chapter 2. Preliminaries and Notation

Minimize $\sum_{i=1}^{n} x_i$

Where, $\forall i \in [n]$ $x_i \in \{0, 1\}$

Subject to $\forall ij \in E$ $x_i + x_j \geq 1$

(a) as an integer (linear) program.

Minimize $\sum_{i=1}^{n} x_i$

Where, $\forall i \in [n]$ $x_i \in \mathbb{R}$

Subject to $\forall ij \in E$ $(1 - x_i)(1 - x_j) = 0$

$\forall i \in [n]$ $x_i^2 = x_i$

(b) as a quadratic program.

Figure 2.1: Formulations of Vertex Cover for a graph $G = ([n], E)$

It is not hard to see that solving integer or quadratic programs is \textbf{NP}-hard, in fact this follows from \textbf{NP}-completeness of Vertex Cover and the reductions implicit in Figure 2.1. Although the integer/quadratic programs that formulate Vertex Cover cannot be solved in polynomial time they can be “relaxed” to similar linear/vector programs that can be solved in polynomial time. Indeed these relaxations serve as the starting point of many important approximation algorithms.

A linear program is very much like an integer program except that the domain of the variables is $\mathbb{R}$ in place of $\mathbb{N}$. A vector program should be compared to a quadratic program. There the variables are $v_1, \ldots, v_n$ and the domain of $v_i$ is $\mathbb{R}^n$ and each constraint is of the form $\sum_{ij} a_{ij} v_i \cdot v_j \geq b$ for some $b, a_{11}, \ldots, a_{nn} \in \mathbb{R}$. The objective is to minimize or maximize $\sum_{ij} c_{ij} v_i \cdot v_j$. Notice that because only the inner products of the vectors are important the domain can be taken to be $\mathbb{R}^m$ for any value $m \geq n$; but one cannot insist that $v_i$’s take low-dimensional values. We can alternatively take the variables to be entries of a symmetric, positive semidefinite matrix $X = [x_{ij}]$ with the constraints and objective function being linear polynomials of $x_{ij}$’s. For this reason the names vector program and semidefinite program (SDP) are used interchangeably.

One can relax an integer program into a linear program by simply changing the domain of the variables to be the set of real numbers. Similarly one can relax a quadratic program to a vector program. The first step is to “homogenize” the objective function and the constraints of the program by adding a new variable $x_0$ and changing all the linear and constant terms into quadratic terms by multiplying them by $x_0$ and $x_0^2$; for example the constraint $x_1x_2 + x_2x_3 + x_3 + 2 \geq 0$ would become $x_1x_2 + x_2x_3 + x_0x_3 + 2x_0^2 \geq 0$. The next step is to add the constraint $x_0^2 = 1$ to the program. The last step is to change the domain of all variables to be $\mathbb{R}^{n+1}$ (where $n$ is the number of the variables in the original quadratic program) and to change each product $x_ix_j$ into the inner product of the two variables. We often also rename the variables from $x_i$’s into $v_i$’s to make it clear that they are vectors. For example, the relaxations of the integer and quadratic programs of Vertex Cover seen in Figure 2.1 are the canonical linear programming and SDP relaxations of Vertex Cover seen in Figure 2.2. Both these relaxations have been studied extensively.
In both cases the solutions of the original integer/quadratic programming formulation are also solution for the resulting linear/semidefinite programing relaxation. In the case of semidefinite programming relaxation in particular the solutions of the original quadratic formulation are precisely the one dimensional solutions of the semidefinite relaxation.

Minimize \( \sum_{i=1}^{n} x_i \)

Where, \( \forall i \in [n] \) \( x_i \in [0, 1] \)

Subject to \( \forall ij \in E \) \( x_i + x_j \geq 1 \)

(a) The linear programming relaxation.

Minimize \( \sum_{i=1}^{n} v_0 \cdot v_i \)

Where, \( \forall i \in \{0 \ldots n\} \) \( v_i \in \mathbb{R}^{n+1} \)

Subject to \( \forall ij \in E \) \( (v_0 - v_i) \cdot (v_0 - v_j) = 0 \)

\( v_0 \cdot v_0 = 1 \)

\( \forall i \in [n] \) \( v_i \cdot v_i = v_i \cdot v_0 \)

(b) The vector (semidefinite) programming relaxation.

Figure 2.2: Canonical relaxations of VERTEX COVER for a graph \( G = ([n], E) \)

The best approximation algorithms for many combinatorial optimization problems are based on linear or semidefinite relaxations. Such algorithms often start by solving a relaxation of the problem and then “rounding” the solution of the relaxation into a solution for the problem. As an example consider Algorithm 1 for VERTEX COVER. Lines 2 to 7 are the “rounding” step of the algorithm; this step (here as well as in many other approximation algorithms) does not look at anything except the solution \( x_1, \ldots, x_n \) of the relaxation.

Algorithm 1: A 2-approximation algorithm for VERTEX COVER

Input: A graph \( G = ([n], E) \) with \( n \) vertices.

Output: A vertex cover \( S \subseteq [n] \) of \( G \).

1 \( (x_1, \ldots, x_n) \leftarrow \text{LPSolve}(\text{The LP shown in Figure 2.2(a)}) \)

2 \( S \leftarrow \emptyset \)

3 for \( i = 1, \ldots, n \) do

4 \( \quad \text{if } x_i \geq \frac{1}{2} \text{ then} \)

5 \( \quad \quad S \leftarrow S \cup \{i\} \)

6 \( \quad \text{end} \)

7 \( \text{end} \)

8 return \( S \)

It is not hard to see that Algorithm 1 is a 2-approximation algorithm for VERTEX COVER. On one hand, if \( ij \in E(G) \) then \( x_i + x_j \geq 1 \) so at least one of \( x_i \) and \( x_j \) is more than or equal to 1/2, so one of \( i \) and \( j \) will be in \( S \), i.e. \( S \) is always a vertex cover of \( G \). On the other hand the size of the vertex cover output by the algorithm is at most twice the objective value of the LP in Figure 2.2(a) which in turn is at most the objective value of the IP in Figure 2.1(a), i.e. the size of
the minimum vertex cover of $G$. Notice how the size of the vertex cover found by the algorithm was directly compared to the objective value of the LP relaxation and only through that to the objective value of the IP formulation. Indeed most approximation algorithms based on linear/semidefinite programming are analyzed in such a way, i.e. the objective value achieved by the approximation algorithm is related to the objective value of the relaxation and only through that to the objective value of the original optimization problem. This motivates the following standard definition is central to the study of LP and SDP relaxations of combinatorial optimization problems.

**Definition 2.5.** The integrality gap (IG) of a relaxation for a minimization problem is defined as,

$$ IG = \sup_{\Psi} \frac{\text{opt}(\Psi)}{\text{relax}(\Psi)}, $$

where the supremum is over all instances, $\Psi$, of the problem, $\text{opt}(\Psi)$ is the optimal objective value for the instance and $\text{relax}(\Psi)$ is the value of the linear/semidefinite relaxation. Similarly, for a maximization problem the integrality gap of a relaxation is defined as,

$$ IG = \sup_{\Psi} \frac{\text{relax}(\Psi)}{\text{opt}(\Psi)}. $$

In either case the integrality gap is at least 1.

As an example let us consider the LP relaxation of Vertex Cover from Figure 2.2(a). Notice that the analysis of Algorithm 1 shows that the integrality of this relaxation is at most 2. It is not hard to see that the integrality gap of this relaxation is also at least 2. To see this consider the graph $G = K_n$ the complete graph on $n$ vertices. It is clear that the smallest vertex cover of $G$ has size $n - 1$ on the other hand $x_1 = \cdots = x_n = 1/2$ is a solution to the LP relaxation with objective value $n/2$. Hence, the integrality gap of the relaxation is at least $2 - 2/n$. Taking $n \to \infty$ shows that the integrality gap is at least 2. Notice that this shows that any algorithm based on the linear program which is analyzed by comparing the objective value of its output with the objective value of the LP (as opposed to that of the IP) has approximation factor at least 2. As we mentioned previously most approximation algorithms based on LP/SDP relaxations are analyzed this way so a lower bound on the integrality gap of a relaxation rules out a big family of approximation algorithms.\(^2\)

One important drawback of the concept of integrality gap is that it depends on the relaxation considered. In particular, one could imagine that although the integrality gap of the LP relaxation in Figure 2.2(a) is 2 a very small modification would decrease its integrality gap significantly, or that the integrality gap of the SDP relaxation in Figure 2.2(b) might be significantly less than 2. Fortunately, there are systematic ways of starting from a relatively weak

\(^2\)For the above LP relaxation of Vertex Cover we can show a stronger statement. Notice that $x_1 = \cdots = x_n = 1/2$ is also optimal solution of the relaxation for $G = C_n$, the cycle on $n$ vertices. Furthermore, notice that the size of the minimum vertex cover of $K_n$ and $C_n$ are $n - 1$ and $\lfloor n/2 \rfloor$ respectively. So, any algorithm whose rounding step only uses the LP solution $x_1, \ldots, x_n$ is either wrong for $K_n$ or has approximation ratio at least $2 - 4/n$ for $C_n$, no matter how it is analyzed.
relaxation like the one in Figure 2.2(a) and “strengthening” it to get stronger and stronger relaxations of the same problem (e.g. the one in Figure 2.2(b).) These methods, called Lift-and-Project systems are discussed in Section 2.3. Proving an integrality gap for a relaxation derived in such a way rules out approximation algorithms based on a rich class of LP/SDP relaxations.

We next turn to the standard relaxations of MAX CSP\((P)\). For any fixed predicate \(P : \{-1,1\}^k \rightarrow \{0,1\}\), MAX CSP\((P)\) has a natural SDP relaxation that can be seen in Figure 2.3(a). The essence of this relaxation is that each \(I_{S,\omega}\) is a distribution, often called a local distribution, over all possible assignments to the variables in set \(S\) as enforced by (2.5). Whenever \(S_1\) and \(S_2\) intersect, (2.6) guarantees that their marginal distributions on the intersection agree. Also, (2.7) and (2.8) ensure that \(v_0 \cdot v_i\) and \(v_i \cdot v_j\) are equal to \(E[x_i]\) and \(E[x_i x_j]\) respectively where the expectations are according to the local distributions (i.e. the bias of variable \(x_i\) and the correlation of the variables \(x_i\) and \(x_j\).) The clauses of the instance are \(C_1, \ldots, C_m\), with \(C_i\) being an application of \(P\) (possibly with some variables negated) on the set of variables \(T_i\). The objective function is the fraction of the clauses that are satisfied.

Observe that the reason this SDP is not an exact formulation but a relaxation is that these distributions are defined only on sets of size up to \(k\).\(^3\) We note that approximation algorithms based on this relaxation often solve the relaxation and then assign a value to the variable \(x_i\) only based on the vectors \(v_0\) and \(v_i\). In particular, the variables \(I_{S,\omega}\) are often not used directly, rather the existence of the local distributions that agree in marginals with \(v_i\)'s is used in the proof of the approximation factor of the algorithm.

The relaxation in Figure 2.3(a) can be easily extended to MAX \(\kappa\)-CSP\(_q\)(\(P\)). In particular, for any \(P : [q]^k \rightarrow \{0,1\}\) Figure 2.3(b) shows the standard SDP relaxation of MAX \(\kappa\)-CSP\(_q\)(\(P\)).

2.3 Lift and Project Systems

As we mentioned in the previous section the integrality gap of a relaxation sets the limit on the approximation ratio of algorithms based on it. It is then natural in the search for the approximability threshold of a combinatorial optimization problem to look for a relaxation with the best possible integrality gap. One way to find such a good relaxation would be to start with a canonical relaxation of the problem and then “strengthen” it, i.e. add extra constraints to it that are satisfied by all “integral solutions” (i.e. solutions of the original integer/quadratic program) but are not satisfied by some of the “fractional” (i.e. non-integral) solutions of the relaxation.

As an example consider the SDP relaxation of VERTEX COVER seen in Figure 2.2(b). For all \(i, j \in [n]\) we can add the constraint \(v_i \cdot v_j \geq 0\). Notice that for any integral solution, i.e. for

\(^3\)We will see that the idea of having variables \(I_{S,\omega}\) to express a “local” distribution over the variables in every set \(S\) of small size (\(|S| \leq k\) here) is closely related to the idea behind the Sherali-Adams hierarchies. In fact the relaxation in Figure 2.3(a) is almost identical to a Sherali-Adams SDP relaxation.
Maximize \[
\frac{1}{m} \sum_{i=1}^{m} \sum_{\omega \in \{-1, 1\}^T_i} C_i(\omega) \ I_{T_i, \omega}
\]

Where,

\[
\forall S \in \binom{[n]}{\leq k}, \omega \in \{-1, 1\}^S \quad I_{S, \omega} \in [0, 1]
\]

\[
\forall i \in [n] \quad \mathbf{v}_i \in \mathbb{S}^n
\]

\[
\mathbf{v}_0 \in \mathbb{S}^n
\]

Subject to

\[
\forall S \in \binom{[n]}{\leq k} \quad \sum_{\omega \in \{-1, 1\}^S} I_{S, \omega} = 1 \quad (2.5)
\]

\[
\forall S \subset S' \in \binom{[n]}{\leq k}, \omega \in \{-1, 1\}^S \quad \sum_{\omega' \in \{-1, 1\}^{S' \setminus S}} I_{S', \omega' \circ \omega} = I_{S, \omega} \quad (2.6)
\]

\[
\forall i \in [n] \quad \mathbf{v}_0 \cdot \mathbf{v}_i = I_{\{i\}, (1)} - I_{\{i\}, (1)} \quad (2.7)
\]

\[
\forall i, j \in [n] \quad \mathbf{v}_i \cdot \mathbf{v}_j = I_{\{i, j\}, (1, 1)} + I_{\{i, j\}, (1, -1)} - I_{\{i, j\}, (-1, -1)} \quad (2.8)
\]

(a) Standard SDP relaxation of MAX CSP\((P)\)

Maximize \[
\frac{1}{m} \sum_{i=1}^{m} \sum_{\omega \in \{-1, 1\}^T_i} C_i(\omega) \ I_{T_i, \omega}
\]

Where,

\[
\forall S \in \binom{[n]}{\leq k}, \omega \in \{0, 1\}^S \quad I_{S, \omega} \in [0, 1]
\]

\[
\forall i \in [n], l \in [q] \quad \mathbf{v}_{i, l} \in \mathbb{S}^n
\]

\[
\mathbf{v}_0 \in \mathbb{S}^n
\]

Subject to

\[
\forall S \in \binom{[n]}{\leq k} \quad \sum_{\omega \in \{0, 1\}^S} I_{S, \omega} = 1
\]

\[
\forall S \subset S' \in \binom{[n]}{\leq k}, \omega \in \{0, 1\}^S \quad \sum_{\omega' \in \{0, 1\}^{S' \setminus S}} I_{S', \omega' \circ \omega} = I_{S, \omega}
\]

\[
\forall i \in [n], l \in [q] \quad \mathbf{v}_{0} \cdot \mathbf{v}_{i, l} = I_{\{i\}, (1)}
\]

\[
\forall i, j \in [n], l, l' \in [q] \quad \mathbf{v}_{i, l} \cdot \mathbf{v}_{j, l'} = I_{\{i, j\}, (l, l')}
\]

(b) Standard SDP relaxation of MAX \(\kappa\)-CSP\(q\)(\(P\))

Figure 2.3: Standard SDP relaxations of MAX \(\kappa\)-CSP\(P\) and MAX \(\kappa\)-CSP\(q\)(\(P\))
any solution of the quadratic program,

\[ v_i \cdot v_j = (v_i \cdot v_0)(v_j \cdot v_0) = (v_i \cdot v_i)(v_j \cdot v_j) \geq 0, \]

so if we add these constraints to the canonical SDP relaxation we get another relaxation. In fact the constraints \( v_i \cdot v_j \geq 0 \) are one kind of “triangle inequality” constraints\(^4\) and if one adds all the triangle inequality constraints to the SDP relaxation of Figure 2.2(b) one gets the relaxation used by Karakostas [Kar05] to get his \( 2 - O(1/\sqrt{\log n}) \)-approximation algorithm for \textsc{Vertex Cover}.

Now imagine that we want to design an approximation algorithm for \textsc{Vertex Cover} with a better approximation factor. One approach is to come up with another set of constraints that are valid for integral solutions and add them to the above relaxation. In fact triangle inequalities themselves have been used to strengthen the relaxations and devise better approximation algorithms for many problems other than \textsc{Vertex Cover}, the celebrated result of Arora, Rao and Vazirani [ARV09] being a prime example.

This approach however is rather ad-hoc as one never quite knows which of a big family of constraints would be useful for a particular optimization problem. For example, for \textsc{Vertex Cover} problem, Goemans and Kleinberg [GK98] showed that the integrality gap of the SDP relaxation in Figure 2.2(b) is at least \( 2 - o(1) \). They then asked if adding (a subset of) the triangle inequalities would decrease the integrality gap. Charikar [Cha02] (see also the discussion in [HMM07]) answered this question negatively. This chain of results continued with various researchers studying if any of a number of inequalities would decrease the integrality gap of the relaxation to \( 2 - \epsilon \) for some positive constant \( \epsilon > 0 \); see e.g. [HMM07, GMT08].

The so called “Lift and Project” systems are systematic ways to start from a relaxation of a problem (which we call the original/canonical relaxation) and strengthen it, i.e. find and add valid constraints to it, to arrive at a new relaxation. The resulting relaxation would have a smaller set of fractional solutions but it will still have all integral solutions of the original relaxation. These systems have a parameter called level which controls how many (and what kind of) valid constraints are present in the resulting program. When the level parameter is 0 the resulting program is simply the original relaxation. The resulting relaxation for level \( k + 1 \) is stronger than (i.e. has all the constraints present in) the resulting relaxation for level \( k \); hence the resulting relaxations can be thought of as a hierarchy of increasingly tighter relaxations. Resulting relaxations for larger level parameter, i.e. stronger relaxations, (in the same system) are referred to as higher levels of the hierarchy.

The commonly studied lift and project systems include the Lovász-Schrijver LP, Lovász-Schrijver SDP [LS91], Sherali-Adams LP, Sherali-Adams SDP [SA90]\(^5\), and Lasserre [Las02] hierarchy.

---

\(^4\)The reason being that if one defines \( w_i \)'s as \( w_i = v_0 - 2v_i \) then the constraints can be rewritten as \( ||w_0 - w_i||^2 + ||w_0 - w_j||^2 \geq ||w_i - w_j||^2 \).

\(^5\)What is often called the Sherali-Adams SDP hierarchy is in fact a “mixed hierarchy”. In particular, it is the result of adding a simple positive-semidefiniteness constraint to the Sherali-Adams LP relaxation. The Sherali-Adams LP hierarchy is due to Sherali and Adams [SA90] while (a close variant) of the Sherali-Adams SDP hierarchy is due to Sherali and Adams [SA90].
archies; see [Mon03] for a comparison. When the starting relaxation is clear from the context, specially when it is the canonical LP relaxation of the problem, the resulting relaxation for level parameter $l$ of a hierarchy is often called the level-$l$ relaxation, e.g. the level-$l$ Sherali-Adams SDP relaxation of Vertex Cover. For all these hierarchies the level $n$ relaxation is a perfect formulation of the problem and has integrality gap 1. On the other hand it is possible to find a solution of the level $l$ relaxation in time $n^{O(l)}$. It is worthwhile to note that the level-$l$ Lovász-Schrijver LP relaxation is weaker than the level-$O(l)$ Sherali-Adams LP relaxation which is in turn weaker than the level-$O(l)$ Sherali-Adams SDP relaxation. Furthermore, level-$l$ of all of these relaxations is weaker than level-$O(l)$ Lasserre relaxation.

The main power of these hierarchies is that they automatically include many of the specific constraints used in many approximation algorithms. For example, if one starts with the canonical LP relaxation of Vertex Cover seen in Figure 2.2(a) and applies Sherali-Adams LP or Sherali-Adams SDP system of level 2 one gets a relaxation which has the triangle inequality constraints. If one applies the same systems with level 3 the resulting relaxation will also have the “pentagonal” constraints studied in [HMM07]. And finally if one applies level 2 of the Lovász-Schrijver SDP, Sherali-Adams SDP, or Lasserre hierarchies the resulting relaxation will be stronger than the SDP relaxation of Vertex Cover seen in Figure 2.2(b). On one hand this indicates that the hierarchies could be useful in designing approximation algorithms; see Section 2.3.3 for some example of approximation algorithms which use them. On the other hand, if one believes that the approximability threshold of a problem is $\alpha$ but cannot prove this an alternative goal would be to show that the level-$l$ relaxation in one of these hierarchies for a large value of $l$ has integrality gap at least $\alpha$. For example, as mentioned before it is believed that the approximation threshold for Vertex Cover is 2 but such a theorem has proven very hard to establish assuming only $P \neq NP$, and a string of papers have studied the integrality gap of level-$l$ relaxation of all the previous hierarchies for various values of $l$. We note that this line of research effectively started with the work of Arora et al. [ABL06] on the integrality gap of the Lovász-Schrijver LP relaxation of Vertex Cover among others; see also the conference version [ABL02].

Almost all the results in this thesis involve the Sherali-Adams SDP hierarchy. In Section 2.3.1 we give a formal definition of this hierarchy. This formal definition is rather cumbersome so in Section 2.3.2 we give a simpler characterization which is equivalent to the definition for the problems we will study. We then proceed to present the level-$l$ Sherali-Adams SDP relaxation for some of the problems studied in this thesis. In Section 2.3.3 we present some of the previous work on the lift and project systems; results which are particularly relevant to a specific chapter however are delayed until those chapters.

---

6Here $n$ is the number of variables in the original relaxation.

7Khot and Regev [KR08] show that assuming UGC the approximation threshold for Vertex Cover is 2.
2.3.1 A formal definition of the Sherali-Adams hierarchy

Sherali and Adams [SA90] proposed a systematic procedure for starting from the LP relaxation of a 0/1 integer program and strengthening it to decrease its integrality gap. There are multiple equivalent ways to define this system but we will (mostly) follow the definition and notation of Laurent [Lau03]. We will need the following notation. For $y \in \mathbb{R}^{\wp([n])}$, $t \leq n$ and $U \in \wp([n])$ we define the matrices

\[(M_t(y))_{I,J} := y_{I \cup J}, \quad \forall I,J \in \binom{[n]}{\leq t}\]

\[(M_U(y))_{I,J} := y_{I \cup J}, \quad \forall I,J \in \wp(U).\]

We will sometimes use the shorthands $y_0 = y_{\emptyset}, y_i = y_{\{i\}}$. From the definitions above it is clear that $M_n(y) = M_{[n]}(y)$, furthermore the definition of $M_t(y)$ extends to $y \in \mathbb{R}^{\binom{[n]}{\leq t}}$ and that of $M_U(y)$ to $y \in \mathbb{R}^{\binom{[n]}{|U|}}$. For any $A \subseteq [n]$ we define the “shifting” $y_{+A}$ of $y \in \mathbb{R}^{\wp([n])}$, as a vector in $\mathbb{R}^{\wp([n])}$, defined as $(y_{+A})_I = y_{A \cup I}$.

To describe the Sherali-Adams hierarchy we will (for now) focus on VERTEX COVER, its IP formulation in Figure 2.1(a) and its LP relaxation in Figure 2.2(a). Consider an integral solution $x$ of the relaxation for some $n$-vertex graph $G = (V,E)$, and define $y \in \mathbb{R}^{\wp([n])}$ as $y_I = \prod_{i \in I} x_i$. Then it is easy to check that

\[
\forall i \in V \quad x_i = y_{\{i\}},
\]

\[
M_{[n]}(y) = y \otimes y \succeq 0,
\]

\[
\forall ij \in E \quad M_{[n]}(y_{+\{i\}} + y_{+\{j\}} - y) = M_{[n]}(x_i y + x_j y - y) = (x_i + x_j - 1)M_{[n]}(y) \succeq 0
\]

In other words the above constraints are satisfied by any integral solution. In fact, Sherali and Adams [SA90] show that the only fractional solutions satisfying these constraints are the convex combinations of integral solutions and therefore adding them to LP relaxation reduces the integrality gap to 1. However, the resulting program has exponentially many variables and constraints and is not known to be solvable in polynomial time.

Sherali and Adams proposed the following relaxation, which has variables $y \in \mathbb{R}^{\binom{[n]}{\leq t}}$, for some fixed $t$.

**Definition 2.6.** The following mathematical program is known as the level-$t$ Sherali-Adams LP relaxation for VERTEX COVER.

\[
\begin{align*}
\min \quad & \sum_{i \in V} y_{\{i\}} \\
\text{s.t.} \quad & M_U(y_{+\{i\}} + y_{+\{j\}} - y) \succeq 0 \quad \forall ij \in E, \forall U \in \binom{[n]}{\leq t-1} \\
& M_U(y) \succeq 0 \quad \forall U \in \binom{[n]}{t-1} \\
& y_{\{i\}} \in [0,1] \quad \forall i \in V,
\end{align*}
\]

The level-$t$ Sherali-Adams LP relaxation for any combinatorial program is defined in a similar way.
way. The only difference being that if the original relaxation has a constraint $\sum_i a_i x_i \geq b$ then the first constraint of 2.12 is replaced by,

$$M_{U}(\sum_i a_i y_{+\{i\}} - by) \succeq 0.$$  

Interestingly, the above relaxation is a linear program in disguise; see [Lau03]. We do not need to take advantage of this fact however as we will use a simpler definition for Sherali-Adams hierarchy discussed in Section 2.3.2.

We now turn to the definition of the Sherali-Adams SDP relaxation. Notice that the main difference between the relaxations in Figure 2.2 is that the one in Figure 2.2(b) has the values $v_i \cdot v_j$ meant to encode the value $x_i x_j$. In fact, as discussed in Section 2.2 one could reformulate the SDP relaxation by taking $X = [X_{ij}]$ to be the (matrix valued) variable in the program and add the constraint $X \succeq 0$. This motivates the following relaxation for Vertex Cover and other optimization problems which is stronger than both the Sherali-Adams LP relaxation and (often) the canonical SDP relaxation of the problem.

**Definition 2.7.** Level-$t$ Sherali-Adams SDP relaxation of Vertex Cover is defined as follows.

$$\begin{align*}
\min & \sum_{i \in V} y_i \\
\text{s.t. } & M_{U}(y_{+\{i\}} + y_{+\{j\}} - y) \succeq 0 \quad \forall ij \in E, \forall U \in \left(\binom{[n]}{\leq t - 1}\right) \\
& M_{U}(y) \succeq 0 \quad \forall U \in \left(\binom{[n]}{\leq t}\right) \\
& M_{1}(y) \succeq 0 \\
& y_i \in [0, 1] \quad \forall i \in V,
\end{align*}$$

The level-$t$ Sherali-Adams SDP relaxation for any combinatorial program is defined as an extension of this definition as in Definition 2.6.

We conclude this subsection by defining the Lasserre system which is the strongest system among the ones discussed so far. One of the most challenging problems in the area is to prove tight integrality gaps in this system. The main difference between the Lasserre and Sherali-Adams systems is the constraint $M_t(y) \succeq 0$ that is present in the Lasserre but not in the Sherali-Adams system. Notice that for all $U \in \left(\binom{[n]}{\leq t}\right)$, $M_U(y)$ is a minor of $M_t(y)$, so $M_t(y) \succeq 0$ is a stronger requirement than $M_U(y) \succeq 0$. So the level-$t$ Sherali-Adams SDP relaxation is stronger than the level-$t$ Sherali-Adams LP relaxation but weaker than the level-$t$ Lasserre relaxation.

**Definition 2.8.** Level-$t$ Lasserre relaxation of Vertex Cover is defined as follows.

$$\begin{align*}
\min & \sum_{i \in V} y_i \\
\text{s.t. } & M_{t-1}(y_{+\{i\}} + y_{+\{j\}} - y) \succeq 0 \quad \forall ij \in E \\
& M_{t}(y) \succeq 0 \\
& y_i \in [0, 1] \quad \forall i \in V,
\end{align*}$$
The level-$t$ Lasserre relaxation for any combinatorial program is defined as an extension of this definition as in Definition 2.6.

### 2.3.2 An alternative definition of the Sherali-Adams Hierarchy

In this subsection we will give an alternative way of defining Sherali-Adams LP and Sherali-Adams SDP hierarchies. This alternative definition was first discussed by [dlVKM07] and later [CMM09, GM08] among others. We need the following two definitions.

**Definition 2.9.** Consider a combinatorial optimization problem where the $n$ variables take values from domain $[q]$. For an integer $t \geq 1$, a family of (consistent) local distributions is a set of distributions $\{D(S)\}_{S \subseteq [n]}$ one for each subset of $[n]$ of size at most $t$ which are consistent on their marginals. In other words,

$$\forall S \subseteq T \in \binom{n}{\leq t}, \quad D(S) = \text{Mar}_S D(T). \tag{2.14}$$

**Definition 2.10.** Consider a combinatorial optimization problem where $n$ variables take values from domain $\{-1, 1\}$ or $\{0, 1\}$. For an integer $t \geq 2$ we say that the vectors $u_0, u_1, \ldots, u_n$ and a family of local distributions $\{D(S)\}_{S \subseteq [n]}$ are compatible if,

$$\forall i \in [n], \quad u_0 \cdot u_i = \mathbb{E}_{x \sim D\{i\}} [x_i], \tag{2.15}$$

$$\forall i, j \in [n], \quad u_i \cdot u_j = \mathbb{E}_{x \sim D\{i,j\}} [x_i x_j]. \tag{2.16}$$

in other words the distributions and the vectors “agree on the marginals.” When the domain of the variables is $[q]$ we say that the vectors $u_0, \{u_{i,a}\}_{i \in [n], a \in [q]}$ and a family of local distributions $\{D(S)\}_{S \subseteq [n]}$ are compatible if,

$$\forall i \in [n], a \in [q], \quad u_0 \cdot u_{i,a} = \mathbb{P}_{x \sim D\{i\}} [x_i = a], \tag{2.17}$$

$$\forall i, j \in [n], a, b \in [q], \quad u_{i,a} \cdot u_{j,b} = \mathbb{P}_{x \sim D\{i,j\}} [x_i = a \land x_j = b]. \tag{2.18}$$

We are now ready to state a fact which essentially gives a reformulation of the Sherali-Adams LP hierarchy.

**Fact 2.1.** Consider a combinatorial optimization problem and its canonical LP relaxation. Assume that each constraint of the relaxation involves at most $k$ variables and $l \geq 2$ is an integer and let $m = l + \max(0, k - 1)$. Suppose that we can associate a distribution $D(U)$ to every set $U \in \binom{S}{\leq m}$ such that $\{D(U)\}$ is a family of consistent local distributions. Furthermore, assume that these distributions are “local” solutions, i.e. if $\sum_i a_i x_i \geq b$ is a constraint of the
LP relaxation with support \( S = \{ i : a_i \neq 0 \} \) (\( |S| \leq k \)), then,

\[
\Pr_{x \sim \mathcal{D}(S)} \left[ \sum_i a_i x_i \geq b \right] = 1.
\] (2.19)

Define the vector \( y \in \mathbb{R}^{[n]} \) as \( y_U = \mathbb{E}_{x \sim \mathcal{D}(U)} \left[ \prod_{i \in U} x_i \right] \). Then \( y \) is a solution of the level-l Sherali-Adams LP relaxation of the problem, i.e. it satisfies the conditions of Definition 2.6.

Similarly adding a simple condition would result in a reformulation of the Sherali-Adams SDP hierarchy.

**Fact 2.2.** Consider a combinatorial optimization problem, its canonical LP relaxation, \( k, l, m \) and \( \{ \mathcal{D}(U) \} \) satisfying the conditions of Fact 2.1. If there is a set of vectors \( u_0, u_1, \ldots, u_n \) or \( u_0, \{ u_{i,a} \}_{i \in [n], a \in [q]} \) (depending on the domain of the variables in the problem) that are compatible with \( \{ \mathcal{D}(U) \} \) per Definition 2.10 then \( y \in \mathbb{R}^{[n]} \) defined in Fact 2.1 together with the set of vectors is a solution of the level-l Sherali-Adams SDP relaxation of the problem, i.e. it satisfies the conditions of Definition 2.7.

Notice that when the domain of the variables in the problem are \( \{0,1\} \) or \( \{-1,1\} \) we can rewrite the condition of Fact 2.2 as follows. Define an \((n+1) \times (n+1)\) matrix \( M_1 = [m_{I,J}] \) whose rows and columns are indexed by \( \emptyset, \{1\}, \ldots, \{n\} \) and

\[
m_{I,J} = \mathbb{E}_{U \sim \mathcal{D}(U)} \left[ \left( \prod_{i \in I} x_i \right) \left( \prod_{j \in J} x_j \right) \right].
\]

The extra condition in Fact 2.2 is equivalent to \( M_1 \succeq 0 \).

### 2.3.3 Previous work on Lift-and-Project hierarchies

In this section we present some of the (many) previous works on Lift and Project hierarchies. It is not hard to check that the relaxations used in the best approximation algorithms for many combinatorial optimization problems are weaker than their level-\( k \) Sherali-Adams SDP relaxation for a small constant \( k \), say 4; examples include Max Cut [GW95], Max 3-Sat [KZ97], Sparsest cut [ARV09], and Vertex Cover [Kar05].

Fernández de la Vega and Kenyon-Mathieu [dlVKM07] have provided a PTAS for Max Cut in dense graphs using Sherali-Adams LP hierarchy. Magen and Moharrami [MM09] show how to get a Sherali-Adams based PTAS for Vertex Cover and Independent Set in minor-free graphs. Bateni, Charikar and Guruswami [BCG09] that the integrality gap of the level-\( k \) Sherali-Adams LP relaxation of the MaxMin allocation problem is at most \( n^{1/k} \). Chlamtac [Chl07] and Chlamtac and Singh [CS08] gave an approximation algorithm for Independent Set in hypergraphs based on the Lasserre hierarchy, with the performance depending on the number of levels. Recently, an \( O(n^{1/4}) \) approximation for Densest \( k \)-Subgraph was shown by Bhaskara.
et al. [BCC10], using a level-\(O(\log n)\) Lovász-Schrijver LP relaxation. Barak et al. [BRS11] present subexponential time algorithms for \textsc{Unique Label Cover} problem using the Sherali-Adams SDP hierarchy. Other examples of algorithms using Lift-and-Project hierarchies include [KMN11, GS11].

Lower bounds in Lift-and-Project hierarchies amount to showing that the integrality gap remains large even after many levels of the hierarchy. Studying such lower bounds was initiated by the seminal work of Arora et al. [ABL02, ABLT06]. Integrality gaps for \(\Omega(n)\) levels can be seen as unconditional lower bounds (as they rule out even exponential time algorithms obtained by the hierarchy) in a restricted (but still interesting) model of computation. Considerable effort was invested in proving such lower bounds; see e.g. [FK03, AAT11, Tou05, Tou06, BOGH06, dlVKM07, GMPT10, STT07b, STT07a, Sch08, CMM09, GMT09b, MS09, RS09b, Tul09].

Finally, for some particular constraint satisfaction problems, and modulo the Unique Games Conjecture, no approximation algorithm can perform better than the one obtained by level-\(k\) Sherali-Adams SDP relaxation for a constant \(k\); see [Rag08].

2.4 The Hamming cube and Fourier Representation

The set \(\{-1,1\}^n\) of \(n\) dimensional \(-1,1\) vectors is often called the Hamming cube. For two points \(x, y \in \{-1,1\}^n\) we use \(w_H(x)\), the Hamming weight of \(x\), for the number of \(-1\)'s in \(x\) and \(d_H(x, y)\), the Hamming distance of \(x\) and \(y\) for the number of coordinate \(i\), such that \(x_i \neq y_i\).

Consider the set of real functions with domain \(\{-1,1\}^n\) as a vector space. It is well known that the following set of functions called the characters form a complete basis, often called the Fourier basis, for this space,

\[
\chi_S(x) \overset{\text{def}}{=} \prod_{i \in S} x_i.
\]  

(2.20)

In fact if we define inner products of functions as \(f \cdot g \overset{\text{def}}{=} \mathbb{E}_x[f(x)g(x)]\) this basis will be orthonormal and every function will have a unique Fourier expansion when written in this basis,

\[
f = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S, \quad \hat{f}(S) \overset{\text{def}}{=} f \cdot \chi_S.
\]  

(2.21)

\(\hat{f}(S)\)'s are often called the Fourier coefficients of \(f\). We write \(f^=d\) for the part of the function that is of degree \(d\), i.e.,

\[
f^=d(x) = \sum_{|S|=d} \hat{f}(\{S\}) \chi_S(x).
\]  

(2.22)
An immediate corollary of the orthonormality of this basis is Parseval’s identity,

$$E_x[f(x)g(x)] = f \cdot g = \sum_S \hat{f}(S)\hat{g}(S).$$

One can easily express the Fourier coefficients of $\text{Mar}_S \mu$ in terms of those of $\mu$,

$$\text{Mar}_S \mu(U) = E_{x \in \{-1,1\}^S}[(\text{Mar}_S \mu)(x)\chi_U(x)] = E_{x \in \{-1,1\}^S} \left[ \sum_{z \in \{-1,1\}^{|\mu|}\setminus S} \mu(x \circ z)\chi_U(x) \right]$$

$$= 2^{-|S|} \sum_{x \in \{-1,1\}^S} \sum_{z \in \{-1,1\}^{|\mu|}\setminus S} \mu(x \circ z)\chi_U(x) = 2^{-|S|} \sum_{y \in \{-1,1\}^{|\mu|}} \mu(y)\chi_U(y) = 2^{n - |S|} \hat{\mu}(U). \quad (2.23)$$

In particular, for all $x \in \{-1,1\}^S$ we have

$$(\text{Mar}_S \mu)(x) = \sum_{U \subseteq S} \text{Mar}_S \mu(U)\chi_U(x) = 2^{n - |S|} \sum_{U \subseteq S} \hat{\mu}(U)\chi_U(x). \quad (2.24)$$

It is natural to define the following norm for functions $f : \{-1,1\}^n \to \mathbb{R}$,

$$\|f\|_2 \overset{\text{def}}{=} \sqrt{f \cdot \overline{f}} = \sqrt{E_x[f(x)^2]} = \sqrt{\sum_S \hat{f}(S)^2}. \quad (2.25)$$

### 2.5 Borsuk and Frankl-Rödl Graphs

Our integrality gap instances in Chapters 6 and 7 are (closely related to) the Frankl-Rödl graphs. These graphs are parameterized by an integer $m$ which is considered growing and a real parameter $0 < \gamma < 1$.

**Definition 2.11** (Frankl-Rödl graphs). The Frankl-Rödl graph $G^n_\gamma$ is the graph with vertices $\{-1,1\}^m$ where two vertices $i,j \in \{-1,1\}^m$ are adjacent iff $d_H(i,j) = (1 - \gamma)m$.

These graphs exhibit an interesting “extremal” combinatorial property. While $G^m_0$ is a perfect matching and thus $VC(G^m_0) = |V(G^m_0)|/2$, a beautiful theorem of Frankl and Rödl states that for slightly larger $\gamma$, any vertex cover of $G^m_\gamma$ is very large. The fact that such a small geometric perturbation results in a drastic change in the vertex cover size has led to the use of Frankl-Rödl graphs as integrality gap instances in a series of results [GK98, Cha02, GMPT10, GMT08, GMT09a]. The following theorem is of particular relevance to us.

**Theorem 2.3** ([GMPT10]; slight modification of Theorem 1.4 of [FR87]). Let $m$ be an integer and let $\gamma = \Theta(\sqrt{\log m/m})$ be a sufficiently small number for which $\gamma m$ is an even integer. Then any vertex cover of $G^n_\gamma$ contains at least a $1 - o(1)$ fraction of the vertices.
We often think of the vertices of the Frankl-Rödl graphs as (scaled and) embedded on the unit sphere $S^{m-1}$. In this sense the Frankl-Rödl graphs are subgraphs of the infinite Borsuk graphs.

**Definition 2.12. (Borsuk graphs)** The Borsuk graph $B^m_\delta$ is an infinite graph with vertex set $S^{m-1}$. Two vertices $x, y$ are adjacent if they are nearly antipodal, i.e. $\|x + y\| \leq 2\sqrt{\delta}$. 

Chapter 3

SDP Gaps from Pairwise Independence\(^1\)

In this chapter we consider the MAX k-CSP\(_q(P)\) problem of Definition 2.2. Recall that a predicate \(P : [q]^k \rightarrow \{0,1\}\) is called promising (Definition 2.4) if the set of assignments accepted by \(P\) contains the support of a balanced pairwise independent distribution. Austrin and Mossel [AM09] showed the following result modulo UGC: For any promising \(P\), MAX k-CSP\(_q(P)\) is hard to approximate beyond the trivial algorithm which assigns a uniformly random value to the variables.

We show that for the same \(P\), the level-\(\Omega(n)\) Sherali-Adams SDP relaxation of MAX k-CSP\(_q(P)\) has an integrality gap matching the above hardness of approximation. Our results can be viewed as an unconditional analogue of the one in [AM09] in a restricted computational model, namely strong SDP relaxations. For our proof we introduce a new generalization of techniques to define consistent “local distributions” over partial assignments to variables in the problem, which is often the crux of proving lower bounds for in the Sherali-Adams SDP hierarchy.

Other than being a generalization of MAX k-SAT, MAX k-Lin\(_q\) and other similar optimization problems the study of the hardness of approximation of MAX k-CSP\(_q(P)\) is also motivated by the study of MAX k-CSP\(_q\). In fact, the best hardness of approximation results of MAX k-CSP\(_q\) choose a suitable \(P\) and proceed to prove hardness of approximating MAX k-CSP\(_q(P)\). The MAX k-CSP\(_q\) problem is NP-hard for \(k \geq 2\), and a lot of effort has been devoted to determining its true inapproximability. For the case of Boolean variables \((q = 2)\), Samorodnitsky and Trevisan [ST00] proved that the problem is NP-hard to approximate better than a factor of \(2^k/2^{2\sqrt{k}}\), which was improved to \(2^k/2^{\sqrt{2k}}\) by Engebretsen and Holmerin [EH05]. Later Samorodnitsky and Trevisan [ST06] showed that it is Unique-Games-hard to approximate the same problem with factor better than \(2^k/2^{\lceil \log k + 1 \rceil}\). For general \(q\), Guruswami and Raghavendra [GR08] showed a UGC-hardness ratio of \(q^k/kq^2\) when \(q\) is a prime.

As mentioned above Austrin and Mossel [AM09] show that if \(P : [q]^k \rightarrow \{0,1\}\) is a promis-

\(^1\)Results in this chapter appear in [BGMT12]
ing predicate then MAX $k$-CSP$_q(P)$ is UGC-hard to approximate better than a factor of $q^k/|P^{-1}(1)|$. Using appropriate choices for the predicate $P$, this then implies hardness ratios of $q^k/kq^2(1+o(1))$ for MAX $k$-CSP$_q$ for general $q \geq 2$, $q^k/kq(q-1)$ when $q$ is a prime power, and $2^k/(k + O(k^{0.525}))$ for $q = 2$; thus superseding all the previously mentioned results, except those of [ST00, EH05] which are incomparable as they do not rely on UGC.

**Previous Work**

There is a wealth of work on the MAX $k$-CSP$_q$ problem and its variants the most relevant of which to our result is the previously mentioned paper of Austrin and Mossel [AM09]. In terms of integrality gaps for strong relaxations of MAX $k$-CSP$_q(P)$, for some specific predicates $P$, strong lower bounds ($\Omega(n)$ levels) were proved recently for the Lasserre hierarchy (which is stronger than the Sherali-Adams SDP hierarchy studied here) by [Sch08] and [Tul09]. They show a factor 2 integrality gap for MAX $k$-XOR and factor $2^k/2k$ integrality gap for MAX $k$-CSP.$^2$

In a beautiful result, Raghavendra [Rag08] showed a general connection between integrality gaps and UG-hardness results. His result essentially shows that for MAX $k$-CSP$(P)$, the integrality gap of a program obtained by $k$ levels of the Sherali-Adams SDP hierarchy is $I$, then the MAX $k$-CSP$(P)$ is UG-hard to approximate better than a factor of $I$. However, in our case the hardness is already known (by the work of Austrin and Mossel), and we are interested in finding the integrality gap for programs obtained by $\Omega(n)$ levels.

**Our Result and Techniques**

Both previously known results in the Lasserre hierarchy and previous analogues in the Lovász-Schrijver hierarchy seemed to heavily rely on the structure of the predicate for which the integrality gap was proved, in particular, the predicate is always some system of linear equations. It was not clear if the techniques could be extended using only the fact that the predicate is promising which is a much weaker condition. In this chapter, we try to explore this issue, proving $\Omega(n)$ level gaps for the Sherali-Adams SDP hierarchy for any promising predicate.

**Theorem 3.1.** Let $P : [q]^k \rightarrow \{0,1\}$ be a promising predicate (see Definition 2.4.) Then for every constant $\zeta > 0$, there exist $c = c(q, k, \zeta) > 0$ such that for large enough $n$, the integrality gap of the level-$cn$ Sherali-Adams SDP relaxation of MAX $k$-CSP$_q(P)$ is at least $q^k/|P^{-1}(1)| - \zeta$.

**Remark 3.2.** We note that weaker integrality gaps for these predicates also follow, via reductions, from the corresponding integrality gap results for Unique Games. In particular, a $(\log \log n)^{\Omega(1)}$-level gap for the SDP hierarchy discussed above follows from the recent results of Raghavendra and Steurer [RS09b]. Also, $\Omega(n^\delta)$-level gaps (where $\delta \rightarrow 0$ as $\zeta \rightarrow 0$) for the

---

$^2$Grigoriev [Gri01] shows a similar lowerbound for MAX $k$-XOR. In his work the lowerbound is for the degree parameter of the, so-called, “Positivstellenstatz Calculus Proofs” It is now known that such lowerbounds imply integrality gap lowerbounds for the Lasserre hierarchy.
(weaker) **Sherali-Adams LP** hierarchy can be deduced from the results of Charikar, Makarychev and Makarychev [CMM09].

A first step in achieving our result is to reduce the problem of a level-$t$ gap to a question about family of distributions over assignments associated with sets of variables of size at most $t$ (see also the discussion in Section 2.3.) These distributions should be (a) supported only on satisfying (partial) assignments, (b) should be consistent among themselves, in the sense that for $S_1 \subseteq S_2$ which are subsets of variables, the distributions over $S_1$ and $S_2$ should be equal on $S_1$, and (c) should be balanced and pairwise-independent. The first requirement guarantees that the solution achieves objective value that corresponds to satisfying all the constraints of the instance. The second requirement implies feasibility for the **Sherali-Adams LP** constraints, while the last one makes it (significantly) easy to produce vectors satisfying the semidefinite constraints.

The second step is to come up with these distributions! We explain why the simple method of picking a uniform distribution (or a reweighting of it according to the pairwise independent distribution that is supported by $P$) over the satisfying assignments cannot work. Instead we introduce the notion of “advice sets”. These are sets on which it is “safe” to define such simple distributions. The actual distribution for a set $S$ we use is then the one induced on $S$ by a simple distribution defined on the advice-set of $S$. Getting such advice sets heavily relies on notions of expansion of the constraint graph of the instance discussed in Section 3.1. In particular, we use the fact that random instances have inherently good expansion properties. At the same time, such instances are highly unsatisfiable, ensuring that the resulting integrality gap is large.

Arguing that it is indeed “safe” to use simple distributions over the advice sets relies on the fact that the predicate $P$ in question is promising, namely $P^{-1}(1)$ contains the support of a balanced pairwise independent distribution. We find it interesting and somewhat curious that the condition of pairwise independence comes up in this context for a reason very different than in the case of UG-hardness. Here, it represents the limit to which the expansion properties of a random CSP instance can be pushed to define the local distributions.

### 3.1 Expanding CSP Instances

For an instance $\Psi$ of MAX $\kappa$-CSP$_q$, define its constraint graph $G_\Psi$, as the following bipartite graph with sides $L$ and $R$. The left hand side $L$ consists of a vertex for each constraint $C_i$. The right hand side $R$ consists of a vertex for every variable $x_j$. There is an edge between a constraint-vertex $i$ and a variable-vertex $j$, whenever variable $x_j$ appears in constraint $C_i$. When it is clear from the context, we will abbreviate $G_\Psi$ by $G$.

For $C_i \in L$ we denote by $\Gamma(C_i) \subseteq R$ the neighbours $\Gamma(C_i)$ of $C_i$ in $R$. For a set of constraints $\mathcal{C} \subseteq L$, $\Gamma(\mathcal{C})$ denotes $\bigcup_{C_i \in \mathcal{C}} \Gamma(C_i)$. Similarly, for $u \in R$ we denote by $\Gamma(u) \subseteq L$ the neighbours of $u$ in $L$.

For $S \subseteq R$, we call a constraint $C_i \in L$, $S$-dominated if $\Gamma(C_i) \subseteq S$ and use $\mathcal{C}(S)$ for the
set of all $S$-dominated constraints. We denote by $G|_{-S}$ the subgraph of $G$ that we get after removing $S$ and all $S$-dominated constraints.

Our results in this chapter rely on the expansion properties of the graph $G_\Psi$ for a random instance $\Psi$. We will defer the precise definition of these properties to Lemma 3.7 but the following definition plays a central role in them.

**Definition 3.1.** Consider a bipartite graph $G = (V,E)$ with partition $L, R$. The *boundary expansion* of $X \subset L$ is the value $|\partial X|/|X|$, where $\partial X = \{u \in R : |\Gamma(u) \cap X| = 1\}$. $G$ is $(r,e)$ boundary expanding if the boundary expansion for all (nonempty) subsets of $L$ of size at most $r$ is at least $e$.

### 3.2 Towards Defining Consistent Distributions

To construct valid solutions for the Sherali-Adams SDP hierarchy, we need to define distributions over every set $S$ of bounded size as is required by Facts 2.1 and 2.2. Since we will deal with promising predicates supported by some distribution $\mu$, in order to satisfy consistency between distributions we will heavily rely on the fact that $\mu$ is a balanced pairwise independent distribution.

For now imagine that $\mu$ is the uniform distribution over $P^{-1}(1)$; the intuition for the general case is not significantly different. It is instructive to think of $q = 2$ and the predicate $P$ being $\text{MAX~k-XOR}$, $k \geq 3$. Observe that the uniform distribution over $P^{-1}(1)$ is pairwise independent and balanced. A first attempt would be to define for every $S$, the distribution $D(S)$ as the uniform distribution over all consistent assignments of $S$, i.e. assignments to variables in $S$ that satisfy all clauses that have all their variables in $S$. We argue that such distributions are in general problematic. This follows from the fact that satisfying assignments are not always extendible. Indeed, consider two constraints $C_{i_1}, C_{i_2} \in L$ that share a common variable $j \in R$. Set $S_2 = T_{i_1} \cup T_{i_2}$, and $S_1 = S_2 \setminus \{j\}$.\(^3\) Assuming that the support of no other constraint is contained in $S_2$, we get that distribution $D(S_1)$ maps any variable in $S_1$ to $\{0,1\}$ with probability $1/2$ independently, but some of these assignments are not extendible to $S_2$ which means that $D(S_2)$ has to assign probability zero to them.

Thus, to define $D(S)$, we cannot simply sample assignments satisfying all constraints in $\mathcal{C}(S)$ with probabilities given by $\mu$. In fact the above example shows that any attempt to blindly assign a set $S$ with a distribution that is supported on all satisfying assignments for $S$ is bound to fail. At the same time it seems hard to reason about a distribution that uses a totally different concept. To overcome this obstacle, we take a two step approach:

1. For a set $S$ we define a superset $\overline{S}$ such that $\overline{S}$ is “global enough” to contain sufficient information, while it is also “local enough” so that $\mathcal{C}(\overline{S})$ is not too large. We define $\overline{S}$ so that if we remove $\overline{S}$ and $\mathcal{C}(\overline{S})$ from $G_\Psi$, the remaining graph, $G|_{-\overline{S}}$, still has good expansion. In Section 3.2.1 we discuss how to find $\overline{S}$ from $S$.

\(^3\)Remember that $T_{i_1}$ is the set of variables appearing in $C_{i_1}$ and $T_{i_2}$ is the set of variables appearing in $C_{i_2}$.
2. When \( \mu \) is the uniform distribution over \( P^{-1}(1) \), the distribution \( D(S) \) is going to be the uniform distribution over satisfying assignments in \( S \). In the case that \( \mu \) is not uniform over \( P^{-1}(1) \), we give a natural generalization to the above uniformity. We show how to define distributions, which we denote by \( P_\mu(S) \), such that for \( S_1 \subseteq S_2 \), the distributions \( P_\mu(S_1) \) and \( P_\mu(S_2) \) are guaranteed to be consistent if \( G|_{-S_1} \) has good expansion. This appears in Section 3.2.2.

We then combine the two techniques and define \( D(S) \) according to \( P_\mu(S) \). This is done in Section 3.3.

### 3.2.1 Finding Advice-Sets

We now give an algorithm below to obtain a superset \( S \) for a given set \( S \), which we call the advice set of \( S \). It is inspired by the “expansion correction” procedure in [BOGH+06].

#### Algorithm 2: Algorithm Advice

**Input:** An \((r,e_1)\) boundary expanding bipartite graph \( G = (L,R,E), e_2 \in (0,e_1), S \subseteq R, |S| < (e_1-e_2)r \), with some (arbitrary) ordering \( S = \{x_1, \ldots, x_t\} \).

**Output:** The advice-set \( \overline{S} \)

1. \( \overline{S} \leftarrow \emptyset \)
2. \( \xi \leftarrow r \)

3. for \( j = 1, \ldots, |S| \) do

4. \( M_j \leftarrow \emptyset \)

5. \( \overline{S} \leftarrow \overline{S} \cup \{x_j\} \)

6. if \( G|_{-\overline{S}} \) is not \((\xi, e_2)\) boundary expanding then

7. Find a maximal \( M_j \subset L \) in \( G|_{-\overline{S}} \), such that \( |M_j| \leq \xi \) and \( |\partial M_j| \leq e_2|M_j| \) in \( G|_{-\overline{S}} \)

8. \( \overline{S} \leftarrow \overline{S} \cup \Gamma(M_j) \)

9. \( \xi \leftarrow \xi - |M_j| \)

end

12. return \( \overline{S} \)

**Theorem 3.3.** For any \((r,e_1)\)-boundary expanding, \( k \) left regular, bipartite graph \( G = (L,R,E) \) and \( 0 < e_2 < e_1 \), Algorithm Advice, when run with inputs \( G, e_1, e_2, r, \) and \( S \subseteq R \), returns \( \overline{S} \subseteq R \) such that

(a) \( G|_{-\overline{S}} \) is \((\xi_S, e_2)\) boundary expanding,

(b) \( \xi_S \geq r - \frac{|S|}{e_1-e_2}, \) and

(c) \( |\overline{S}| \leq \frac{(2e_1+k-e_2)|S|}{2(e_1-e_2)}. \)
Proof. Let \( \xi_S \) be the value of \( \xi \) when the loop terminates. From the bounded size of \( M_j \) and how \( \xi \) changes at each iteration we know that \( \xi \) remains non-negative throughout the execution of the while loop, and in particular \( \xi_S \geq 0 \). Note that at step \( j \), all the neighbours of \( M_j \) are added to the set \( \overline{S} \) so no member of \( M_j \) will be in \( G_{-\overline{S}} \) after the \( j \)th step. In particular, all the sets \( M_j \) will be disjoint.

In order to prove (a) we will prove the following loop invariant: \( G|_{-\overline{S}} = (\xi, e_2) \) boundary expanding. Indeed, note that the input graph \( G \) is \((\xi, e_1)\) boundary expanding so the invariant holds at the beginning. At step \( j \) consider the set \( \overline{S} \cup \{x_j\} \), and suppose that \( G_{-(\overline{S} \cup \{x_j\})} \) is not \((\xi, e_2)\) boundary expanding. We find maximal \( M_j \), \(|M_j| \leq \xi \), such that \(|\partial M_j| \leq e_2|M_j|\).

We claim that \( G_{-(\overline{S} \cup \{x_j\}) \cup (M_j)} \) is \((\xi - |M_j|, e_2)\) boundary expanding. Assuming the contrary, there must be \( M' \subset L \) such that \(|M'| \leq \xi - |M_j| \) and \(|\partial M'| \leq e_2|M'|\). As we mentioned above, \( M_j \) will be disjoint from the left vertices of \( G_{-(\overline{S} \cup \{x_j\}) \cup (M_j)} \), and in particular it will be disjoint from \( M' \). Consider then \( M_j \cup M' \) and note that \(|M_j \cup M'| \leq \xi \). More importantly (right before we added \( \Gamma(M_j) \) to \( \overline{S} \)) \(|\partial(M_j \cup M')| \leq e_2|M_j| + e_2|M'| = e_2|M_j \cup M'| \) contradicting the maximality of \( M_j \); (a) follows.

To show (b) we consider the set \( M = \bigcup_{j=1}^j M_j \) and upper bound and lower bound its boundary expansion in \( G \) in two different ways. Notice that as we mentioned \( M_j \)'s are disjoint, so \(|M| = \sum_{j=1}^j |M_j| = r - \xi_S \). Since \( G \) is \((r, e_1)\) boundary expanding, the set \( M \) has at least \( e_1(r - \xi_S) \) boundary neighbours in \( G \). But each member of \( \partial M \) is in the boundary of exactly one of the \( M_j \)'s, so it will be counted towards the boundary expansion of \( M_j \) in \( G_{-\overline{S}} \) in the \( j \)th iteration of the loop (for some \( j \)). Given that \( M_j \)'s have boundary expansion at most \( e_2 \) we have,

\[
e_1(r - \xi_S) \leq |\partial M| \leq |S| + \sum_j e_2|M_j| = |S| + e_2(r - \xi_S),
\]

which readily implies (b).

Finally note that \( \overline{S} \) consists of \( S \) union the neighbours of all \( M_j \)'s. But given that \( M_j \)'s have boundary expansion at most \( e_2 \) they cannot have a big neighbour set. In particular,

\[
|\overline{S}| \leq |S| + \sum_j |\Gamma(M_j)| = |S| + \sum_j (|\partial M_j| + |\Gamma(M_j) \setminus \partial M_j|) \leq |S| + \sum_j (e_2 + \frac{k - e_2}{2})|M_j| = |S| + \frac{e_2 + k}{2} \sum_j |M_j| \leq |S| + \frac{(e_2 + k)|S|}{2(e_1 - e_2)} = \frac{2e_1 + k - e_2)|S|}{2(e_1 - e_2)},
\]

which proves (c). \( \blacksquare \)

### 3.2.2 Defining the Distributions \( \mathcal{P}_\mu(S) \)

We now define for every set \( S \), a distribution \( \mathcal{P}_\mu(S) \) such that for any \( \alpha \in [q]^S \), \( \Pr_{\mathcal{P}_\mu(S)}[\alpha] > 0 \) only if \( \alpha \) satisfies all the constraints in \( \mathcal{C}(S) \). For a constraint \( C_i \) with set of inputs \( T_i \), defined
as \( C_i(x_{i_1}, \ldots, x_{i_k}) \equiv P(x_{i_1} + a_{i_1}, \ldots, x_{i_k} + a_{i_k}) \), let \( \mu_i : [q]^{T_i} \rightarrow [0, 1] \) denote the distribution

\[
\mu_i(x_{i_1}, \ldots, x_{i_k}) = \mu(x_{i_1} + a_{i_1}, \ldots, x_{i_k} + a_{i_k})
\]

so that the support of \( \mu_i \) is contained in \( C_i^{-1}(1) \). We then define the distribution \( \mathcal{P}_\mu(S) \) by picking each assignment \( \alpha \in [q]^S \) with probability proportional to \( \prod_{C_i \in \mathcal{C}(S)} \mu_i(\alpha(T_i)) \). Formally,

\[
\Pr_{\mathcal{P}_\mu(S)}[\alpha] = \frac{1}{Z_S} \cdot \prod_{C_i \in \mathcal{C}(S)} \mu_i(\alpha(T_i))
\]

(3.1)

where \( \alpha(T_i) \) is the restriction of \( \alpha \) to \( T_i \) and \( Z_S \) is a normalization factor given by

\[
Z_S = \sum_{\alpha \in [q]^S} \prod_{C_i \in \mathcal{C}(S)} \mu_i(\alpha(T_i)).
\]

To understand the distribution, it is easier to think of the special case when \( \mu \) is just the uniform distribution on \( P^{-1}(1) \) (like in the case of MAX \( \kappa \)-XOR). Then \( \mathcal{P}_\mu(S) \) is simply the uniform distribution on assignments satisfying all the constraints in \( \mathcal{C}(S) \). When \( \mu \) is not uniform, then the probabilities are weighted by the product of the values \( \mu_i(\alpha(T_i)) \) for all the constraints.\(^4\) But, crucially, we still have the property that if \( \Pr_{\mathcal{P}_\mu(S)}[\alpha] > 0 \), then \( \alpha \) satisfies all the constraints in \( \mathcal{C}(S) \).

In order for the distribution \( \mathcal{P}_\mu(S) \) to be well defined, we need to ensure that \( Z_S > 0 \). The following lemma shows how to calculate \( Z_S \) if \( G \) is sufficiently expanding, and simultaneously proves that if \( S_1 \subseteq S_2 \), and if \( G|_{-S_1} \) is sufficiently expanding, then \( \mathcal{P}_\mu(S_1) \) is consistent with \( \mathcal{P}_\mu(S_2) \) over \( S_1 \).

**Lemma 3.4.** Let \( \Psi \) be a MAX \( \kappa \)-CSP\((P) \) instance as above and \( S_1 \subseteq S_2 \) be two sets of variables such that both \( G \) and \( G|_{-S_1} \) are \((r, k-3+\delta)\) boundary expanding for some \( \delta > 0 \) and \( |\mathcal{C}(S_2)| \leq r \). Then \( Z_{S_2} = q^{|S_2|}/q^{k|\mathcal{C}(S_2)|} \), and for any \( \alpha_1 \in [q]^{S_1} \)

\[
\sum_{\substack{\alpha_2 \in [q]^{S_2} \\ \alpha_2(S_1) = \alpha_1}} \Pr_{\mathcal{P}_\mu(S_2)}[\alpha_2] = \Pr_{\mathcal{P}_\mu(S_1)}[\alpha_1].
\]

(3.2)

**Proof.** Let \( \mathcal{C} = \mathcal{C}(S_2) \setminus \mathcal{C}(S_1) \) be the set of \( t \) constraints dominated by \( S_2 \) but not \( S_1 \). Without loss of generality let \( \mathcal{C} = \{C_1, \ldots, C_t\} \) with \( C_i \) being on the set of variables \( T_i \) some of which might be set by \( \alpha_1 \). Note that any \( \alpha_2 \) consistent with \( \alpha_1 \) can be written as \( \alpha_1 \circ \alpha \) for some \( \alpha \in [q]^{S_2 \setminus S_1} \). We will show a way to calculate a sum similar to the left hand side of (3.2). Note that these calculations are meaningful even if \( Z_{S_1} \) or \( Z_{S_2} \) are zero, in which case both sides are

\(^4\)Note however that \( \mathcal{P}_\mu(S) \) is not a product distribution because different constraints in \( \mathcal{C}(S) \) may share variables.
simply zero. Taking \( S_1 = \emptyset \) will then give us the value of \( Z_{S_2} \).

\[
Z_{S_2} \cdot \sum_{\alpha_2 \in [q]^{S_2} \setminus S_1} \Pr_{\mathcal{P}_\mathcal{S}}[\alpha_2] = \sum_{\alpha \in [q]^{S_2} \setminus S_1} \prod_{C_i \in \mathcal{C}(S_2)} \mu_i((\alpha_1 \circ \alpha)(T_i)) \\
= \left( \prod_{C_i \in \mathcal{C}(S_1)} \mu_i(\alpha_1(T_i)) \right) \sum_{\alpha \in [q]^{S_2} \setminus S_1} \prod_{j=1}^t \mu_j((\alpha_1 \circ \alpha)(T_j)) \\
= (Z_{S_1} \cdot \Pr_{\mathcal{P}_\mathcal{S}}[\alpha_1]) \sum_{\alpha \in [q]^{S_2} \setminus S_1} \prod_{j=1}^t \mu_j((\alpha_1 \circ \alpha)(T_j))
\]

The following claim lets us calculate this last sum conveniently using the expansion of \( G|_{-S_1} \).

**Claim 3.5.** Let \( \mathcal{C} \) be as above. Then there exists an ordering \( C_{i_1}, \ldots, C_{i_t} \) of constraints in \( \mathcal{C} \) and a partition of \( S_2 \setminus S_1 \) into sets of variables \( F_1, \ldots, F_t \) and \( F_{t+1} \) such that for all \( j \leq t \), \( F_j \subseteq T_{i_j} \), \( |F_j| \geq k - 2 \), and

\[
\forall j \ F_j \cap \left( \cup_{l>j} T_{i_l} \right) = \emptyset.
\]

**Proof.** We build the sets \( F_j \) inductively using the fact that \( G|_{-S_1} \) is \((r, k - 3 + \delta)\) boundary expanding.

Start with the set of constraints \( C_1 = \mathcal{C} \). Since \( |C_1| = |\mathcal{C}(S_2) \setminus \mathcal{C}(S_1)| \leq r \), \( C_1 \) should be expanding in \( G_{-S_1} \) i.e. \( |\partial(C_1) \setminus S_1| \geq (k - 3 + \delta)|C_1| \). Hence, there exists \( C_j \subseteq C_1 \) contributing at least \( k - 2 \) variables to the boundary of \( C_1 \), i.e. \( |T_j \cap (\partial(C_1) \setminus S_1)| \geq k - 2 \). Let \( T_j \cap (\partial(C_1) \setminus S_1) = F_1 \) and \( i_1 = j \). We then take \( C_2 = C_1 \setminus \{C_{i_1}\} \) and continue in the same way. What is left from \( S_2 \setminus S_1 \) after the last step will be \( F_{t+1} \). It is not hard to check that \( F_{t+1} \) is the portion of \( S_2 \setminus S_1 \) that is not in any \( C_i \) although we will not use this.

Since at every step, we have \( F_j \subseteq \partial(C_j) \setminus S_1 \), and for all \( l > j \ C_l \subseteq C_j \), \( F_j \) shares no variables with \( \Gamma(C_l) \) for \( l > j \). Hence, we get \( F_j \cap \left( \cup_{l>j} T_{i_l} \right) = \emptyset \) as claimed.

Using this decomposition, we can reorder the sum and split it as,

\[
\sum_{\alpha \in [q]^{S_2} \setminus S_1} \prod_{j=1}^t \mu_j(\alpha_1 \circ \alpha(T_j)) = \sum_{\beta_{t+1} \in [q]^{F_{t+1}}} \sum_{\beta_t \in [q]^{F_t}} \mu_{t+1} \sum_{\beta_{t-1} \in [q]^{F_{t-1}}} \mu_{t} \cdots \sum_{\beta_2 \in [q]^{F_2}} \mu_{2} \sum_{\beta_1 \in [q]^{F_1}} \mu_{1}
\]

where the input to each \( \mu_{i_1} \) depends on \( \alpha_1 \) and \( \beta_j, \ldots, \beta_{t+1} \) but not on \( \beta_1, \ldots, \beta_{j-1} \).

We now reduce the expression from right to left. Since \( F_1 \) contains at least \( k - 2 \) variables and \( \mu_{i_1} \) is a balanced pairwise independent distribution,

\[
\sum_{\beta_1 \in [q]^{F_1}} \mu_{i_1} = \Pr_{\mu}[(\alpha_1 \circ \beta_2 \cdots \circ \beta_t)(T_{i_1} \setminus F_1)] = \frac{1}{q^{k-|F_1|}}
\]

irrespective of the values assigned by \( \alpha_1 \circ \beta_2 \circ \cdots \circ \beta_t \) to the remaining (at most 2) variables in
Continuing in this fashion from right to left, we get that

\[ \sum_{\alpha \in [q]^{S_2 \setminus S_1}} \prod_{j=1}^{t} \mu_i\left( (\alpha_1 \circ \alpha)(T_i) \right) = \frac{1}{q^{|S_2 \setminus S_1| - k|C(S_2) \setminus C(S_1)|}} \]

Hence, we get that

\[ Z_{S_2} \cdot \sum_{\alpha_2 \in [q]^{S_2}} \Pr_{P_\mu(S_2)}[\alpha_2] = \left( Z_{S_1} \cdot \frac{q^{|S_2 \setminus S_1| - k|C(S_2) \setminus C(S_1)|}}{q^{|K(S_2) \setminus C(S_1)|}} \right) \Pr_{P_\mu(S_1)}[\alpha_1]. \quad (3.3) \]

Now, since we know that \( G \) is \((r, k - 3 + \delta)\)-boundary expanding, we can replace \( S_1 \) by \( \emptyset \) in the above calculation to get,

\[ Z_{S_2} = \sum_{\alpha \in [q]^{S_2}} \prod_{C_i \in C(S_2)} \mu_i\left( (\alpha_1 \circ \alpha)(T_i) \right) = q^{|S_2| - k|C(S_2)|}, \]

as claimed. Plugging in the value of \( Z_{S_2} \) and \( Z_{S_1} \) into (3.3) will show that,

\[ \sum_{\alpha_2 \in [q]^{S_2}} \Pr_{P_\mu(S_2)}[\alpha_2] = \Pr_{P_\mu(S_1)}[\alpha_1], \]

which proves the lemma.

We are now almost ready to describe our Sherali-Adams SDP solution. The only other property of \( P_\mu(S) \) that we need to show is that it is pairwise-independent and balanced.

**Claim 3.6.** Let \( G \) be a \((r, k - 3 + \epsilon)\)-boundary expanding constraint graph, with \( \epsilon > 2/3 \) in which no two constraints share more than one variable. Then for any \( S \subset R, |S| \leq r \), the distribution \( P_\mu(S) \) is a pairwise-independent and balanced distribution. That is, for all \( i_1, i_2 \in S, j_1, j_2 \in [q], \)

\[ \Pr_{\alpha \sim P_\mu(S)}[\alpha_{i_1} = j_1] = 1/q, \quad \Pr_{\alpha \sim P_\mu(S)}[(\alpha_{i_1} = j_1) \land (\alpha_{i_2} = j_2)] = 1/q^2. \]

**Proof.** Let \( S \) be a subset of the variables, \( i, j \in S \) and \( \epsilon' = \min(1, \epsilon - 2/3) \). We will first show that \( G_{\{i,j\}} \) is \((r, k - 3 + \epsilon')\) boundary expanding.

Let \( C \) be a set of constraints with \( |C| \leq r \). When \( |C| = 1 \), \( C \) has \( k \) boundary neighbours in \( G \) and hence at least \( k - 2 \geq (k - 3 + \epsilon')|C| \) boundary neighbours in \( G_{\{i,j\}} \). When \( |C| = 2 \), the number of boundary neighbours in \( G \) must be at least \( 2k - 2 \) since the two constraints in \( C \) can share at most one variable. It follows that \( |\partial C| \) is at least \( 2k - 2 - 2 \geq (k - 3 + \epsilon')|C| \) in \( G_{\{i,j\}} \).

Finally, when \( 3 \leq |C| \leq r \), we get that the size of \( |\partial C| \) in \( G_{\{i,j\}} \) is at least \((k - 3 + \epsilon)|C| - 2\) as \( G \) is \((r, k - 3 + \epsilon)\) boundary expanding. This proves the claim as \((k - 3 + \epsilon)|C| - 2 = (k - 3 + \epsilon - 2/|C|)|C| \geq (k - 3 + \epsilon')|C| \).
It follows from applying Lemma 3.4 with \( S_2 = S \) and \( S_1 = \{i, j\} \) that \( P_\mu(S) \) agrees with \( P_\mu(\{i, j\}) \) on the set \( \{i, j\} \). Now note that for \( k = 2 \) the only promising predicate is the one accepting everything for which the theorem is trivial, so one can assume that \( k > 2 \) and \( C(\{i, j\}) \) is empty. It follows that \( P_\mu(\{i, j\}) \) is the uniform distribution on \( [q]^{\{i,j\}} \) which completes the proof.

### 3.3 Constructing the Integrality Gap

We now show how to construct integrality gaps using the ideas in the previous section. For a given promising predicate \( P \), our integrality gap instance will be a random instance \( \Psi \) of the \( \text{MAX } k\text{-CSP}_{q}(P) \) problem, conditioned on the event that no two constraints sharing more than one variable. To generate a random instance with \( m \) constraints, for every constraint \( C_i \), we randomly select a \( k \)-tuple of distinct variables \( T_i = \{x_{i1}, \ldots, x_{ik}\} \) and \( a_{i1}, \ldots, a_{ik} \in [q] \), and put \( C_i \equiv P(x_{i1} + a_{i1}, \ldots, x_{ik} + a_{ik}) \). It is well known and used in various works on integrality gaps and proof complexity (e.g. [BOGH+06, AAT11, STT07a, Sch08]), that random instances of CSPs are both highly unsatisfiable and highly expanding. We capture the properties we need in the lemma below. The proof uses standard arguments and is left for Section 3.4.

**Lemma 3.7.** Let \( \epsilon, \delta > 0 \) and a predicate \( P : [q]^k \to \{0, 1\} \) be given. Then there exist \( \gamma = O(q^k \log q/\epsilon^2) \), \( \eta = \Omega((1/\gamma)^{10/\delta}) \) and \( N \in \mathbb{N} \), such that if \( n \geq N \) and \( \Psi \) is a random instance of \( \text{MAX } k\text{-CSP}(P) \) with \( m = \gamma n \) constraints, then with probability \( \exp(-O(k^4 \gamma^2)) \) we have

1. \( \text{OPT}(\Psi) \leq \frac{|P^{-1}(1)|}{q^k} (1 + \epsilon) \cdot m \).

2. For any set \( C \) of constraints with \( |C| \leq \eta m \), \( |\partial(C)| \geq (k - 2 - \delta)|C| \).

3. No two constraints in \( \Psi \) share more than one variable.

Let \( \Psi \) be an instance of \( \text{MAX } k\text{-CSP}_{q} \) on \( n \) variables for which \( G_\Psi \) is \( (\eta m, k - 2 - \delta) \) boundary expanding for some \( \delta < 1/4 \), as in Lemma 3.7. For such a \( \Psi \), we now define the distributions \( \mathcal{D}(S) \). In the rest of this chapter we assume \( k \geq 3 \) as for \( k = 2 \) the only promising predicate is the one satisfying all assignments for which the theorem is trivial.

For a set \( S \) of size at most \( t = \eta m/6k \), let \( \overline{S} \) be the subset of variables output by the algorithm Advice when run with input \( S \) and parameters \( r = \eta m, \epsilon_1 = (k - 2 - \delta), \epsilon_2 = (k - 8/3 - \delta) \) on the graph \( G_\Psi \). Theorem 3.3 shows that

\[
|\overline{S}| \leq (2k - 4 - 2\delta + k - k + 8/3 + \delta)|S|/(4/3) = (6k - 4 - 3\delta)|S|/4 \leq \eta m/4,
\]

and also,

\[
\xi_S \geq \eta m - |S|/(2/3) = \eta m - \eta m/4k > 3\eta m/4.
\]
We then use (3.1) to define the distribution $D(S)$ for sets $S$ of size at most $\eta n/6k$ as

$$\Pr_{D(S)}[\alpha] = \text{Mar}_{S} P_{\mu}(S) = \sum_{\beta \in [q]^{S}} \Pr_{P_{\mu}(S)}[\beta].$$

Using the properties of the distributions $P_{\mu}(S)$, we can now prove that the distributions $D(S)$ are consistent.

**Claim 3.8.** For any two sets $S_1 \subseteq S_2 \subseteq [n]$ with $|S_2| \leq t = \eta n/6k$, the distributions $D(S_1), D(S_2)$ are equal on $S_1$.

**Proof.** The distributions $D(S_1), D(S_2)$ are defined according to $P_{\mu}(S_1)$ and $P_{\mu}(S_2)$ respectively. To prove the claim, we show that $P_{\mu}(S_1)$ and $P_{\mu}(S_2)$ are equal to the distribution $P_{\mu}(S_1 \cup S_2)$ on $S_1, S_2$ respectively (note that it need not be the case that $S_1 \subseteq S_2$).

Let $S_3 = S_1 \cup S_2$. Since $|S_1|, |S_2| \leq \eta n/4$, we have $|S_3| \leq \eta n/2$. We will first show that $|\partial C|/|C| \leq 3\eta n/4$. Assume to the contrary, and let $C$ be any subset of $C(S_3)$ of size $3\eta n/4$. Given that $k \geq 3$ and $\delta < 1/4$ we would have the following bound on the boundary of $C$,

$$|\partial C|/|C| \leq |\Gamma(C)|/|C| \leq |S_3|/|C| \leq (1/2)/(3/4) = 2/3 < k - 2 - \delta,$$

which would contradict boundary expansion of $G$.

Now, by Theorem 3.3, we know that both $G|_{S_1}$ and $G|_{S_2}$ are $(3\eta n/4, k - 8/3 - \delta)$ boundary expanding. Thus, using Lemma 3.4 for the pairs $S_1 \subseteq S_3$ and $S_2 \subseteq S_3$, we get that

$$\Pr_{D(S_1)}[\alpha_1] = \sum_{\beta_1 \in [q]^{S_1}, \beta_1(S_1) = \alpha_1} \Pr_{P_{\mu}(S_1)}[\beta_1] = \sum_{\beta_2 \in [q]^{S_2}, \beta_2(S_1) = \alpha_1} \Pr_{P_{\mu}(S_2)}[\beta_2] = \Pr_{D(S_2)}[\alpha_2]$$

which shows that $D(S_1)$ and $D(S_2)$ are equal on $S_1$.

It is now easy to prove the main result.

**Theorem 3.9.** Let $P : [q]^k \to \{0, 1\}$ be a promising predicate. Then for every constant $\zeta > 0$, there exist $c = c(q, k, \zeta)$, such that for large enough $n$, the integrality gap of MAX $k$-CSP($P$) for the program obtained by $cn$ levels of the Sherali-Adams SDP hierarchy is at least $\frac{q^k}{|P^{-1}(1)|} - \zeta$.

**Proof.** We take $\epsilon = \zeta/q^k$, $\delta = 1/4$ and consider a random instance $\Psi$ of MAX $k$-CSP($P$) with $m = \gamma n$ as given by Lemma 3.7. From that lemma $\text{OPT}(\Psi) \leq \frac{|P^{-1}(1)|}{q^k}(1 + \epsilon) \cdot m$.

On the other hand, by Claim 3.8 we can define distributions $D(S)$ over every set of at most $\eta n/6k$ variables such that for $S_1 \subseteq S_2$, $D(S_1)$ and $D(S_2)$ are consistent over $S_1$. Also, Claim 3.6
implies that $\mathcal{P}_\mu(S)$ hence $\mathcal{D}(S)$ is pairwise-independent and balanced for any $S$. We can now construct the SDP vectors with inner products according to the distributions $\mathcal{D}(S) = \mathcal{P}_\mu(S)$ for an instance satisfying the above property using the following simple fact.

**Claim 3.10.** There exists vectors $\{v_{i,j}\}_{i \in [n], j \in [q]}$ and $v_0$ satisfying:

1. For all $i \in [n]$ and $j_1 \neq j_2$, $v_{i,j_1} \cdot v_{i,j_2} = 0$.
2. For all $i \in [n], j \in [q]$, $v_{i,j} \cdot v_0 = \|v_{i,j}\|^2 = 1/q$.
3. For all $i_1 \neq i_2 \in [n]$ and $j_1, j_2 \in [q]$, $v_{i_1,j_1} \cdot v_{i_2,j_2} = 1/q^2$.

**Proof.** Let $e_1, \ldots, e_n$ be an orthonormal basis for $\mathbb{R}^n$ and $u_0, \ldots, u_{q-1}$ be vertices of a $q - 1$ dimensional simplex satisfying $u_{j_1} \cdot u_{j_2} = -1/(q - 1)$ when $j_1 \neq j_2$ and 1 otherwise. Let $v_0 \in \mathbb{R}^{n(q-1)+1}$ be a unit vector such that $v_0 \cdot e_i \otimes u_j = 0$ for all $i \in [n]$ and $j \in [q]$. We then define $v_{i,j}$ as

$$v_{i,j} := \frac{1}{q} v_0 + \frac{\sqrt{q-1}}{q} (e_i \otimes u_j)$$

It is easy to check that $v_0 \cdot v_{i_1,j_1} = 1/q$ and $v_{i_1,j_1} \cdot v_{i_2,j_2} = 1/q^2$ for all $i_1 \neq i_2 \in [n]$ and $j_1, j_2 \in [q]$. Also, for $j_1 \neq j_2$

$$v_{i_1,j_1} \cdot v_{i_2,j_2} = \frac{1}{q^2} + \frac{q-1}{q^2} \cdot \left(\frac{-1}{q-1}\right) = 0$$

which proves the claim. 

By Fact 2.2 this gives a feasible solution to the level $\eta n/6k$ Sherali-Adams SDP relaxation\(^5\) of MAX $k$-CSP\(_q\)(P). Also, by definition of $\mathcal{D}(S)$, we have that $\Pr_{\mathcal{D}(S)}[\alpha] > 0$ only if $\alpha$ satisfies all constraints in $\mathcal{C}(S) \supseteq \mathcal{C}(S)$. Hence, the value of FRAC($\Psi$) is given by

$$\sum_{i=1}^{m} \sum_{\alpha \in [q]^{T_i}} C_i(\alpha) X_{(T_i, \alpha)} = \sum_{i=1}^{m} \sum_{\alpha \in [q]^{T_i}} C_i(\alpha) \Pr_{\mathcal{D}(T_i)}[\alpha] = \sum_{i=1}^{m} \sum_{\alpha \in [q]^{T_i}} \Pr_{\mathcal{D}(T_i)}[\alpha] = m.$$

Thus, the integrality gap after $\eta n/6k$ levels is at least

$$\frac{\text{FRAC}(\Psi)}{\text{OPT}(\Psi)} = \frac{q^k}{|P^{-1}(1)|/(1+\epsilon)} \geq \frac{q^k}{|P^{-1}(1)|} - \zeta.$$

### 3.4 Proof of Lemma 3.7

We prove Lemma 3.7, restated below for convenience.

\(^5\)to be precise the level is $\eta n/6k - O(k)$ but given that $n$ is growing and $k$ is a constant this does not matter.
Lemma 3.7 (restated). Let $\epsilon, \delta > 0$ and a predicate $P : [q]^k \to \{0,1\}$ be given. Then there exist $\gamma = O(q^k \log q/\epsilon^2)$, $\eta = \Omega((1/\gamma)^{10/\delta})$ and $N \in \mathbb{N}$, such that if $n \geq N$ and $\Psi$ is a random instance of MAX $k$-CSP ($P$) with $m = \gamma n$ constraints, then with probability $\exp(-O(k^4 \cdot \gamma^2))$

1. $\text{OPT}(\Psi) \leq \frac{|P^{-1}(1)|}{q^k} (1 + \epsilon) \cdot m$.

2. For any set $\mathcal{C}$ of constraints with $|\mathcal{C}| \leq \eta m$, we have $|\partial(\mathcal{C})| \geq (k - 2 - \delta)|\mathcal{C}|$.

3. No two constraints in $\Psi$ share more than one variable.

Proof. Let $\alpha \in [q]^n$ be any fixed assignment. For a fixed $\alpha$, the events that a constraint $C_i$ is satisfied are independent and happen with probability $|P^{-1}(1)|/q^k$ each. Hence, by Chernoff bound the probability over the choice of $\Psi$ that $\alpha$ satisfies more than $|P^{-1}(1)|/(1 + \epsilon) \cdot \gamma n$ constraints is at most $\exp(-\epsilon^2 \gamma n |P^{-1}(1)|/3q^k)$. By a union bound, the probability that any assignment satisfies more than $|P^{-1}(1)|/(1 + \epsilon) \cdot \gamma n$ constraints is at most $q^n \cdot \exp(\frac{-\epsilon^2 \gamma n |P^{-1}(1)|}{3q^k}) = \exp(n \ln q - \frac{\epsilon^2 \gamma n |P^{-1}(1)|}{3q^k})$ which is $o(1)$ for $\gamma = \frac{6q^k \ln q}{\epsilon^2}$.

For showing the boundary expansion, we note that it suffices to show that the constraints have large expansion i.e. each set of $s$ constraints (for $s \leq \eta m$) contains at least $(k - 1 - \delta/2)s$ variables. Since each non-boundary variable occurs in at least two constraints, we get that that the number of boundary variables must be at least $2(k - 1 - \delta/2)s - ks = (k - 2 - \delta)s$.

To show this, consider the probability that a set of $s$ constraints contains at most $cs$ variables, where $c = k - 1 - \delta$. This is upper bounded by

$$\left(\frac{n}{cs}\right) \cdot \left(\frac{(cs)}{s}\right) \cdot s! \left(\frac{\gamma n}{s}\right) \cdot \left(\frac{n}{k}\right)^{-s}$$

Here $\left(\frac{n}{cs}\right)$ is the number of ways of choosing the $cs$ variables involved, $\left(\frac{(cs)}{s}\right)$ is the number of ways of picking $s$ tuples out of all possible $k$-tuples on $cs$ variables and $s! \left(\frac{\gamma n}{s}\right)$ is the number of ways of selecting the $s$ constraints. The number $\left(\frac{n}{k}\right)^s$ is simply the number of ways of picking $s$ of these $k$-tuples in an unconstrained way. Using $\left(\frac{a}{b}\right)^b \leq \left(\frac{a}{b}\right)^a \leq \left(\frac{a+1}{b+1}\right)^b$, $s! \leq s^a$ and collecting terms, we can bound this expression by

$$\left(\frac{s}{n}\right)^{\delta s/2} \left(e^{2k+1-\delta/2} k^{1+\delta/2} \gamma\right)^s \leq \left(\frac{s}{n}\right)^{\delta s/2} \gamma^s = \left(\frac{s \gamma^{10/\delta}}{n}\right)^{\delta s/2}$$

We need to show that the probability that a set of $s$ constraints contains less than $cs$ variables for any $s \leq \eta m$ is $o(1)$. Thus, we sum this probability over all $s \leq \eta m$ to get

$$\sum_{s=1}^{\eta m} \left(\frac{s \gamma^{10/\delta}}{n}\right)^{\delta s/2} = \sum_{s=1}^{\ln^2 n} \left(\frac{s \gamma^{10/\delta}}{n}\right)^{\delta s/2} + \sum_{s=\ln^2 n+1}^{\eta m} \left(\frac{s \gamma^{10/\delta}}{n}\right)^{\delta s/2} \leq O\left(\frac{\gamma^{10/\delta}}{n^{\delta/2}} \ln^2 n\right) + O\left(\left(\frac{\eta \gamma^{10/\delta}}{(\delta/2) \ln^2 n}\right)^{\delta s/2}\right)$$
The first term is $o(1)$ and is small for large $n$. The second term is also $o(1)$ for $\eta = 1/(100\gamma^{10/\delta})$.

Thus, we get the first two properties with probability $1 - o(1)$. Finally, the probability that there are no two constraints sharing two variables must be at least $\prod_{i=1,\ldots,m}(1 - O(i \cdot k^4/n^2))$ because when we choose the $i$th constraint, by wanting it to not share two variables with another previously chosen constraint, we are forbidding any of the $\binom{k}{2}$ pairs of variables in the $i$th constraint from being equal to any of the $(i - 1) \cdot \binom{k}{2}$ pairs in the previously chosen ones. Now using that for small enough $x$, $1 - x > \exp(-2x)$, the probability is at least $\exp(-O((\sum_{i=1,\ldots,m} i \cdot k^4)/n^2)) = \exp(-O((k^4 \cdot m^2)/n^2)) = \exp(-O(k^4\gamma^2))$. ■
Chapter 4

On Quadratic Threshold CSPs

In Chapter 3 we showed that if a predicate $P : \{q\}^k \rightarrow \{0, 1\}$ is promising, then the level-$\Theta(n)$ Sherali-Adams SDP relaxation of MAX $\kappa$-CSP$_q(P)$ has integrality gap $q^k / |P^{-1}(1)|$. We also mentioned that Austrin and Mossel [AM09] show that if $P$ is promising then assuming UGC MAX $\kappa$-CSP$_q(P)$ is hard to approximate within a factor better than $q^k / |P^{-1}(1)|$, i.e. $P$ is approximation resistant. In this chapter we show a weak converse of these two results for $q = 2$.

In particular, we show that for a natural subclass of binary predicates $P :: \{-1, 1\}^k \rightarrow \{-1, 1\}$ which are not promising MAX $\kappa$-CSP$_q(P)$ can be approximated to within a factor better than $2^k / |P^{-1}(1)|$, i.e. these predicates are not approximation resistant.

It is worth mentioning that our algorithms for MAX $\kappa$-CSP$_q(P)$ for these predicates use the standard SDP relaxation of the problem (see Figure 2.3(b)), so we are also proving an upper bound on the integrality gap of this relaxation. As mentioned in Section 2.2 the standard SDP relaxation of MAX $\kappa$-CSP$_q(P)$ incorporates the idea of local distributions from the Sherali-Adams hierarchy. In fact the analysis of our algorithm heavily relies on the existence of these distributions.

Our specific interest is predicates of the form $P(x) = \frac{1+\text{sign}(Q(x))}{2}$ where $Q : \mathbb{R}^k \rightarrow \mathbb{R}$ is a quadratic polynomial with no constant term$^2$, i.e., $Q(x) = \sum_{i=1}^{k} a_i x_i + \sum_{i\neq j} b_{ij} x_i x_j$ for some set of coefficients $a_1, \ldots, a_n$ and $b_{11}, \ldots, b_{nn}$. Notice that if a predicate $P$ that is defined as the sign of a quadratic polynomial $Q$ in this fashion it is not promising. This is simply because if $P$ is promising supported by a balanced, pairwise independent distribution $\mu$ then, $E_{x \sim \mu}[Q(x)] = 0$ which implies that $Pr[x \sim \mu]\text{sign}(Q(x)) = -1 > 0$ which is a contradiction. In fact, Austrin and Håstad [AH11] show that signs of quadratic forms are precisely the “minimal non-promising” predicate.

Theorem 4.1 (Theorem 3.2 in [AH11]). A predicate $P : \{-1, 1\}^k \rightarrow \{0, 1\}$ is not promising if and only if there exists a quadratic polynomial $Q : \{-1, 1\}^k \rightarrow \mathbb{R}$ with no constant term that is positive on support of $P$, i.e. $P^{-1}(1)$. In other words, $P$ implies a predicate $P'$ of the form

\footnote{Results in this chapter appear in [ABM10]}
Chapter 4. On Quadratic Threshold CSPs

1 + \text{sign}(Q) \over 2.

Theorem 4.1 is our main motivation for studying these predicates.

Given that the main tool for approximation algorithms, semidefinite programming, works by optimizing quadratic forms, it seemed natural and intuitive to hope that predicates of this form are always approximable. This however turns out to be false; Håstad [Hås09] constructs a predicate that is the sign of a quadratic polynomial and still approximation resistant. Loosely speaking, the main crux is that semidefinite programming is good for optimizing the degree-2 part of the Fourier expansion of a predicate, which Unfortunately can behave very differently from $P$ itself or the quadratic polynomial used to define $P$ (we elaborate on this below.) However, it turns out that when we restrict our attention to the special case of symmetric predicates, this cannot happen, and we can obtain an approximation algorithm, which is the first result of this chapter.

**Theorem 4.2.** Let $P : \{-1,1\}^k \rightarrow \{0,1\}$ be a predicate that is of the form $P(x) = \frac{1 + \text{sign}(Q(x))}{2}$ where $Q$ is a symmetric quadratic polynomial with no constant term. Then $P$ is not approximation resistant.

A very natural special case of the signs of quadratic polynomials is the case when $P(x) = \frac{1 + \text{sign} \left( \sum a_i x_i \right)}{2}$ is simply the sign of a linear form, i.e., a linear threshold function. While we cannot prove that linear threshold predicates are approximable in general, we do believe this is the case, and make the following conjecture.

**Conjecture 4.3.** Let $P : \{-1,1\}^k \rightarrow \{0,1\}$ be a predicate that is a sign of a linear form with no constant term. Then $P$ is not approximation resistant.

We view the resolution of this conjecture as a very natural and interesting open problem. As in the quadratic case, the difficulty stems from the fact that the low-degree part of $P$ can be unrelated to the linear form used to define $P$. Specifically, it can be the case that the low-degree part of the arithmetization of $P$ vanishes or becomes negative for some inputs where the linear/quadratic polynomial is positive (i.e. accepting inputs), and unfortunately this seems to make the standard SDP approach fail. The perhaps most extreme case of this phenomenon is exhibited by the predicate Monarchy : $\{-1,1\}^k \rightarrow \{0,1\}$ suggested by Håstad [Hås09], in which the first variable (the “monarch”) decides the outcome, unless all the other variables unite against it. In other words,

$$
\text{Monarchy}(x) = \frac{1 + \text{sign}((k-2)x_1 + \sum_{i=2}^{k} x_i)}{2}.
$$

(4.1)

Observe that for the input $x_1 = -1, x_2 = \ldots = x_k = 1$, the linear part of the Fourier expansion of Monarchy takes value $-1 + o_k(1)$, whereas the linear form used to define monarchy is positive, hence the value of the predicate is 1. Again, we stress that this means that known algorithms and techniques do not apply. However, in this case we are still able to achieve an approximation algorithm, which is our second result.
**Theorem 4.4.** The predicate Monarchy is not approximation resistant.

This shows that there is some hope in overcoming the apparent barriers to proving Conjecture 4.3.

A recent related work of Cheraghchi et al. [CHIS12] also studies the approximability of predicates defined by linear forms. However, their focus is on establishing precise quantitative bounds on the approximability, rather than the more qualitative distinction between approximable and approximation resistant predicates. As such, their results only apply to predicates which were already known to be approximable. Specifically, they consider “Majority-like” predicates $P$ where the linear part of the Fourier expansion of $P$ behaves similarly to $P$ itself in the sense explained above.

**Techniques**

The starting point of both our algorithms is the standard SDP relaxation of $\text{MAX } k\text{-CSP}(P)$. The main difficulty in rounding the solutions of these SDPs is that current rounding algorithms only offer very local analysis; i.e. they offer no way to analyze the joint distribution of the outcome of the rounding for $k$ variables, when $k > 2$.

Unfortunately, arguing about the outcome of $k$ variables together seems essential to understanding the performance of the algorithm for $\text{MAX } k\text{-CSP}(P)$ as each constraint depends on $k$ variables. Indeed, even a local argument would have to argue about the outcome of the rounding algorithm for $k$ variables together.

To prove Theorem 4.2, we give a new, simpler proof of a theorem of Hast [Has05a], giving a general condition on the low-degree part of the Fourier expansion of $P$ which guarantees a predicate is approximable (Theorem 4.6). We then show that this condition holds for predicates which are defined by symmetric quadratic polynomials. The basic idea behind our new algorithm is as follows. First, observe that the SDP solution in which all vectors are perpendicular is easy to analyze when the usual hyperplane rounding is employed, as in this case the obtained integral values are distributed uniformly. This motivates the following approach: start with the perpendicular configuration and then perturb the vectors in the direction of the optimal SDP solution. This perturbation acts as a differentiation operator, and as such allows for a “linear snapshot” of what is typically a complicated system. For each clause we analyze the probability that hyperplane rounding outputs a satisfying assignment, as a function of the inner products of vectors involved. Now, the object of interest is the gradient of this function at “zero”. The hope is that since the optimal SDP solution (almost) satisfies most clauses, it has a positive inner product with the gradient, and so can act as a global recipe that works for all clauses. It is important to stress that since we are only concerned with getting an approximation algorithm that works slightly better than random we can get away with this linear simplification. We show that this condition on the gradient translates into a condition on the low-degree part of the Fourier expansion of the predicate.

---

3Interestingly, when some modest value $k = 3$ is used, often some numerical methods are employed to complete the analysis; see [KZ97, Zwi98].
As it turns out, the predicate Monarchy which we tackle in Theorem 4.4 does not exhibit the aforementioned desirable property. In other words, the gradient above does not generally have a positive inner product with an optimal SDP solution. Instead, we show that when all vectors are sufficiently far from \( \pm v_0 \) it is possible to get a similar guarantee on the gradient using high (but not too high) moments of the vectors. We can then handle vectors which are very close to \( \pm v_0 \) separately by rounding them deterministically to \( \pm 1 \).

**Organization**

The rest of this chapter is organized as follows. Firstly, we point the reader to Chapter 2 for a formal definition of the MAX \( k \)-CSP\((P)\) problem and its standard SDP relaxation. In Section 4.1 we give our new algorithm for rounding solutions of this SDP relaxation and characterize the predicates for which it gives an approximation ratio better than a random assignment. We leave the proof of a technical lemma needed there for Section 4.4. We then take a closer look at signs of symmetric quadratic forms in Section 4.2 and show that these satisfy the condition of the previous section, proving Theorem 4.2. In Section 4.3 we give the approximation algorithm for the Monarchy predicate and the ideas behind its somewhat tedious analysis.

### 4.1 \((\epsilon, \eta)\)-hyperplane rounding

In this section we define a rounding scheme for the semidefinite program of MAX \( k \)-CSP\((P)\) and proceed to analyze its performance. The rounding scheme is based on the usual hyperplane rounding but is more flexible in that it uses two parameters \( \epsilon \geq 0 \) and \( \eta \) where \( \epsilon \) is a sufficiently small constant and \( \eta \) is an arbitrary real number. We will then formalize a (sufficient) condition involving \( P \) and \( \eta \) under which our approximation algorithm has approximation factor better than that of a random assignment. In the next section we show that this condition is satisfied (for some \( \eta \)) by signs of symmetric quadratic polynomials.

Given an instance of MAX \( k \)-CSP\((P)\), our algorithm first solves the standard SDP relaxation of the problem see in Figure 2.3(a). It then employs the rounding scheme outlined below as Algorithm 3 to get an integral solution.

Note that when \( \epsilon = 0 \) the rounding scheme above simply to assigns all \( x_i \)'s uniformly and independently at random which satisfies a \( \frac{|P^{-1}(1)|}{2^k} \) fraction of all clauses in expectation. For non-zero \( \epsilon \), \( \eta \) will determine how much weight is given to the position of \( v_0 \) compared to the correlation of the variables.

Notice that in the pursuit of a rounding algorithm that has approximation ratio better than \( \frac{2^k}{|P^{-1}(1)|} \) it is possible to assume that the optimal integral solution is arbitrary close to 1 as otherwise random assignment already delivers an approximation factor better than \( \frac{2^k}{|P^{-1}(1)|} \). In particular, the optimal vector solution can be assumed to be that good. This observation
Theorem 4.5. For any fixed $\eta$, the probability that $P(x_1, \ldots, x_k)$ is satisfied by the assignment behaves as follows at $\epsilon = 0$:

$$\Pr[(x_1, \ldots, x_k) \in P^{-1}(1)] = \frac{|P^{-1}(1)|}{2^k}$$

$$\frac{d}{d\epsilon} \Pr[(x_1, \ldots, x_k) \in P^{-1}(1)] = \frac{2\eta}{\sqrt{2\pi}} \sum_{i=1}^k \hat{P}(\{i\})v_i \cdot v_i + \frac{2}{\pi} \sum_{i<j} \hat{P}(\{i, j\})v_i \cdot v_j.$$  \hspace{1cm} (4.2)

Proof. The first claim follows from the definitions as we discussed. To prove (4.2), we first introduce some notation. For an assignment $\omega \in \{-1, 1\}^k$ define the function $p_\omega$ on $(k+1) \times (k+1)$ semidefinite matrices as follows. For a positive semidefinite matrix $A_{(k+1) \times (k+1)}$, consider a set of vectors $w_0, \ldots, w_k$ whose Gram matrix is $A$, and consider running steps 2-6 of Algorithm 3 on $w_i$'s instead of $w_i$'s. Define $p_\omega(A)$ as the probability that the rounding procedure outputs $\omega$. In what follows, we will use $A^* = A^*(\epsilon)$ to denote the Gram matrix of

\footnote{In general, without loss of generality we can assume that the current clause is on $k$ \textit{variables} as opposed to $k$ \textit{literals}. This is simply because the rounding algorithm we are considering is symmetric, i.e. if $v_j = -v_i$ then the distribution of $x_j$ is the same as the distribution of $-x_i$.}
the set of vectors $w_1, \ldots, w_n$ used by the algorithm which depends on $\epsilon$. Clearly, we have,

$$\Pr[(x_1, \ldots, x_k) \in P^{-1}(1)] = \sum_{\omega \in P^{-1}(1)} \Pr[(x_1, \ldots, x_k) = \omega] = \sum_{\omega \in P^{-1}(1)} p_\omega(A^*)$$

We start by computing $\frac{d}{d\epsilon} p_\omega(A^*)$ using the chain rule.

$$\frac{d}{d\epsilon} p_\omega(A^*(\epsilon)) = \sum_{i<j} \left( \frac{d}{d\epsilon} a_{ij}^* \bigg|_{\epsilon=0} \cdot \frac{\partial}{\partial a_{ij}} p_\omega(A) \right) = \sum_{i<j} v_i \cdot v_j \frac{\partial}{\partial a_{ij}} p_\omega(A) \bigg|_{A=I}, \quad (4.3)$$

where, when we talk about $\frac{\partial}{\partial a_{ij}} p_\omega(A)$ we consider $A$ a symmetric positive semidefinite matrix so $a_{ij}$ changes with $a_{ji}$. Now to compute $\frac{\partial}{\partial a_{ij}} p_\omega(A) \bigg|_{A=I}$ we compute a formula for $p_\omega(A)$ where $A$ is equal to $I$ in every entry except the $ij$ and $ji$ entries where it is $a_{ij}$. Define $J_{ij}$ as the matrix which is zero on every coordinate except coordinates $ij$ and $ji$ where it is 1. Then,

$$\frac{\partial}{\partial a_{ij}} p_\omega(A) \bigg|_{A=I} = \frac{d}{dt} p_\omega(I + tJ_{ij}) \bigg|_{t=0}. \quad (4.3)$$

Note that for $t \in [-1, 1]$, $I + tJ$ is positive semidefinite, so $p_\omega(I + tJ)$ is well-defined. Now observe that in the case of $p_\omega(I + tJ_{ij})$ the geometric realization of $w_0, \ldots, w'_k$ is simple; In particular, all the vectors are perpendicular except the pair $w'_i$ and $w'_j$, so $p_\omega(I + tJ_{ij})$ is easy to compute in this case. Depending on if one of $i$ and $j$ are zero we have two cases.

First consider $i = 0$; the value of all variables will be assigned independently, and all but $x_j$ will be assigned uniformly random values. It is easy to see that for any $j \in [k]$ and any $\omega \in \{-1, 1\}^k$,

$$p_\omega(I + tJ_{0j}) = 2^{-(k-1)} \cdot \begin{cases} \Pr[w'_j \cdot g \geq -\eta t] & \text{if } \omega_j = 1 \\ \Pr[w'_j \cdot g \leq -\eta t] & \text{if } \omega_j = -1 \end{cases} \begin{cases} 1 - \Phi(-\eta t) & \omega_j = 1 \\ \Phi(-\eta t) & \omega_j = -1 \end{cases} = 2^{-(k-1)} \cdot \left(1 + \omega_j - \omega_j \Phi(-\eta t)\right).$$

Differentiating, we get

$$\frac{d}{dt} p_\omega(I + tJ_{0j}) = -2^{-(k-1)} \omega_j \frac{d}{dt} \Phi(-\eta t) = 2^{-(k-1)} \eta \omega_j e^{-\eta^2 t^2 / 2} \frac{2^{-(k-1)} \eta \omega_j e^{-\eta^2 t^2 / 2}}{\sqrt{2\pi}},$$

from which we get the identity

$$\frac{\partial}{\partial a_{0j}} p_\omega(A) \bigg|_{A=I} = \frac{2^{-(k-1)} \eta \omega_j}{\sqrt{2\pi}}. \quad (4.4)$$
Let us then consider the case where both \( i \) and \( j \) are non-zero. In this case, each variable is going to be assigned a value in \( \{-1, 1\} \) uniformly at random, and all these assignments are independent except the assignments of the \( i \)th and the \( j \)th variable. In particular, a simple analysis similar to that of [GW95] shows that for all \( i \neq j \in [k] \),

\[
p_\omega(I + tJ_{ij}) = 2^{-(k-1)} \cdot \begin{cases} 
\frac{1}{\pi} \arccos t 
& \omega_i \neq \omega_j \\
1 - \frac{1}{\pi} \arccos t 
& \omega_i = \omega_j 
\end{cases} 
- \frac{1}{\pi} \omega_i \omega_j \arccos t 
\]

Differentiating this expression, we have

\[
\frac{d}{dt} p_\omega(I + tJ_{ij}) = -\frac{2^{-(k-1)} \omega_i \omega_j}{\pi} \cdot \frac{d}{dt} \arccos t = \frac{2^{-(k-1)} \omega_i \omega_j}{\pi \sqrt{1 - t^2}} 
\]

and we can conclude that

\[
\frac{\partial}{\partial a_{ij}} p_\omega(A) \bigg|_{A=I} = 2^{-(k-1)} \frac{\omega_i \omega_j}{\pi}. 
\]

Now combining (4.3) with (4.4) and (4.5) we get,

\[
\frac{d}{de} \Pr[(x_1, \ldots, x_k) \in P^{-1}(1)] = \sum_{\omega \in P^{-1}(1)} \sum_{i < j} v_i \cdot v_j \frac{\partial}{\partial a_{ij}} p_\omega(A) \bigg|_{A=I} 
\]

\[
= \sum_{1 \leq i \leq k} v_0 \cdot v_i \frac{2^{-(k-1)} \eta}{\sqrt{2\pi}} \sum_{\omega \in P^{-1}(1)} \omega_i + \sum_{1 \leq i < j \leq k} v_i \cdot v_j \frac{2^{-(k-1)}}{\pi} \sum_{\omega \in P^{-1}(1)} \omega_i \omega_j 
\]

\[
= \sum_{1 \leq i \leq k} v_0 \cdot v_i \frac{2\eta}{\sqrt{2\pi}} \mathbb{E}[\omega_i P(\omega)] + \sum_{1 \leq i < j \leq k} v_i \cdot v_j \frac{2}{\pi} \mathbb{E}[\omega_i \omega_j P(\omega)] 
\]

\[
= \frac{2\eta}{\sqrt{2\pi}} \sum_{i=1}^{k} \hat{P}([i]) v_0 \cdot v_i + \frac{2}{\pi} \sum_{i < j} \hat{P}([i, j]) v_i \cdot v_j. 
\]

Which completes the proof.

Now, the inner products \( v_i \cdot v_j \) are equal to the moments of the local distributions \( I_{(i,j),*} \), which in turn agree with those of the local distribution \( I_{[k],*} \). It follows that,

\[
\frac{2\eta}{\sqrt{2\pi}} \sum_{i=1}^{k} \hat{P}([i]) v_0 \cdot v_i + \frac{2}{\pi} \sum_{i < j} \hat{P}([i, j]) v_i \cdot v_j = \mathbb{E}_{\omega \sim I_{[k],*}} \left[ \frac{2\eta}{\sqrt{2\pi}} P_{=1}(\omega) + \frac{2}{\pi} P_{=2}(\omega) \right]. 
\]

Thus, in order for the derivative in (4.2) to be positive for all possible values of the \( v_i \)'s that have SDP objective value 1, it is necessary and sufficient that \( \frac{2\eta}{\sqrt{2\pi}} P_{=1}(\omega) + \frac{2}{\pi} P_{=2}(\omega) \) is positive for every \( \omega \in P^{-1}(1) \). This leads us to the following theorem formulating a condition under which our rounding algorithm works.
Theorem 4.6. Suppose that there exists an \( \eta \in \mathbb{R} \) such that
\[
\frac{2\eta}{\sqrt{2\pi}} P^{=1}(\omega) + \frac{2}{\pi} P^{=2}(\omega) > 0
\]
for every \( \omega \in P^{-1}(1) \). Then \( P \) is approximable.

As mentioned in the Techniques section, this theorem is not new. It was previously found by Hast [Has05]. However, his algorithm and analysis are completely different from ours using different algorithms to optimize the linear and quadratic parts of the predicate, and case analysis depending on the behaviour of the integral solution. We believe that our algorithm is considerably more direct, and its analysis is simpler.

The general strategy for the proof, which can be found below, is as follows. We will concentrate on a clause that is almost satisfied by the SDP solution. By Equation 4.7 and Theorem 4.5 the first derivative of the probability that this clause is satisfied by the rounded solution is at least some positive global constant (say \( \delta \)) at \( \epsilon = 0 \). We will then show that for small enough \( \epsilon \) the second derivative of this probability is bounded in absolute value by, say, \( \Gamma \) at any point in \([0, \epsilon]\). Now we can apply Taylor’s theorem to show that if \( \epsilon \) is small enough the probability of success is at least
\[
|P^{-1}(1)|^2 \kappa^2 + \frac{\delta \epsilon}{1 - \delta / \Gamma}
\]
which for \( \epsilon = \delta / \Gamma \) is at least
\[
|P^{-1}(1)|^2 \kappa^2 + \frac{\delta^2}{2 \Gamma}.
\]

Proof of Theorem 4.6. Consider the optimal vector solution \( v_0, \ldots, v_n \in S^n \). Note that the optimal integral solution will have objective value less than or equal to that of \( v_0, \ldots, v_n \). So if we fix a constant \( \delta_2 \), we can always assume that the vector solution achieves objective value at least \( 1 - \delta_2 \). Otherwise, a random assignment to the variables will achieve an objective value of \( |P^{-1}(1)|^2 / 2^k \) and approximation factor \((1 - \delta_2) / |P^{-1}(1)|^2 = 2^k / |P^{-1}(1)|^2(1 - \delta_2) \). So, in this case even a random assignment shows that the predicate is not approximation resistant. From here on we assume that the vector solution achieves an objective value of \( 1 - \delta_2 \), where \( \delta_2 \) is some constant to be set later. Now, applying a simple Markov type inequality one can see that at least a \( 1 - \sqrt{\delta_2} \) fraction of the clauses must have SDP value at least \( |P^{-1}(1)|^2 + \delta \epsilon \Gamma^2 / 2 \) which for \( \epsilon = \delta / \Gamma \) is at least \( |P^{-1}(1)|^2 + \delta^2 / 2 \Gamma \).

5In particular, if \( \delta_2 \leq \delta_1^2 / 4(\delta_1 - \min(s, 0))^2 \)
that the first derivative (at \( \epsilon = 0 \)) of the probability that \( P \) is satisfied by the rounded solution is bounded from below by the constant \( \delta_1/2 \).

All that remains is to show that the second derivative of this probability cannot be too large in absolute value. We will need the following lemma about the second derivative of the orthant probability of normal random variables.

**Lemma 4.7.** For fixed \( k \), define the function \( \text{ort}(\nu, \Sigma) \) as the orthant probability of the multivariate normal distribution with mean \( \nu \) and covariance matrix \( \Sigma \), where \( \nu \in \mathbb{R}^k \) and \( \Sigma \) is a positive semidefinite matrix. That is,

\[
\text{ort}(\nu, \Sigma) \overset{\text{def}}{=} \Pr_{x \sim N(\nu, \Sigma)} \left[ \forall i \in [k] \ x_i \geq 0 \right].
\]

There exists a global constant \( \Gamma \) that upper bounds all the second partial derivatives of \( \text{ort}(\cdot) \) when \( \Sigma \) is close to \( I \). In particular, for all \( i_1, j_1, i_2, j_2 \in [k] \), all vectors \( \nu \in \mathbb{R}^k \) and all positive definite matrices \( \Sigma \) satisfying

\[
|I - \Sigma|_\infty, |\nu|_\infty < \kappa,
\]

we have,

\[
\left| \frac{\partial^2}{\partial \Sigma_{i_1 j_1} \partial \Sigma_{i_2 j_2}} \text{ort}(\nu, \Sigma) \right| < \Gamma, \quad \left| \frac{\partial^2}{\partial \Sigma_{i_1 j_1} \partial \nu_{i_2}} \text{ort}(\nu, \Sigma) \right| < \Gamma, \quad \left| \frac{\partial^2}{\partial \nu_{i_1} \partial \nu_{i_2}} \text{ort}(\nu, \Sigma) \right| < \Gamma.
\]

The proof of this lemma is rather technical and is left for Section 4.4, but the general outline is as follows. First we write down the orthant probability as an integral of the probability density function over the positive orthant. Then we observe that each of the inner integrals as well as the probability density function and their partial derivatives are continuous, so we can apply Leibniz’s integral rule iteratively to move the differentiation under the integral. We then differentiate the probability density function and the result will be in the form of the expectation of a degree 2 expression in \( x_i \)’s under the same distribution. We can then bound this expression in terms of the means and correlations of the variables.

Now, similar to the proof of Theorem 4.5 we can write,

\[
\frac{d^2}{d\epsilon^2} \Pr[(x_1, \ldots, x_k) \in P^{-1}(1)] = \frac{d}{d\epsilon} \sum_{\omega \in P^{-1}(1)} \sum_{0 \leq i < j \leq k} v_i \cdot v_j \frac{\partial}{\partial a_{ij}} p_{\omega}(A)
\]

\[
= \sum_{\omega \in P^{-1}(1)} \sum_{0 \leq i < j \leq k} v_i \cdot v_j \sum_{0 \leq i' < j' \leq k} v_{i'} \cdot v_{j'} \frac{\partial^2}{\partial a_{ij} \partial a_{i'j'}} p_{\omega}(A).
\]

One can think of \((g \cdot w_i + \eta w_0 \cdot w_i)_{i=1}^k\) as a set of \( k \) joint Gaussian random variables. In particular for a fixed \( \omega \) define \( \nu \in \mathbb{R}^n \) and a positive definite matrix \( \Sigma \) as follows. For
\[ 1 \leq i < j \leq k, \]

\[ \nu_i = \eta A_0 \omega_i = \epsilon \eta \omega_i \mathbf{v}_0 \cdot \mathbf{v}_i, \quad \Sigma_{ii} = 1, \quad \Sigma_{ij} = \Sigma_{ji} = A_{ij} \omega_i \omega_j = \epsilon \omega_i \omega_j \mathbf{v}_i \cdot \mathbf{v}_j. \]

It is easy to verify that \( p_\omega(A) \) is indeed the orthant probability of Gaussian distribution with mean \( \nu \) and correlation matrix \( \Sigma \). So according to Lemma 4.7 and (4.8) for \( \epsilon \leq \min(\kappa, \kappa/|\eta|) \),

\[ \left| \frac{d^2}{d \epsilon^2} \Pr[(x_1, \ldots, x_k) \in P^{-1}(1)] \right| \leq 2^k k^4 \Gamma, \]

where \( 2^k k^4 \) is a bound on the number of terms in (4.8) and \( \kappa \) and \( \Gamma \) are constants only depending on \( k \). Now for every such \( \epsilon_0 \) according to Taylor’s theorem for some \( 0 \leq \epsilon' \leq \epsilon_0 \),

\[
\Pr[(x_1, \ldots, x_k) \in P^{-1}(1)]\bigg|_{\epsilon=\epsilon_0} = \frac{|P^{-1}(1)|}{2^k} + \epsilon_0 \frac{d}{d \epsilon} \Pr[(x_1, \ldots, x_k) \in P^{-1}(1)]\bigg|_{\epsilon=0} + \frac{\epsilon_0^2}{2} \frac{d^2}{d \epsilon^2} \Pr[(x_1, \ldots, x_k) \in P^{-1}(1)]\bigg|_{\epsilon=\epsilon'} 
\geq \frac{|P^{-1}(1)|}{2^k} + \epsilon_0 \delta_1/2 - \epsilon_0^2 2^k k^4 \Gamma/2
\]

Setting \( \epsilon_0 \) appropriately\(^6\), this is at least \( \frac{|P^{-1}(1)|}{2^k} + \delta_3 \) for \( \delta_3 = \epsilon_0 \delta_1/4 \) which crucially does \textit{not} depend on \( \delta_2 \). This shows that each clause for which the vector solution gets a value of \( (1 - \sqrt{\delta_2}) \) is going to be satisfied by the rounded solution with probability at least \( \frac{|P^{-1}(1)|}{2^k} + \delta_3 \). As these constitute a \( 1 - \sqrt{\delta_2} \) fraction of all clauses, the overall expected value of the objective function for the rounded solution is at least

\[
(1 - \sqrt{\delta_2}) \left( \frac{|P^{-1}(1)|}{2^k} + \delta_3 \right) \geq \frac{|P^{-1}(1)|}{2^k} + \delta_3 - \sqrt{\delta_2}.
\]

If we set \( \delta_2 < (\delta_3/2)^2 \), this is at least \( \frac{|P^{-1}(1)|}{2^k} + \delta_3/2 \), which provides a lower bound on the approximation ratio of the algorithm on instances with optimal value at least \( 1 - \delta_2 \). This completes the proof. \( \square \)

### 4.2 Signs of Symmetric Quadratic Polynomials

In this section we study signs of symmetric quadratic polynomials, and give a proof of Theorem 4.2. Consider a predicate \( P : \{-1, 1\}^k \to \{0, 1\} \) that is the sign of a symmetric quadratic polynomial with no constant term, i.e.,

\[
P(x) = \frac{1 + \text{sign}(\alpha \sum_i x_i + \beta \sum_{i<j} x_i x_j)}{2}
\]

---

\(^6\epsilon_0 = \min(\kappa, \kappa/|\eta|, 2^{-k} k^{-2} \delta_1/2 \Gamma) \) will do
for some constants $\alpha$ and $\beta$. We would like to apply the $(\epsilon, \eta)$-rounding scheme to MAX k-CSP$(P)$, which in turn requires us to understand the low-degree Fourier coefficients of $P$. Note that because of symmetry, the value of a Fourier coefficient $\hat{P}(S)$ depends only on $|S|$.

We will prove that “morally” $\beta$ has the same sign as the degree-2 Fourier coefficient of $P$ and that if one of them is 0 then so is the other. This statement is not quite true (consider for instance the predicate $P(x_1, x_2) = \frac{1 + \text{sign}(x_1 + x_2)}{2} = \frac{1 + x_1 + x_2 + x_1 x_2}{4}$), however we will show that you can always slightly adjust $\beta$ without changing $P$ so that the above statement becomes true.

**Theorem 4.8.** For any $P$ of the above form, there exists $\beta'$ with the property that $\beta' \hat{P}([1, 2]) \geq 0$ and $\beta' = 0$ iff $\hat{P}([1, 2]) = 0$, satisfying

$$P(x) = \frac{1 + \text{sign}(\alpha \sum x_i + \beta' \sum x_i x_j)}{2}.$$ 

**Proof.** Let us define

$$P_\beta(x) = \frac{1 + \text{sign}(\alpha \sum x_i + \beta \sum x_i x_j)}{2}$$

where we consider $\alpha$ fixed and $\beta$ as a variable. First, we have the following claim:

**Claim 4.9.** $\hat{P}_\beta([1, 2])$ is a monotonically non-decreasing function in $\beta$. Furthermore, if $P_{\beta_1} \neq P_{\beta_2}$ then $\hat{P}_{\beta_1}([1, 2]) \neq \hat{P}_{\beta_2}([1, 2])$.

**Proof.** Fix two arbitrary values $\beta_1 < \beta_2$ of $\beta$, and let $\Delta P : \{-1, 1\}^k \to \{-1, 0, 1\}$ be the difference $\Delta P = P_{\beta_2} - P_{\beta_1}$. Consider an input $x \in \{-1, 1\}^k$. It follows from the definition of $P_\beta$ that if $\Delta P(x) > 0$ then $\sum_{i<j} x_i x_j > 0$, and similarly if $\Delta P(x) < 0$ then $\sum_{i<j} x_i x_j < 0$. Now since $\Delta P$ is symmetric, the level-2 Fourier coefficient of $\Delta P$ equals

$$\hat{\Delta P}([1, 2]) = \frac{1}{\binom{n}{2}} \sum_{i<j} \Delta P(i, j) = \frac{1}{\binom{n}{2}} \mathbb{E}_x \left[ \Delta P(x) \sum_{i<j} x_i x_j \right] \geq 0,$$

with equality holding only if $\Delta P$ is zero everywhere, i.e., if $P_{\beta_1} = P_{\beta_2}$. Given that $\hat{P}_{\beta_1}([1, 2]) = \hat{P}_{\beta_2}([1, 2]) + \hat{\Delta P}([1, 2])$ this completes the proof of the Claim. 

Returning to the proof of Theorem 4.8 suppose first that either $\alpha = 0$ or $k$ is odd. It is easy to check that in these two cases, $\hat{P}_0([1, 2]) = 0$: if $\alpha = 0$ the function $P_0$ is constant, and if $\alpha \neq 0$ but $k$ is odd the function $P_0$ is odd. Consider the set of values $B = \{ \beta | \hat{P}_\beta([1, 2]) = 0 \}$. From Claim 4.9 it follows that $B$ is an interval (possibly consisting of a single point) and that the function $P_\beta$ is the same for all $\beta \in B$. For $\beta < 0$, $\beta \not\in B$, Claim 4.9 shows that $\hat{P}_\beta([1, 2]) < 0$ and so has the same sign as $\beta$, and similarly for $\beta > 0$, $\beta \not\in B$. For $\beta \in B$ we see that $P_\beta = P_0$ so we can set $\beta' = 0$.

The remaining case is that of even $k$, and $\alpha \neq 0$. Notice that if $|\beta|$ is sufficiently small compared to $|\alpha|$, say $|\beta| \leq |\alpha|/k^2$, then $P_\beta(x)$ only differs from $P_0(x)$ on balanced inputs, i.e. when $\sum x_i = 0$. Let $B$ be the set of all such sufficiently small $\beta$. For $\beta \in B$, the only
Chapter 4. On Quadratic Threshold CSPs

contribution to \( \hat{P}_\beta(\{1, 2\}) \) comes from points \( x \) that are balanced. The reason is that for all other \( x \), the contribution \( \text{sign}(\alpha \sum x_i x_1 x_2) \) to \( \hat{P}_0(\{1, 2\}) \) is cancelled by the contribution from the point \(-x\).

When \( x \) is balanced, we have \( \sum_{i<j} x_i x_j = 2^{(n/2)} - (n/2)^2 = -\frac{n}{2} < 0 \) and therefore \( \text{sign}(\alpha \sum x_i + \beta \sum x_i x_j) = \text{sign}(\beta) \), implying

\[
\sum_{i<j} \hat{P}_\beta(\{i, j\}) = 2^{-k} \sum_{x: \sum x_i = 0} \text{sign}(\beta) x_i x_j,
\]

which has the same sign as \(-\text{sign}(\beta)\). Thus we see that if \( \beta > 0 \), \( \beta \in B \) we have \( \hat{P}_\beta(\{1, 2\}) > 0 \), and if \( \beta \leq 0 \), \( \beta \in B \) we have \( \hat{P}_\beta(\{1, 2\}) < 0 \).

Now Claim 4.9 implies that whenever \( \beta \neq 0 \) (not necessarily in \( B \)) we can simply take \( \beta' = \beta \), and that when \( \beta = 0 \) we can take \( \beta' \) to be a negative value close to 0 (e.g., \( \beta' = -|\alpha|/k^2 \)).

We are now ready to prove Theorem 4.2.

**Theorem 4.2** (restated). Let \( P : \{-1, 1\}^k \rightarrow \{0, 1\} \) be a predicate that is of the form \( P(x) = \frac{1 + \text{sign}(Q(x))}{2} \) where \( Q \) is a symmetric quadratic polynomial with no constant term. Then \( P \) is not approximation resistant.

**Proof.** Without loss of generality, we can take \( Q(x) = \alpha \sum x_i + \beta \sum x_i x_j \) where \( \beta \) satisfies the property of \( \beta' \) in Theorem 4.8.

If \( \hat{P}(\{1, 2\}) = \beta = 0 \), we set \( \eta = \alpha/\hat{P}(\{1\}) \) note that in this case we can assume that \( \alpha \), and hence also \( \hat{P}(\{1\}) \) is non-zero as otherwise \( P \) is the trivial predicate that is always false. We then have, for every \( x \in P^{-1}(1) \),

\[
\frac{2\eta}{\sqrt{2\pi}} P^{=1}(x) + \frac{2}{\pi} P^{=2}(x) = \frac{2\alpha}{\sqrt{2\pi}} \sum x_i,
\]

which is positive by the definition of \( P \). If \( \hat{P}(\{1, 2\}) \neq 0 \), we set \( \eta = \sqrt{\frac{2}{\pi P(\{1\})}} \cdot \hat{P}(\{1, 2\})/\beta \). In this case for every \( x \in P^{-1}(1) \),

\[
\frac{2\eta}{\sqrt{2\pi}} P^{=1}(x) + \frac{2}{\pi} P^{=2}(x) = \frac{2\hat{P}(\{1, 2\})}{\pi \beta} \left( \alpha \sum x_i + \beta \sum x_i x_j \right) > 0,
\]

since \( \beta \) agrees with \( \hat{P}(\{1, 2\}) \) in sign and \( Q(x) > 0 \). In either cases, using Theorem 4.6 and the respective choices of \( \eta \) we conclude that \( P \) is approximable.

**4.3 Monarchy**

In this section we prove that for \( k > 4 \) the Monarchy predicate is not approximation resistant. Notice that Monarchy is defined only for \( k > 2 \), and that the case \( k = 3 \) coincides with the
Chapter 4. On Quadratic Threshold CSPs

predicate majority that is known to be approximable. Further, the case $k = 4$ is handled by the exhaustive classification of Hast [Has05a].

Just like the algorithm for symmetric predicates we first solve the natural semidefinite program of Monarchy, and then use a rounding algorithm to construct an integral solution from the vectors. The rounding algorithm, which is shown below as Algorithm 4, has two parameters $\epsilon > 0$ and an odd positive integer $\ell$, both depending on $k$ to be fixed in the proof.

Algorithm 4: Rounding SDP solutions for Monarchy

**Input:** “biases” $b_1 = v_0 \cdot v_1, \ldots, b_n = v_0 \cdot v_n$.

**Output:** $x_1, \ldots, x_n \in \{-1, 1\}$.

**Parameter:** $\epsilon > 0$, $\ell$ an odd integer.

1. Choose a threshold $\tau \in \left[1/(2k^2), 1/k^2\right]$ uniformly at random.

2. for $i \leftarrow 1$ to $n$ do

3. if $b_i > 1 - \tau$ then

4. $x_i \leftarrow 1$

5. else if $b_i < -1 + \tau$ then

6. $x_i \leftarrow -1$

7. else

8. set $x_i$ (independent of all other $x_j$’s), randomly to $-1$ or $1$ such that $E[x_i] = \epsilon b_i^\ell$.

9. end

10. end

Remark. As the reader may have observed, the “geometric” power of SDP is not used in the above rounding scheme, and indeed a linear programming relaxation of the problem would suffice here. However, in the interest of consistency and being able to describe the techniques in a language comparable to Theorem 4.2 we elected to use the SDP framework.

We will first discuss the intuition behind the analysis of the algorithm ignoring, for now, the greedy ingredient, i.e. lines 3-6. Notice that for $\epsilon = 0$ the rounding produces a uniformly random assignment to the variables, hence the expected value of the obtained solution is $1/2$. From the discussion of the previous section, as long as $\epsilon > 0$ is small enough, the probability of success for a clause is essentially only affected by the degree-one Fourier coefficients of Monarchy. Now, fix a clause and assume that the SDP solution completely satisfies it. Specifically, consider the clause Monarchy$(x_1, \ldots, x_k)$, and define $b_1, \ldots, b_k$ as the corresponding biases. Notice that all Fourier coefficients of Monarchy are positive. This implies that the rounding scheme above will succeed with probability that is essentially $1/2$ plus some positive linear combination of the $\epsilon b_i^\ell$. Our objective is then to fix $\ell$ that would make the value of this combination positive (and independent from $n$). It turns out that the maximal $|b_i|$ in magnitude (call it $b_j$) is always positive in this case. Oversimplifying, imagine that $|b_j| \geq |b_i| + \xi$ for all $i$ different than $j$ where

---

7In the notation of [Has05a], Monarchy on 4 variables is the predicate 00000001011111111, which is listed as approximable in Table 6.6. We remark that this is not surprising since Monarchy in this case is simply a majority in which $x_1$ serves as a tie-breaker variable.
ξ is some positive constant. In this setting it is easy to take \( \ell \) (a function of \( k \)) that makes the effect of all \( b_i \) other than \( b_j \) vanish, ensuring a positive addition to the probability as desired so that overall the expected fraction of satisfied clauses is more than 1/2.

More realistically, the above slack \( \xi \) does not generally exist. However, we can show that a similar condition holds provided that the \(|b_i|\) are bounded away from 1. This condition suffices to prove that the rounding algorithm works for clauses that do not have any variables with bias very close to \( \pm 1 \). The case where there are \( b_i \) that are very close to 1 in magnitude is where the greedy ingredient of the algorithm, i.e. lines 3-6, is used; we can show that when \( \tau \) is roughly \( 1/k^2 \), this ingredient works. In particular, we can show that for each clause, if lines 3-6 are used to round one of the variables, they are used to round essentially every variable in the clause. Also, if this happens, the clause is going to be satisfied with high probability by the rounded solution.

The last complication stems from the fact that the clauses are generally not completely satisfied by the SDP solution. However, as in Section 4.2 a standard averaging argument implies that it is enough to deal with clauses that are *almost* satisfied by the SDP solution. For any such clause the SDP induces a probability distribution on the variables that is mostly supported on satisfying assignments, compared to *only* on satisfying assignments in the above ideal setting. As such, the corresponding \( b_i \)’s can be thought of as a perturbed version of the biases in that ideal setting. Unfortunately, the greedy ingredient of the algorithm is very sensitive to such small perturbations. In particular, if the biases are very close to the set threshold, \( \tau \), a small perturbation can break the method. To avoid this, we choose the actual threshold randomly, and we argue that only a small fraction of the clauses end up in such unfortunate configurations.

This completes the high level description of the proof of our second result.

**Theorem 4.4** (restated). *The predicate Monarchy is not approximation resistant.*

**Proof.** As with the case of symmetric predicates we can assume that the objective value of the SDP solution is at least \( 1 - \delta \) for a fixed constant \( \delta \) to be set later, and we can focus on clauses with SDP value at least \( 1 - \sqrt{\delta} \). Again, similar to our analysis of the symmetric predicates we consider one of these constraints and without loss of generality assume that this constraint is on \( x_1, \ldots, x_k \). Remember that the variables \( I_{[k],*} \) define a distribution, say \( \mu \), on \( \{-1, 1\}^k \) such that

\[
\Pr_{y \sim \mu}[\text{Monarchy}(y) = 1] \geq 1 - \sqrt{\delta}, \quad \forall i \ b_i = \mathbb{E}_{y \sim \mu}[y_i].
\]

Given that we choose \( \tau \) uniformly at random in an interval of length \( 1/(2k^2) \), for any particular clause the probability that \(|b_1|\) is of distance less than \( 2\sqrt{\delta} \) from \( 1 - \tau \) is at most \( 8\sqrt{\delta}k^2 \) and in particular, there are in expectation no more than a \( 8\sqrt{\delta}k^2 \) fraction of clauses for which the following does not hold.

\[
|b_1| \notin [1 - \tau - 2\sqrt{\delta}, 1 - \tau + 2\sqrt{\delta}]. \tag{4.9}
\]

We will assume that (4.9) holds for our clause.
Given that $\mu$ is almost completely supported on points that satisfy Monarchy, we will first prove a few properties of distributions supported on satisfying points of Monarchy.

**Lemma 4.10.** For a distribution $\nu$ on $\{-1, 1\}^k$, completely supported on satisfying points of monarchy, i.e. $\Pr_{y \sim \nu}[\text{Monarchy}(y) = 1] = 1$, let 

$$\forall i \quad b_i \overset{\text{def}}{=} \mathbb{E}_{y \sim \nu}[y_i].$$

Then, 

$$\forall i > 1 \quad b_i \geq -b_1,$$  \hspace{1cm} (4.10)

$$\frac{1}{k-1} \sum_{i>1} b_i \geq -b_1 + (1 + b_1)/(k-1).$$  \hspace{1cm} (4.11)

**Proof.** These two properties are linear conditions on the distribution, so if we check them for all points satisfying Monarchy they will follow for every distribution by convexity. They are two types of points in $\text{Monarchy}^{-1}(1)$. There is the point $(-1, 1, \ldots, 1)$, and there are points of the form $(1, z_2, \ldots, z_n)$ where not all $z_i$'s are $-1$. One can check that (4.10) and (4.11) hold for both kinds of points. \hfill \blacksquare

Given that the distribution $\mu$ has (at least) mass $1 - \sqrt{\delta}$ on the support of Monarchy, Lemma 4.10 immediately implies that, for our $b_1, \ldots, b_n$,

$$\forall i > 1 \quad b_i \geq -b_1 - 2\sqrt{\delta},$$  \hspace{1cm} (4.12)

$$\frac{1}{k-1} \sum_{i>1} b_i \geq -b_1 + (1 + b_1)/(k-1) - 2\sqrt{\delta}.$$  \hspace{1cm} (4.13)

Now, as promised, we show that the deterministic rounding of the variables which are extremely biased does not interfere with the randomized rounding of the rest of the variables.

**Lemma 4.11.** For any clause for which the SDP value is at least $1 - \sqrt{\delta}$, and $b_1$ satisfies the range requirement of (4.9), one of the following happens,

1. $x_1$ is deterministically set to $-1$ and all the rest of $x_i$’s are deterministically set to $1$.
2. $x_1$ is deterministically set to $1$, and at least two of the other $x_i$’s are not deterministically set to $-1$.
3. $x_1$ is deterministically set to $1$, and for some $i > 1$, $b_i \geq 1 - 3/2(k-2)$.
4. $x_1$ is not set deterministically, and no other $x_i$ is deterministically set to $-1$.

**Proof.** The proof is by case analysis based on how $x_1$ is rounded.
First, assume that \( x_1 \) is deterministically set to \(-1\). It follows from (4.12) that all the other \( b_i \)'s are at least \(-b_1 - 2\sqrt{\delta} \) so by the assumption in (4.9) we know that we are in case 1 of the theorem and we are done.

Now, assume that \( x_1 \) is deterministically set to 1. If for two distinct \( i \)'s \( b_i > -1 + \tau \) we are in case 2 of the theorem and we are done. Otherwise, assume that \( b_j \) is the biggest of \( b_i \)'s, and in particular all other \( b_i \)'s are at most \(-1 + \tau \), we have,

\[
\frac{1}{k-1} \sum_{i>1} b_i \geq -b_1 + (1 + b_1)/(k-1) - 2\sqrt{\delta} \quad \text{by (4.13)}
\]

\[
\frac{1}{k-1} \sum_{i>1} b_i \leq \frac{b_j}{k-1} + \frac{k-2}{k-1}(-1 + \tau) \quad \text{by assumption}
\]

\[
\Rightarrow b_j \geq -(k-2)b_1 + 1 - 2\sqrt{\delta}(k-1) - (k-2)(-1 + \tau)
\]

\[
= 1 - (k-2)(b_1 - 1 + \tau) - 2\sqrt{\delta}(k-1)
\]

\[
\geq 1 - (k-2)\tau - 2\sqrt{\delta}(k-1)
\]

\[
> 1 - 1/(k-2) - 2\sqrt{\delta}(k-1) \geq 1 - 1.5/(k-2), \quad \text{by } \tau \leq 1/k^2
\]

where the last step assumes \( \delta < 1/16(k-1)^2(k-2)^2 \); we have shown that we are in the third case and we are done.

Finally, assume that \( x_1 \) is not deterministically rounded, i.e. \(-1 + \tau \leq b_1 \leq 1 - \tau \). It follows from (4.9) that in fact, \( b_1 < 1 - \tau - 2\sqrt{\delta} \). So, one can use (4.12) to deduce that for all \( i > 1 \),

\[
b_i \geq -b_1 - 2\sqrt{\delta} > -1 + \tau + 2\sqrt{\delta} - 2\sqrt{\delta} = -1 + \tau.
\]

So, we are in case 4 of the theorem and we are done. \( \blacksquare \)

We can now look at different cases and show that the parameters \( \epsilon \) and \( \ell \) can be set appropriately such that in all the cases the rounded \( x_1, \ldots, x_k \) satisfy the predicate with probability at least \( 1/2 + \gamma \) for some constant \( \gamma \). Intuitively, in the first three cases the clause is deterministically rounded and has a high probability of being satisfied, while the forth case is when the analysis of the clause needs arguments about the absolute values of the biases, and the clause is satisfied with probability slightly above 1/2. We will handle the first three cases first.

**Lemma 4.12.** If one of the three first cases of Lemma 4.11 happen for a clause then it is satisfied in the rounded solution with probability at least 1, \( 1 - (1 + \epsilon)^2/4 \), and \( 1/2 + \epsilon 2^{-\ell-1} \) respectively. In particular, if \( \epsilon \) and \( \ell \) are constants only depending on \( k \) and \( \epsilon < \sqrt{2} - 1 \) the clause is satisfied with probability at least \( 1/2 + \gamma \) for some constant \( \gamma \) independent of \( n \).

**Proof.** The first two cases are easy: in the first case the clause is always satisfied by \( x_1, \ldots, x_k \) while in the second case it is satisfied if and only if at least one \( i > 1 \), \( x_i \) is set to +1. Given that at least two of these \( x_i \)'s are not deterministically rounded to \(-1\), the clause is unsatisfied with probability at most \((1 + \epsilon)^2/4\).
Now, assume that a clause is in the third case. Then, we know that for some $i > 1$, $b_i \geq 1 - 3/2(k-2)$, so the clause is satisfied with probability at least $\frac{1+\epsilon(1-3/2(k-2))}{2} \geq \frac{1}{2} + \epsilon 2^{-\ell-1}$, where we have used $k \geq 5$.

All that remains is to show that in the fourth case the clause is satisfied with some probability greater than $1/2$ and to find the suitable values of $\epsilon$ and $\ell$ in the process. This is formulated as the next lemma.

**Lemma 4.13.** There are constants $\epsilon = \epsilon(k)$, $\ell = \ell(k)$, and $\gamma = \gamma(k)$, such that, for any small enough $\delta$, any clause for which the fourth case of Lemma 4.11 holds is satisfied with probability at least $1/2 + \gamma$, if we round with parameters $\epsilon$ and $\ell$.

**Proof.** We can assume that none of the $x_i$’s are deterministically rounded as being rounded to 1 only helps us. We consider two cases: either $b_1 \leq 0$, or $b_1 \geq 0$. But first let us write the probability that this clause is satisfied in terms of the Fourier coefficients.

$$\Pr[\text{Monarchy}(x) = 1] = \text{Monarchy}(\emptyset) + \sum_{i=1}^{k} \text{Monarchy}({\{i}\}}) eb_i^\ell + \sum_{i<j} \text{Monarchy}({\{i, j}\}}) e^2 b_i^\ell b_j^\ell + \cdots$$

$$\geq \text{Monarchy}(\emptyset) + \epsilon \sum_{i=1}^{k} \text{Monarchy}({\{i}\}}) b_i^\ell - \epsilon^2 2^k \max_{S:|S|>1} |\text{Monarchy}(S)|$$

$$= 1/2 + \epsilon \sum_{i=1}^{k} \text{Monarchy}({\{i}\}}) b_i^\ell - \epsilon^2 2^k.$$

So all we have to do is to find a value of $\ell$ and a positive lower bound for $\sum_{i=1}^{k} \text{Monarchy}({\{i}\}}) b_i^\ell$ that holds for all valid $b_i$’s; the term $-\epsilon^2 2^k$ can be ignored for small enough $\epsilon$. It is easy to see that

$$\text{Monarchy}({\{1\}}) = 1/2 - 2^{1-k} \quad \forall i > 1 \quad \text{Monarchy}({\{i\}}) = 2^{1-k}.$$

Define,

$$C \overset{\text{def}}{=} \frac{\text{Monarchy}({\{1\}})}{\text{Monarchy}({\{2\}})} \approx 2^{k-2}.$$

$$f(b_1, \ldots, b_k) \overset{\text{def}}{=} \frac{1}{\text{Monarchy}({\{2\}})} \sum_{i=1}^{k} \text{Monarchy}({\{i\}}) b_i^\ell = C b_1^\ell + \sum_{i>1} b_i^\ell.$$

We have to lower bound $f(b_1, \ldots, b_k)$. 

First, assume $b_1 \leq (2k - 4)^{-3}/2$. We know that,

\[
f(b_1, \ldots, b_k) = Cb_1^\ell + \sum_{i>1} b_i^\ell \geq Cb_1^\ell + (k - 1)\left(\frac{1}{k - 1} \sum_i b_i\right)^\ell\]

by concavity of $x^\ell$,

\[
\geq Cb_1^\ell + (k - 1)(-b_1 + (1 + b_1)/(k - 1) - 4\sqrt{\delta})^\ell\]

by (4.13),

\[
\geq Cb_1^\ell + (k - 1)(-b_1 + \tau/(k - 1) - 4\sqrt{\delta})^\ell\]

as we are in case 4,

\[
\geq Cb_1^\ell + (k - 1)(-b_1 + 1/2(k - 2)^2(k - 1) - 4\sqrt{\delta})^\ell\]

by choice of $\tau$,

\[
\geq Cb_1^\ell + (k - 1)(-b_1 + 1/4(k - 2)^3 - 4\sqrt{\delta})^\ell
\]

assuming $\delta < 2^{-5}(k - 2)^{-6}$.

Now if $b_1 \geq 0$ we are done as $f(b_1, \ldots, b_k)$ would be at least $(k - 1)(2^k - 4)^{-3\ell/2 - \ell}$ and any constant $\ell$ will do the job. Otherwise, note that the expression inside the parenthesis is at least $(2k - 4)^{-3}$ bigger than $b_1$ in absolute value. So, if we take $\ell$ to be big enough the second expression is going to dominate the expression. Specifically, first assume $|b_1| \geq (4k - 8)^{-3},$

\[
f(b_1, \ldots, b_k) > Cb_1^\ell + (k - 1)(-b_1 + (2k - 4)^{-3})^\ell
\]

\[
\geq Cb_1^\ell - b_1^\ell(k - 1) + (k - 1)b_1^\ell - 1(2k - 4)^{-3}
\]

\[
= ((C - k + 1)b_1 + \ell(k - 1)(2k - 4)^{-3})b_1^\ell - 1
\]

\[
\geq ((C - k + 1)b_1 + \ell(k - 1)(2k - 4)^{-3})(4k - 8)^{-3\ell + 3}
\]

as $\ell - 1$ is even

\[
> -(C - k + 1) + \ell(k - 1)(2k - 4)^{-3})(4k - 8)^{-3\ell + 3}
\]

by $b_1 < 1$

Which clearly has a constant lower bound if $\ell \geq (C - k + 1)(2k - 4)^3$. Now if $|b_1| < (4k - 8)^{-3}$ we can write,

\[
f(b_1, \ldots, b_k) > Cb_1^\ell + (k - 1)(-b_1 + (2k - 4)^{-3})^\ell
\]

\[
> Cb_1^\ell + (k - 1)(2k - 4)^{-3\ell}
\]

\[
> -C(4k - 8)^{-3\ell} + (k - 1)(2k - 4)^{-3\ell}
\]

\[
= (2k - 4)^{-3\ell}(-C2^{-3\ell} + (k - 1)),
\]

\[
> (2k - 4)^{-3\ell}(k - 2),
\]

as $3\ell > \log_2 C$

which is a constant as long as $\ell$ is some constant. This completes the case of $b_1 \leq (2k - 4)^3/2$. Note that so far we have assumed that $\ell$ is some constant and at least $\max(\log_2 C/3, (C - k + 1)(2k - 4)^3)$.

Let's assume $b_1 > (2k - 4)^{-3}/2$. One can write,

\[
f(b_1, \ldots, b_k) = Cb_1^\ell + \sum_{i>1} b_i^\ell \geq Cb_1^\ell + (k - 1)(-b_1 - 4\sqrt{\delta})^\ell
\]

by (4.12),

\[
= b_1^\ell \left(\frac{C}{b_1} - (k - 1)(1 + 4\sqrt{\delta}/b_1)^\ell\right)
\]
\[ b_1^\ell \left( C - (k - 1)(1 + 8(2k - 4)^3 \sqrt{\delta})^\ell \right) \]
\[ > b_1^\ell \left( C - (k - 1) \exp(8(2k - 4)^3 \sqrt{\delta})^\ell \right) \]
\[ > b_1^\ell \left( C - (k - 1) \right) \exp(8(2k - 4)^3 \sqrt{\delta})^\ell \]
\[ \geq b_1^\ell \geq (2k - 4)^{-3\ell^2/2}, \]

where to derive the last line we have assumed

\[ \delta \leq (4k - 8)^{-6\ell^2 - 2}(\ln(C - 1) - \ln(k - 1))^2 \]
\[ \sqrt{\delta} \leq (4k - 8)^{-3\ell^2 - 1}(\ln(C - 1) - \ln(k - 1)) \]
\[ 8(2k - 4)^3 \sqrt{\delta} \ell \leq \ln(C - 1) - \ln(k - 1) \]
\[ (k - 1) \exp(8(2k - 4)^3 \sqrt{\delta})^\ell \leq C - 1 \]

So, as long as \( C - 1 > k - 1 \) we can set \( \ell = \max(\log_2 C/3, (C - k + 1)(2k - 4)^3) \) and prove that \( f(b_1, \ldots, b_k) \) has a constant lower bound depending on \( k \), provided that we assume \( \delta < \delta_0 \) where \( \delta_0 \) is another function of \( k \). One can check that \( C > k - 2 \) whenever \( k > 4 \) which completes the proof.

We can now finish the proof of Theorem 4.4. In particular, solve the SDP relaxation of Monarchy, if the objective value is smaller than \( 1 - \delta \) output a uniformly random solution, and if it is bigger apply the rounding algorithm in Figure 4. In the first case the expected objective value of the output is \( 1/2 \), while the optimal solution cannot have objective value more than \( 1 - \delta \), giving rise to approximation ratio \( 2(1 - \delta) = 2 - 2\delta \). In the second case, at least a \( (1 - \sqrt{\delta}) \) fraction of the clauses have objective \( 1 - \sqrt{\delta} \) and from these in expectation at least a \( 1 - 16\sqrt{\delta}(k - 2)^2 \) fraction satisfy (4.9). We can apply Lemma 4.12 and Lemma 4.13 to these. So, the objective function is at least,

\[ (1 - \sqrt{\delta})(1 - 16\sqrt{\delta}(k - 2)^2)(1/2 + \gamma) \geq 1/2 + \gamma - 17\sqrt{\delta}(k - 2)^2. \]

This is clearly more than \( 1/2 + \gamma/2 \) for small enough \( \delta \), which shows that the approximation factor of the algorithm is at most \( 2/(1 + \gamma) < 2 - \gamma \) for a small but constant \( \gamma \) finishing the proof of the theorem.

4.4 Proof of Lemma 4.7

In this section we present the proof Lemma 4.7 restated here for convenience.

**Lemma 4.7 (restated).** For fixed \( k \), define the function \( \text{ort}(\nu, \Sigma) \) as the orthant probability of the multivariate normal distribution with mean \( \nu \) and covariance matrix \( \Sigma \), where \( \nu \in \mathbb{R}^k \) and \( \Sigma_{k \times k} \) is a positive semidefinite matrix. That is,

\[ \text{ort}(\nu, \Sigma) \overset{\text{def}}{=} \Pr_{\mathbf{x} \sim N(\nu, \Sigma)} [\mathbf{x} \geq 0]. \]
There exists a global constant $\Gamma$ that upper bounds all the second partial derivatives of $\text{ort}(\cdot)$ when $\Sigma$ is close to $I$. In particular, for all $k$, there exist $\kappa > 0$ and $\Gamma$, such that for all $i_1, j_1, i_2, j_2 \in [k]$, all vectors $\nu \in \mathbb{R}^k$ and all positive definite matrices $\Sigma_{k \times k}$,

$$|I - \Sigma|_\infty, |\nu|_\infty < \kappa \Rightarrow \left| \frac{\partial^2}{\partial \Sigma_{i_1 j_1} \partial \Sigma_{i_2 j_2}} \text{ort}(\nu, \Sigma) \right| < \Gamma,$$

$$\left| \frac{\partial^2}{\partial \Sigma_{i_1 j_1} \partial \nu_{i_2}} \text{ort}(\nu, \Sigma) \right| < \Gamma,$$

$$\left| \frac{\partial^2}{\partial \nu_{i_1} \partial \nu_{i_2}} \text{ort}(\nu, \Sigma) \right| < \Gamma.$$

**Proof.** We will set $\kappa$ small enough so that $\Sigma$ is full rank. We know that whenever $\Sigma$ is a full rank positive definite matrix, $\text{ort}(\nu, \Sigma) = 1 \left( \frac{2 \pi}{k} \right)^{k/2} |\Sigma|^{1/2} \int_{x_1 = 0}^{+\infty} \cdots \int_{x_k = 0}^{+\infty} \phi(x, \nu, \Sigma) dx_1 \cdots dx_k,$

where,

$$\phi(x, \nu, \Sigma) = \exp \left( -\frac{1}{2} (x - \nu)^\top \Sigma^{-1} (x - \nu) \right) = \exp \left( -\frac{1}{2} \sum_{l,m \in [k]} (x_l - \nu_l)(x_m - \nu_m)(-1)^{l+m}|\Sigma^{ml}|/|\Sigma| \right),$$

here $\phi$ is (a normalization of) the probability density function of the multivariate normal distribution and $\Sigma^{ml}$ is the minor of $\Sigma$ obtained by removing row $m$ and column $l$ and $|\Sigma|$ is the determinant of $\Sigma$. More abstractly, we can write

$$\phi(x, \nu, \Sigma) = \exp \left( \sum_{l,m \in [k]} (x_l - \nu_l)(x_m - \nu_m)p_{lm}(\Sigma)/q(\Sigma) \right),$$

where $p_{lm}$ and $q$ are polynomials of degree $\leq k$ in $\Sigma$ (depending only on $k$), and where $q(\Sigma)$ is bounded away from 0 in a region around $\Sigma$. Let $y_i = x_i - \nu_i$. We can then write

$$\frac{\partial^2}{\partial \Sigma_{ij} \partial \Sigma_{i'j'}} \phi(x, \nu, \Sigma) = \frac{\phi(x, \nu, \Sigma)}{q(\Sigma)^4} \sum_{l,m} y_l y_m (A_{lm,ij,i'j'}(\Sigma) + y_l y_m B_{lm,ij,i'j'}(\Sigma))$$

where $A_{lm,ij,i'j'}$ and $B_{lm,ij,i'j'}$ are polynomials depending only on $p_{lm}$, $q$, $i$, $j$, $i'$ and $j'$. Thus, for all $\Sigma$ in a neighbourhood around $I$, we have

$$\left| \frac{\partial^2}{\partial \Sigma_{ij} \partial \Sigma_{i'j'}} \phi(x, \nu, \Sigma) \right| \leq C \sum_{l,m} (|y_l y_m| + y_l^2 y_m^2) \phi(x, \nu, \Sigma) \leq \sum_{l,m} \left( \frac{1}{2} + \frac{3}{2} y_l^2 y_m^2 \right) \phi(x, \nu, \Sigma),$$

for a constant $C$ (depending only on $k$). By an iterative application of Leibniz Integral Rule
we can bound the second derivative of $\text{ort}(\nu, \Sigma)$ as 

$$\left| \frac{\partial^2}{\partial \Sigma_{ij} \partial \Sigma_{i'j'}} \text{ort}(\nu, \Sigma) \right| \leq \sum_{l,m \in [k]} \frac{C}{(2\pi)^{k/2} |\Sigma|^{1/2}} \int_{x_1 = 0}^{+\infty} \cdots \int_{x_k = 0}^{+\infty} \left( \frac{1}{2} + \frac{3}{2} (x_l - \nu_l)^2 (x_m - \nu_m)^2 \right) \phi(x, \nu, \Sigma) \leq \sum_{l,m \in [k]} \frac{C}{(2\pi)^{k/2} |\Sigma|^{1/2}} \int_{x_1 = -\infty}^{+\infty} \cdots \int_{x_k = -\infty}^{+\infty} \left( \frac{1}{2} + \frac{3}{2} (x_l - \nu_l)^2 (x_m - \nu_m)^2 \right) \phi(x, \nu, \Sigma)$$

$$= C \sum_{l,m \in [k]} \left( \frac{1}{2} + \frac{3}{2} \mathbb{E}_{\nu \sim N(\nu, \Sigma)} \left[ (x_l - \nu_l)^2 (x_m - \nu_m)^2 \right] \right) \leq 19k^2 C,$$

where we have assumed each element of $\Sigma$ is at most 2 in absolute value. The cases of partial derivatives with respect to $\nu_i$’s follows similarly. \hfill \Box
Chapter 5

Extending SDP Integrality Gaps to Sherali-Adams with Applications to Quadratic Programming and Max Cut Gain

In this chapter we show how under certain conditions one can extend an integrality gap lower bound for the canonical SDP relaxation of a problem to the stronger level-$k$ Sherali-Adams SDP relaxation. The value of $k$ depends on properties of the problem. We present two applications, to the Quadratic Programming problem and to the Max Cut Gain problem.

Our technique is inspired by a paper of Raghavendra and Steurer [RS09a] and our result gives a doubly exponential improvement for Quadratic Programming on another result by the same authors [RS09b]. They provide tight integrality gap lower bounds for the level-$(\log \log n)^{\Omega (1)}$ Sherali-Adams SDP relaxation of Quadratic Programming while we prove tight lower bounds for level-$n^{\Omega (1)}$ of the same hierarchy.

Our main problem of interest in this chapter will be Quadratic Programming as defined in Section 2.1.2. Here the input is a matrix $A_{n \times n}$ and the objective is to find $x \in \{-1, 1\}^n$ that maximizes the quadratic form $\sum_{i \neq j} a_{ij}x_ix_j$. The natural semidefinite programming relaxation of this problem replaces the integer ($\pm 1$) valued $x_i$’s with unit vectors, and products with inner products; see Figure 5.1. This problem has applications in correlation clustering and its relaxation is related to the well-known Grothendieck inequality [Gro53]; see [AMNN06, CW04]. Charikar and Wirth [CW04] show that the integrality gap of this relaxation is at most $O(\log n)$ and a matching lower bound was later established by a chain of papers [AMNN06, ABK+05, KO06]. To lower bound the integrality gap of the SDP program in Figure 5.1(b) Khot and O’Donnell [KO06] construct a matrix $A$ and a solution for the SDP relaxation in Figure 5.1(b) that has objective value at least $\Theta(\log n)$ bigger than any integral solution, i.e. any solution of

---

1Results in this chapter appear in [BM10]
the formulation in Figure 5.1(a).

(a) the quadratic program formulation.

(b) The SDP relaxation.

Figure 5.1: Canonical formulation and relaxation of Quadratic Programming for an input matrix $A_{n \times n} = [a_{ij}]$.

It is then natural to ask if strengthening the SDP with a lift and project method will reduce this $\Theta(\log n)$ gap resulting in a better approximation algorithm. To this end in this chapter we investigate the performance of the level-$k$ Sherali-Adams SDP relaxation of Quadratic Programming. We show that for Quadratic Programming the integrality gap is still $\Theta(\log n)$ even for level $n^\delta$ of the Sherali-Adams SDP hierarchy where $\delta > 0$ is an absolute constant. The technique used is somewhat general and can be applied to other problems, and as an example we apply it to the related problem of Max Cut Gain. We can prove the following two theorems.

**Theorem 5.1.** The integrality gap of the level-$N^{\Omega(1)}$ Sherali-Adams SDP relaxation of Quadratic Programming is $\Theta(\log N)$.

**Theorem 5.2.** The integrality gap of the level-$\omega(1)$ Sherali-Adams SDP relaxation of Max Cut Gain is $\omega(1)$. That is for any constants $l,c$ the integrality gap of the level-$l$ Sherali-Adams iSDP of Max Cut Gain is at least $c$.

It should be mentioned that Arora et al. [ABK+05] show that Quadratic Programming is quasi-NP-hard\(^2\) to approximate within a factor of $\log^\gamma n$ for some constant $\gamma > 0$ and Khot and O’Donnel [KO06] show that Max Cut Gain is hard to approximate within any constant factor assuming the Unique Games Conjecture. However, our results do not rely on any such assumptions and are sharp for both problems. Also, given that the Sherali-Adams strengthening can only be solved in polynomial time for constant levels, Theorem 5.1 precludes an algorithm (for Quadratic Programming) based on the standard relaxation with Sherali-Adams strengthening that runs in time $2^{n^\delta}$ for some small constant $\delta > 0$.

**Outline of our approach**

We start with a known integrality gap instance for the SDP relaxation and its SDP solution $\alpha$. Now consider an integral solution $\beta$ to the same instance; obviously $\beta$ is also a solution to level-$k$ Sherali-Adams SDP relaxation of the problem for any $k$. We combine $\alpha$ and $\beta$ into a

\(^2\)i.e. impossible to approximate unless $\text{NP} \subseteq \text{TIME}(2^{\text{poly}(\log n)})$


candidate solution $\gamma$ that has some of the good qualities of both, namely it has high objective value like $\alpha$ and has almost the same behaviour with regards to the Sherali-Adams hierarchy as $\beta$. In particular, the only conditions of the level-$k$ Sherali-Adams SDP relaxation that $\gamma$ violates are some positivity conditions, and those violations can be bounded by some quantity $\xi$. To handle these violations we take a convex combination of $\gamma$ with a solution in which these violated conditions are satisfied with substantial slack. In fact a uniform distribution over all integral points (considered as a solution in the spirit of Fact 2.1) satisfies these positivity constraints with such slack. The weights we need to give to $\gamma$ and the uniform solution in the convex combination will depend on $\xi$ and the amount of slack in the uniform solution which are in turn a function of the level parameter $k$. In turn these two weights determine the objective value of the resulting solution. The result is a trade-off between the level $k$ and the objective value of the resulting solution.

This idea of “smoothening” a solution like the above to satisfy Sherali-Adams constraints is originally due to Raghavendra and Steurer [RS09a]. They use this idea together with dimensionality reduction to give a rounding algorithm that is optimal for a large class of constraint satisfaction problems. In another work [RS09b] the same authors show how to get integrality gaps lower bounds for Sherali-Adams SDP relaxations similar to the ones we show in this chapter. However, in [RS09b] the main vehicle in getting the results are reductions from the Unique Games Problem and “smoothening” is used as a small step.

It is interesting to compare our results with [RS09b]. While [RS09b] gives a more general result in that it applies to a broader spectrum of problems, it only applies to a relatively low level of the Sherali-Adams hierarchy. In particular, for QUADRATIC PROGRAMMING their result is valid for Sherali-Adams SDP relaxation of level up to $(\log \log n)^{\Omega(1)}$ whereas ours is valid up to level $n^{\Omega(1)}$. We note that the two results exhibit different level/integrality gap tradeoffs as well; [RS09b] provides asymptotically the same integrality gap until the “critical level” in which point it breaks down completely. Our results supplies a more smooth tradeoff with the integrality gap dropping “continuously” as the level increases. An additional difference is that our result does not use reductions from the Unique Games reductions and is elementary.

**Organization**

The rest of this chapter is organized as follows. In Section 5.1 we state and prove our main technical lemma. Section 5.2 presents an application of the lemma to QUADRATIC PROGRAMMING. There we prove a lower bound for the integrality gap for the level-$n^\delta$ Sherali-Adams SDP relaxation of this problem for some $\delta > 0$. We leave the proof of a technical lemma concerning the inner products of random vectors used in that section for Section 5.4. In Section 5.3 we present our application to the MAX CUT GAIN problem; in particular we present super-constant integrality gaps for level-$\omega(1)$ Sherali-Adams SDP relaxation of that problem.
5.1 Extending Vector Solutions to Sherali-Adams Hierarchy

In this section we provide the general framework that suggests how vector (or SDP) solutions can be extended to Sherali-Adams solutions of comparable objective value. This framework is captured by the following lemma.

**Lemma 5.3.** Let \( v_0, v_1, \ldots, v_n \in S^{n-1} \) be unit vectors, \( \nu \) be a distribution on \( \{-1, 1\}^n \), \( k \) be a positive integer and \( \epsilon \in (0, 1/2] \) be small enough to satisfy,

\[
\begin{align*}
\text{for all } i \in [n] & \quad 2\epsilon k^2 |v_0 \cdot v_i - \mathbb{E}_{x \sim \nu} [x_i]| \leq 1, \\
\text{for all } i \neq j \in [n] & \quad 2\epsilon k^2 |v_i \cdot v_j - \mathbb{E}_{x \sim \nu} [x_i x_j]| \leq 1.
\end{align*}
\]

Then there exist unit vectors \( u_0, u_1, \ldots, u_n \in S^{n-1} \) and a family of distributions \( \{D(S)\}_{S \subseteq \binom{[n]}{\leq k}} \) such that \( \{D(S)\} \) are a family of consistent local distributions and they are compatible with \( u_i \)'s as per Definitions 2.9 and 2.10. Furthermore, the inner products of \( u_i \)'s are related to those of \( v_i \)'s as follows

\[
\begin{align*}
\text{for all } i \in [n] & \quad u_0 \cdot u_i = \epsilon v_0 \cdot v_i, \quad (5.1) \\
\text{for all } i \neq j \in [n] & \quad u_i \cdot u_j = \epsilon v_i \cdot v_j. \quad (5.2)
\end{align*}
\]

Before we prove the lemma we will briefly describe its use. For simplicity we will only describe the case where \( \nu \) is the uniform distribution on \( \{-1, 1\}^n \). To use the lemma one starts with an integrality gap instance for the canonical SDP relaxation of the problem of interest, say QUADRATIC PROGRAMMING, and its vector solution \( v_0, v_1, \ldots, v_n \). Given that the instance is an integrality gap instance one knows that the objective value attained by \( v_0, v_1, \ldots, v_n \) is much bigger than what is attainable by any integral solution. The simplest thing one can hope for is to show that this vector solution is also a solution to the level-\( k \) Sherali-Adams SDP relaxation, or in light of Fact 2.2 the vectors are compatible with a family of local distributions \( \{D(S)\}_{S \subseteq \binom{[n]}{\leq k}} \). However, this is not generally possible. Instead we use the lemma (in the simplest case with \( \nu \) being the uniform distribution on \( \{-1, 1\}^n \)) to get another set of vectors \( u_0, \ldots, u_n \) which are in fact compatible with some \( \{D(S)\}_{S \subseteq \binom{[n]}{\leq k}} \). Given that in the problems we consider the objective function is a quadratic polynomial, and given the promise of the lemma that inner products of \( u_i \)'s is \( \epsilon \) times that of \( v_i \)'s, it follows that the objective value attained by \( u_0, \ldots, u_n \) is \( \epsilon \) times that attained by \( v_0, \ldots, v_n \). As \( v_0, \ldots, v_n \) are the vector solution for an integrality gap instance, it follows that the integrality gap will decrease by a multiplicative factor of at most \( \epsilon \) when one goes from the canonical SDP relaxation to the level-\( k \) Sherali-Adams SDP relaxation.

How big one can take \( \epsilon \) while satisfying the requirements of the lemma will determine the quality of the resulting integrality gap. In the simplest case one can take \( \nu \) to be the uniform distribution on \( \{-1, 1\}^n \) and argue that in this case the requirements of the lemma
are satisfied as long as $2\epsilon k^2 \leq 1$ and in particular for $\epsilon = 1/2k^2$. In fact our application to the Max Cut Gain problem detailed in section 5.3 follows this simple outline. For Quadratic Programming we can get a better result by taking a close look at the particular structure of $v_i$’s for the integrality gap instance of [KO06] and using a more appropriate distribution for $\nu$.

We will now prove Lemma 5.3.

Proof of Lemma 5.3. Our proof proceeds in two steps. First, we construct a family of functions $\{f_S : \{-1,1\}^S \rightarrow \mathbb{R}\}_{S \subseteq [n]}$ that satisfy all the required conditions except being the probability mass function of distributions. In particular, while for any $S$ the sum of the values of $f_S$ is 1, this function can take negative values for some inputs. The construction of $f_S$ uses the distribution $\nu$ and guarantees that while $f_S$ may take negative values, these values are not too negative. In the second step we take a convex combination of the $f_S$’s for the integrality gap instance of [KO06] and using a more appropriate distribution for $\nu$.

For any subset $S \subseteq [n]$ we define $f_S$ as a “hybrid” object that is using both the vectors $\{v_i\}$’s and the distribution $\nu$. Given that a function is uniquely determined by its Fourier expansion we will define $f_S$ in terms of its Fourier expansion,

$$\widehat{f_S}(\emptyset) = 2^{-|S|} = 2^{n-|S|}\widehat{\nu}(\emptyset),$$
$$\forall i \in S \quad \widehat{f_S}(\{i\}) = 2^{-|S|}v_0 \cdot v_i,$$
$$\forall i \neq j \in S \quad \widehat{f_S}(\{i,j\}) = 2^{-|S|}v_i \cdot v_j,$$
$$\forall T \subseteq S, |T| > 2 \quad \widehat{f_S}(T) = 2^{n-|S|}\widehat{\nu}(T).$$

Comparing the above definition with (2.23), $f_S$ is exactly like the marginal distribution of $\nu$ on the set $S$ except it has different degree one and degree two Fourier coefficients. First, observe that for any $S$, the sum of the values of $f_S$ is 1.

$$\sum_{x \in \{-1,1\}^S} f_S(x) = 2^{|S|} \mathbb{E}_x[f_S(x)] = 2^{|S|}\widehat{f_S}(\emptyset) = 1.$$  
Then observe that by (2.23), for all $U \subseteq T \subseteq S$,

$$\text{Mar}_T f_S(U) = 2^{|S|-|T|}\widehat{f_S}(U) = \widehat{f_T}(U).$$
So, $f_S$ satisfies (2.14). Now observe that,

$$\sum_{x \in \{-1,1\}^S} f_S(x)x_i = 2^{|S|} \mathbb{E}_x[f_S(x)\chi_{\{i\}}(x)] = 2^{|S|}\widehat{f_S}(\{i\}) = v_0 \cdot v_i,$$
$$\sum_{x \in \{-1,1\}^S} f_S(x)x_i x_j = 2^{|S|} \mathbb{E}_x[f_S(x)\chi_{\{i,j\}}(x)] = 2^{|S|}\widehat{f_S}(\{i,j\}) = v_i \cdot v_j.$$
So, $f_S$’s satisfy (2.15) and (2.16) and are compatible with $v_i$’s (except, they are not distribu-
Next we show that $f_S(y)$ cannot be too negative.

$$f_S(y) = \sum_{T \subseteq S} \hat{f}_S(T) \chi_T(y) = 2^{n-|S|} \sum_{T \subseteq S} \hat{\nu}(T) \chi_T(y) + \sum_{T \subseteq S} (\hat{f}_S(T) - 2^{n-|S|} \hat{\nu}(T)) \chi_T(y)$$

$$= (\text{Mar}_S \nu)(y) + \sum_{T \subseteq S} (\hat{f}_S(T) - 2^{n-|S|} \hat{\nu}(T)) \chi_T(y) \geq - \sum_{T \subseteq S} |\hat{f}_S(T) - 2^{n-|S|} \hat{\nu}(T)|$$

$$= - \sum_{i \in S} |\hat{f}_S(\{i\}) - 2^{n-|S|} \hat{\nu}(\{i\})| - \sum_{i \neq j \in S} |\hat{f}_S(\{i, j\}) - 2^{n-|S|} \hat{\nu}(\{i, j\})|$$

$$= -2^{-|S|} \left( \sum_{i \in S} |v_0 \cdot v_i - \mathbb{E}_{x \sim \nu} [x_i] | + \sum_{i \neq j \in S} |v_i \cdot v_j - \mathbb{E}_{x \sim \nu} [x_i x_j] | \right)$$

$$\geq -2^{-|S|} |1| 2\epsilon,$$

where the second line follows from (2.24) and the last step follows from the condition on $\epsilon$ and because $|S| \leq k$. This completes the first step of the proof.

Next, define $\pi$ to be the uniform distribution on $\{-1, 1\}^n$ and $D(S)$ as a convex combination of $f_S$ and $\text{Mar}_S \pi$, i.e.,

$$\forall y \in \{-1, 1\}^S \quad D(S)(y) = \epsilon f_S(y) + (1 - \epsilon)(\text{Mar}_S \pi)(y).$$

$$\forall i \in [n] \quad u_i = \sqrt{\epsilon} \cdot v_i + \sqrt{1 - \epsilon} \cdot w_i.$$

Here, $w_i$'s are defined such that they are perpendicular to all $v_i$'s and each other.

It is easy to check that $u_i$'s are compatible with $D(S)$'s and satisfy all the required properties of the lemma, except that $D(S)$'s can potentially be negative.) In fact (2.14-2.16) are linear conditions and given that they hold for $\{f_S\}$ and $\{v_i\}$ and for $\{\text{Mar}_S \pi\}$ and $\{w_i\}$ they must hold for $\{D(S)\}$ and $\{u_i\}$. Finally, for any $S \in \binom{[n]}{\leq k}$ and $y \in \{-1, 1\}^S$ we have

$$D(S)(y) = \epsilon f_S(y) + (1 - \epsilon)(\text{Mar}_S \pi)(y) \geq -\epsilon 2^{-|S|}/2\epsilon + (1 - \epsilon)2^{-|S|} = 2^{-|S|}(\frac{1}{2} - \epsilon) \geq 0. \quad \blacksquare$$

**Remark.** Some of the conditions in Lemma 5.3 can readily be relaxed. First, notice that the choice of $\pi$ as the uniform distribution is not essential. The only property that was needed for the argument to work was that the marginal distribution of $\pi$ over $S$ assigns positive probability to every member of $\{-1, 1\}^S$ (the existence of compatible $w_i$'s is also required but is true for any distribution on $\{-1, 1\}^n$.) More precisely there is a positive function $\delta(k)$ so that,

$$\forall y \in \{-1, 1\}^S \quad \Pr_{x \sim \text{Mar}_S \pi} [x = y] \geq \delta(k).$$

One would need a stronger condition on $\epsilon$ in this case that depends on $\delta(k)$. The inner product

\[3\]Note that for a distribution $D(S)$ we have $\mathbb{E}_{x \sim D(S)} [x_i] = \sum_{x \in \{-1, 1\}^S} D(S)(x) x_i$. Hence, the above sums are relevant to (2.15) and (2.16).
of the resulting vectors would of course depend on the distribution \( \pi \), namely,

\[
\mathbf{u}_0 \cdot \mathbf{u}_i = \epsilon \mathbf{v}_0 \cdot \mathbf{v}_i + (1 - \epsilon) \mathbb{E}_{x \sim \pi} [x_i],
\]

\[
\mathbf{u}_i \cdot \mathbf{u}_j = \epsilon \mathbf{v}_i \cdot \mathbf{v}_j + (1 - \epsilon) \mathbb{E}_{x \sim \pi} [x_i x_j].
\]

Another observation is that \( \nu \) and \( \pi \) do not have to be true (global) distributions on \( \{-1, 1\}^n \).
Instead, we can start with two families of local distributions, \( \{\pi_S\}, \{\nu_S\} \) on sets of size at most \( k \) and a set of vectors \( \mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_n \) such that \( \pi_S \)'s are compatible with \( \mathbf{w}_i \)'s.

Unfortunately, we are not aware of a way to use either of these two extensions to prove better integrality gap lower bounds.

### 5.2 Application to Quadratic Programming

In this section we prove that the level-\( k \) Sherali-Adams SDP relaxation of QUADRATIC PROGRAMMING has integrality gap \( \Omega(\log N) \), i.e. Theorem 5.1. Showing an integrality gap of \( \Omega(\log N) \) for the canonical SDP relaxation is equivalent to showing that for any large enough \( N \) there exists a matrix \( A_{N \times N} \) and value \( \xi \), such that the following hold; (a) for all \( x \in \{-1, 1\}^N \), \( \sum_{i \neq j} a_{ij} x_i x_j \leq O(\xi/\log(N)) \), (b) there exist unit vectors \( \{\mathbf{v}_i\} \) such that \( \sum_{i \neq j} a_{ij} \mathbf{v}_i \cdot \mathbf{v}_j \geq \Omega(\xi) \).

In order to show an integrality gap for the level-\( k \) Sherali-Adams SDP relaxation one in addition needs to show that \( \{\mathbf{v}_i\} \)'s are compatible with a set of local distributions \( \{\mathcal{D}(S)\}_{S \subseteq \llbracket n \rrbracket_{\leq k}} \).

As discussed earlier we will start with the integrality gap instance of [KO06] and apply Lemma 5.3. The following is a fairly immediate corollary of Theorem 4.4 and Proposition 2.2 in [KO06]. It’s worth noting that Khot and O’Donnell are mainly interested in the existence of integrality gap instances, and the matrix \( A \) is implicit in their work.

**Corollary 5.4** (Theorem 4.4 and Proposition 2.2 from [KO06]). Let \( \xi > 0 \) be sufficiently small and let \( d = 1/\xi^3 \) and \( n = \Theta(d^7) \). Further, let \( m = \frac{1}{\xi} \log(\frac{1}{\xi}) = \Theta(n^{1/21} \log n) \) and \( N = nm \).

Choose \( n \) vectors \( \mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbb{R}^d \) according to the \( d \)-dimensional Gaussian distribution. Then one can define \( A_{N \times N} \) as a function of \( \mathbf{u}_i \)'s such that almost surely the following two conditions hold:

1. \( \sum_{i \neq j} a_{ij} x_i x_j \leq O(\xi/\log(1/\xi)) \) for all \( x \in \{-1, 1\}^N \).
2. There exist unit vectors \( \mathbf{v}_i \) such that \( \sum_{i \neq j} a_{ij} \mathbf{v}_i \cdot \mathbf{v}_j \geq \Omega(\xi) \).

Furthermore, the \( \mathbf{v}_i \)'s are produced in this simple manner based on \( \mathbf{u}_i \)'s. Divide the \( N \) variables into \( n \) classes of size \( m \) each, and assign the vector \( \mathbf{u}_j/\|\mathbf{u}_j\| \) to the variables in the \( j \)th class. Formally, \( \mathbf{v}_i = \mathbf{u}_{[i/m]}/\|\mathbf{u}_{[i/m]}\| \).

We need the following property of the \( \mathbf{v}_i \)'s which follows from well-known properties of random unit vectors. We leave the proof for Section 5.4.
Fact 5.5. In the SDP solution of [KO06], with probability at least $1 - 4/n^4$, for all pairs of indices $1 \leq i, j \leq N$ the inner product $v_i \cdot v_j$ has the following property,

$$v_i \cdot v_j = 1 \quad \text{if } i \text{ and } j \text{ are in the same class, i.e. } [i/m] = [j/m],$$

$$|v_i \cdot v_j| \leq \sqrt{(12 \log n)/d} \quad \text{if } i \text{ and } j \text{ are in different classes, i.e. } [i/m] \neq [j/m].$$

Recall that in Lemma 5.3 a choice of a distribution $\nu$ on $\{-1, 1\}^n$ is required. In particular, if for every pair of variables $i, j$, $E_{x \sim \nu} [x_i x_j]$ is close to $v_i \cdot v_j$ one can choose a big value for $\epsilon$ in the lemma, which in turn means that the resulting $u_i$'s will have inner products close to those of $v_i$'s.

Indeed, the key to Theorem 5.1 is using $\nu$ which is “agreeable” with fact 5.5: two variables will have a large or a small covariance depending on whether they are from the same or different classes. Luckily, this is easily achievable by identifying variables in the same class and assigning values independently across classes. In other words the distribution $\nu$ will choose a random value from $\{-1, 1\}$ for $x_1 = x_2 = \cdots = x_m$, an independently chosen value for $x_{m+1} = \cdots = x_{2m}$, and similarly an independently chosen value for $x_{nm-m+1} = \cdots = x_{nm}$. Such $\nu$ clearly satisfies,

$$E_{x \sim \nu} [x_i x_j] = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are in the same class,} \\ 0 & \text{otherwise.} \end{cases}$$

Consider the vector solution of [KO06], $v_1, \ldots, v_{nm}$ and define $v_0$ as a vector perpendicular to all other $v_i$'s. Consider the distribution $\nu$ defined above and apply Lemma 5.3 for $v_0, \ldots, v_i, \nu, k = d^{0.2}$, and $\epsilon = 1/2$. By Fact 5.5 the inner products of the $v_i$ vectors are close to the corresponding correlations of the distribution $\nu$. It is easy to check that the conditions of Lemma 5.3 are satisfied,

$$2\epsilon k^2 |v_0 \cdot v_i - E_{x \sim \nu} [x_i]| = 2\epsilon k^2 |0 - 0| = 0,$$

$$2\epsilon k^2 |v_i \cdot v_j - E_{x \sim \nu} [x_i x_j]| \leq d^{0.4} \sqrt{(12 \log n)/d} = \sqrt{12 \log n}/d^{0.1} \ll 1,$$

and the lemma applies for large enough $n$ thus by Fact 2.2 the resulting vectors, $u_i$'s, are a solution to the level-$k$ Sherali-Adams SDP relaxation of Quadratic Programming. It is now easy to see that (5.2) implies a big objective value for this solution,

$$\sum_{i \neq j} a_{ij} u_i \cdot u_j = \epsilon \sum_{i \neq j} a_{ij} v_i \cdot v_j \geq \Omega(\xi).$$

It remains to estimate the value of $k$ in terms of $N$:

$$k = d^{0.2} = \Theta(n^{1/35}) = \Theta(N^{21/770}/\log^{1/35} N) = \Omega(N^{1/37}),$$

and we conclude,
Theorem 5.1 (restated). For $\delta = 1/37$, the level-$N^\delta$ Sherali-Adams SDP relaxation of Quadratic Programming has integrality gap $\Omega(\log N)$.

5.3 Application to Max Cut Gain

The Max Cut Gain problem is an important special case of the Quadratic Programming problem where all entries of the matrix $A$ are nonpositive. In other words the input is a nonpositive $n$ by $n$ matrix $A$ and the objective is to find $x \in \{-1,1\}^n$ that maximizes the quadratic form $\sum_{i \neq j} a_{ij} x_i x_j$. This problem gets its name and main motivation from studying algorithms for the Max Cut problem that perform well for graphs with maximum cut value close to half the edges. See [KO06] for a discussion.

Naturally, constructing integrality gap instances for Max Cut Gain is harder than Quadratic Programming. The best integrality gap instances are due to Khot and O’Donnell [KO06] who, for any constant $\Delta$, construct an instance of integrality gap at least $\Delta$. The following is a restatement of their Theorem 4.1 tailored to our application.

Theorem 5.6 (Theorem 4.1 from [KO06]). The standard SDP relaxation for Max Cut Gain has super constant integrality gap. Specifically, for any constant $\xi > 0$, there is a big enough $n$ and a matrix $A_{n \times n}$ such that,

1. $\sum_{i \neq j} a_{ij} x_i x_j \leq \xi / \log \frac{1}{\xi}$ for all $x \in \{-1,1\}^n$.

2. There are unit vectors $v_1, \ldots, v_n$ such that $\sum_{i \neq j} a_{ij} v_i \cdot v_j \geq \Omega(\xi)$.

It should be mentioned that the proof of [KO06] is continuous in nature and it is not entirely clear how $n$ grows as a function of $1/\xi$. However, an integrality gap of $f(n)$ for some function $f(n) = \omega(1)$ is implicit in Theorem 5.6.

Consider the instance from Theorem 5.6 and the function $f(n)$ as above and let $g(n) = \sqrt[3]{f(n)}$. We know that for every $n$, there are unit vectors $v_1, \ldots, v_n$ such that $\sum_{i \neq j} a_{ij} v_i \cdot v_j \geq \Omega(\xi)$. Let $v_0$ be a unit vector perpendicular to all $v_i$’s and set $k = g(n)$, $\epsilon = g(n)^{-2}/2$ and let $\nu$ be the uniform distribution on $\{-1,1\}^n$. Note that for all $i < j$,

$$2\epsilon k^2 |v_0 \cdot v_i - \mathbb{E}_{x \sim \nu}[x_i]| = |v_0 \cdot v_i| = 0,$$

$$2\epsilon k^2 |v_i \cdot v_j - \mathbb{E}_{x \sim \nu}[x_i x_j]| = |v_i \cdot v_j| \leq 1,$$

hence the conditions of Lemma 5.3 hold. Consequently, there are vectors $u_0, u_1, \ldots, u_n$ compatible with a family of local distributions $\{D(S)\}_{S \subseteq \{n\}}$ on subsets of size up to $k = g(n)$.
which satisfy (5.2). Now,
\[ \sum_{i \neq j} a_{ij} u_i \cdot u_j = \epsilon \sum_{i \neq j} a_{ij} v_i \cdot v_j \geq \Omega(\xi/g(n)^2), \]
\[ \forall x \in \{-1, 1\}^n \sum_{i \neq j} a_{ij} x_i x_j \leq O(\xi/f(n)) = O(\xi/g(n)^3) \]
and we obtain,

**Theorem 5.2** (restated). There exists a function \( g(n) = \omega(1) \), such that the level-\( g(n) \) Sherali-Adams SDP relaxation of Max Cut Gain has integrality gap \( \Omega(g(n)) \).

### 5.4 Proof of Fact 5.5

We will need the following well known isoperimetric inequality for the unit sphere. It states that most of the area of \( S^{d-1} \) is concentrated around the equator.

**Lemma 5.7.** For any unit vector \( v \in S^{d-1} \), if a unit vector \( x \in S^{d-1} \) is chosen uniformly at random, \( |v \cdot x| \) is sharply concentrated:
\[
\Pr_x[|v \cdot x| \geq t/\sqrt{d}] \leq 4e^{-t^2/2}.
\]

**Proof.** Define,
\[
f(x) \overset{\text{def}}{=} v \cdot x,
\]
and apply Lévy’s lemma (see Theorem 14.3.2 of [Mat02]) observing that \( f(x) \) is 1-Lipschitz. We will have,
\[
\Pr[|v \cdot x| \geq t/\sqrt{d}] = \Pr[|f(x) - \text{median}(f)| \geq t/\sqrt{d}] \leq 4e^{-(t/\sqrt{d})^2d/2} = 4e^{-t^2/2}.
\]

Now, proving the fact is a matter of looking at the actual definition of the solution vectors and applying lemma 5.7.

**Fact 5.5** (restated). In the SDP solution of [KO06], with probability at least \( 1 - 4/n^4 \), for all pairs of indices \( 1 \leq i, j \leq N \) the inner product \( v_i \cdot v_j \) has the following property,
\[
\begin{align*}
v_i \cdot v_j &= 1 & \text{if } i \text{ and } j \text{ are in the same class}, \\
|v_i \cdot v_j| &\leq \sqrt{(12 \log n)/d} & \text{if } i \text{ and } j \text{ are in different classes}.
\end{align*}
\]

**Proof.** The first case follows from the definition of \( v_i \)'s. For the second case \( v_i \) and \( v_j \) are independent \( d \)-dimensional vectors distributed uniformly on \( S^{d-1} \). Consider a particular choice
of \( v_i \), according to lemma 5.7,

\[
\Pr_{v_j} \left[ |v_i \cdot v_j| \geq \sqrt{\frac{12 \log n}{d}} \right] \leq 4 e^{-6 \log n} = 4n^{-6}.
\]

Applying union bound on all \( n^2 \) pairs of classes shows that the condition of the lemma holds for all pairs with probability at least \( 1 - n^2 4n^{-6} = 1 - 4/n^4 \).
Chapter 6

Tight Gaps for Vertex Cover in the Sherali-Adams SDP Hierarchy

In this chapter we present a tight integrality gap for the Sherali-Adams SDP relaxation of Vertex Cover. More precisely, we show that for any \( \epsilon > 0 \), the level-5 Sherali-Adams SDP relaxation of Vertex Cover has integrality gap at least \( 2 - \epsilon \). To the best of our knowledge this is the first nontrivial tight integrality gap for the Sherali-Adams SDP hierarchy for a combinatorial problem with hard constraints.

For our proof we introduce a new tool to establish Local-Global Discrepancy which uses simple facts from high-dimensional geometry. This allows us to give solutions to the level-\( k \) Sherali-Adams LP relaxation of Vertex Cover with objective value \( n(1/2 + o(1)) \) for any \( n \) vertex subgraph\(^2\) of the Borsuk graph \( B^m_{o(1)} \) for any constant \( k \). Since such graphs with no linear size independent sets exist, this immediately gives a tight integrality gap for any constant level Sherali-Adams LP relaxation of Vertex Cover. Interestingly, Vishwanathan [Vis09], shows that a \( 2 - \epsilon \)-approximation algorithm for Vertex Cover in subgraphs of \( B^m_{o} \) implies a \( 2 - \epsilon^2/4 \)-approximation algorithm for Vertex Cover in all graphs. Furthermore, any hard instance of Vertex Cover should have large subgraphs which are also subgraphs of \( B^m_{o} \).

In order to extend our solution to the level-\( k \) Sherali-Adams SDP relaxation of Vertex Cover, we reduce the extra semidefiniteness requirement of the solution to a condition on the Taylor expansion of a reasonably simple function that we are able to establish up to \( k = 5 \). We conjecture that this condition holds for any constant \( k \) which, if true, would imply that our solution is valid for any constant level Sherali-Adams SDP relaxation.

Motivation for studying Vertex Cover is two-fold. For one thing it is arguably one of the simplest \textbf{NP}-hard problems whose inapproximability remains unresolved. But more importantly, studying Vertex Cover has introduced some very important techniques both in terms of approximation algorithms and hardness of approximation with [DS05] being a prime example. Intuitively this is, at least partly, due to the “hard constraints” of Vertex Cover,

\(^1\)Results in this chapter appear in [BCGM11]
\(^2\)In other words any graph with vector chromatic number \( 2 + o(1) \)
that is the solution has to satisfy a number of inflexible constraints (the edge constraints.) As many of the standard techniques for proving hardness of approximation and integrality gaps produce solutions which satisfy most constraints in an instance, showing tight hardness for Vertex Cover has remained unresolved, unless one assumes UGC; see [KR08].

Our main result is as follows.

**Theorem 6.1.** For every \( \epsilon > 0 \), the level-5 Sherali-Adams SDP relaxation of Vertex Cover has integrality gap \( 2 - \epsilon \).

While tight integrality gaps for weaker or incomparable systems were known, there were no good candidates for Sherali-Adams SDP integrality gap solutions. In particular, while integrality gaps for the closely related but weaker Sherali-Adams LP system for Vertex Cover were known [CMM09], the solution there does not satisfy the required positive semidefiniteness condition. As we explain below, apart from the significance of our new SDP integrality gap, we also believe that our proofs are interesting in their own right.

On our way to prove the above theorem we need to define new solutions for Sherali-Adams LP relaxations of Vertex Cover. One of our contributions is an intuitive and geometric explanation of why this large family of LPs are fooled by a certain family of graphs, the so-called Borsuk graphs. This yields a tight integrality gap lower bound for level-\( \Omega(\sqrt{\log n / \log \log n}) \) Sherali-Adams LP relaxation of Vertex Cover; see Theorem 6.6. Other than being used in our proof of Theorem 6.1, our solution is arguably simpler and more intuitive than the integrality gap of [CMM09] for the same system.\(^3\)

The heart of the problem in showing integrality gaps for Sherali-Adams SDPs is that the proposed solution needs to satisfy a strong positive-semidefiniteness condition. Toward establishing Theorem 6.1, we show how to reduce this condition into a clean analytic statement about a certain function parameterized by \( t \). We are able to show that this analytic statement holds up to \( t = 5 \), hence the level-5 Sherali-Adams SDP gap. We have strong evidence (both theoretical and experimental) that the aforementioned analytic statement holds for any constant value of \( t \), which we explicitly state as a conjecture in Section 6.3.3. To sum up, we have the following second theorem.

**Theorem 6.2.** Assuming Conjecture 6.18, for every constant \( \epsilon > 0 \) and \( t \in \mathbb{N} \), the level-\( t \) Sherali-Adams SDP relaxation for Vertex Cover has integrality gap \( 2 - \epsilon \).

For a brief discussion of the validity of Conjecture 6.18 see Remark 6.3.3.

**Known integrality gaps for Vertex Cover**

Considerable effort has been invested in strong lower bounds for various hierarchies for Vertex Cover. For LP hierarchies, Schoenebeck, Trevisan, and Tulsian [STT07a] show an integrality gap of \( 2 - \epsilon \) for level-\( \Omega(n) \) Lovász-Schrijver LP relaxation of the problem and Charikar,
Makarychev, and Makarychev [CMM09] show the same integrality gap for the stronger Sherali-Adams LP system up to level $\Omega(n^\delta)$ (with $\delta$ going to 0 together with $\epsilon$.) For SDP hierarchies, and in particular for the Lovász-Schrijver SDP system which is stronger than both the LS system and the canonical SDP formulation but incomparable to Sherali-Adams LP/SDP, Georgiou et al. [GMPT10] show an integrality gap of $2 - \epsilon$ for level $\Omega(\sqrt{\log n/\log \log n})$.

The integrality gap of two stronger hierarchies for VERTEX COVER, on the other hand, is open. The first is the Sherali-Adams SDP system which is stronger than the LS system, and the subject of this chapter. The second is the Lasserre system, for which no tight integrality gap for VERTEX COVER is known.\footnote{In fact there are only a few combinatorial problems for which tight Lasserre integrality gaps are known; see [Sch08] and [Tul09] for some notable exceptions.} If one is content with an integrality gap less than 2, Tulsiani [Tul09] proves an integrality gap of $\frac{7}{6}$ for level $\Omega(n^\delta)$ and Schoenebeck [Sch08] an integrality gap of $\sqrt{\gamma}$ for level $\Omega(n)$ of the Lasserre system.

**Organization**

The rest of this chapter is organized as follows. In Section 6.1 we give an outline of our Sherali-Adams SDP solution and its proof of correctness as well as a brief comparison of our techniques with those of the previous work. We present our integrality gap lower bound for the Sherali-Adams LP relaxation of VERTEX COVER in Section 6.2. While such integrality gaps were previously known, ours is much more intuitive. Furthermore, we will use this integrality gap toward our lower bound for Sherali-Adams SDP relaxation of VERTEX COVER. Our Sherali-Adams SDP solution and its proof is presented in Section 6.3.

### 6.1 Outline of Our Method and Comparison to Previous Work

The instance we use for our Sherali-Adams SDP integrality gap is the Frankl-Rödl graph, $G^m_\gamma$; see Definition 2.11. By Theorem 2.3, for $\gamma = \sqrt{\log m/m}$, $G^m_\gamma$ has no vertex cover smaller than $2^m(1 - o(1))$. A tight integrality gap therefore calls for a solution of objective value at most $2^m(1/2 + \epsilon)$, for arbitrarily small $\epsilon > 0$.

Consider the following experiment used to define our solution. A geometric way to obtain a distribution of vertex covers would be to embed $G^m_\gamma$ on the unit sphere and take a sufficiently large spherical cap centered at a random point on the sphere. Of course, given the Frankl-Rödl theorem mentioned above, in doing so we have not achieved much since we are defining a global distribution of vertex covers, and thus its expected size has to be at least $2^m(1 - o(1))$. However, it is useful to understand why these vertex covers are big from a geometric point of view: the height of the spherical cap must be at least $1 + \sqrt{\gamma}$ (as opposed to 1 for a half-sphere.) Now concentration of measure on the sphere (i.e. Lemma 5.7) implies that because $\sqrt{\gamma m} = \omega(1)$ the area of such a cap is a $1 - o(1)$ fraction of the whole sphere. So the probability that any vertex of the graph is in the cap is $1 - o(1)$, which is very large. Had it been the case that
\( \sqrt{\gamma m} = o(1) \) concentration of measure would imply that the area of the cap is \( 1/2 + o(1) \) of that of the sphere and we would have had a small vertex cover.

The main idea is that based on Fact 2.1 one only needs to define probabilities for small sets, i.e. sets of size up to \( t \) if the goal is to show integrality gaps for the level-(\( t - 1 \)) Sherali-Adams LP relaxation. So one can first embed the points in the set in a small dimensional sphere and then repeat the above experiment to define a random vertex cover. The spherical caps that are required in order to cover the edges in these sets have the same height, but now, due to the lower dimension, their area is greatly reduced! Specifically, if the original set has at most \( t \) points, the experiment can be performed in a \( t \)-dimensional sphere and if \( \sqrt{\gamma t} = o(1) \), the probability of any vertex participating in the vertex cover will be no more than \( 1/2 + o(1) \). In particular, \( t = o(\sqrt{m}/\log m) \) would suffice.

It is critical for Fact 2.1 of course that the obtained distributions are consistent. But this is “built-in” in this experiment. Indeed, due to spherical symmetry, the probability that a set of points on a \( t \)-dimensional sphere belong to a random cap of a fixed radius depends only on \( t \), the radius of the cap and the pairwise Euclidean distances of the points in the set. Interestingly this construction works for any graph with vector chromatic number \( 2 + o(1) \). In other words, if \( G \) is an \( n \) vertex graph that can be embedded into the unit sphere so that the end points of any edge are almost antipodes (i.e. a subgraph of the Borsuk graph \( B_{o(1)}^m \)), then there is a sufficiently "low-level" (but non-trivial) Sherali-Adams LP solution of value \( (1/2 + o(1))n \).

Unfortunately, we cannot show that the above solution satisfies the extra constraints imposed by the Sherali-Adams SDP system. Instead we change our solution in several ways to attain positive semidefiniteness. These changes are somewhat technical and we avoid discussing them in detail here. At a high level the changes are (i) we add a small probability of picking the whole graph as the vertex cover. (ii) We apply a transformation of the canonical embedding of the cube in the sphere that ensures that the farthest pairs of vertices are precisely the edges, and also that the inner products have a bias to being positive. This is in contrast to the canonical embedding in which the average inner product is 0.

To get some insight into the rationale of these modifications, first note that the matrix whose positive definiteness we need to prove happens to be highly symmetric. For such symmetric matrices a necessary condition for positive semi-definiteness is that the average entry is at least as large as the square of the diagonal entries. Step (i) above is precisely the tool we need to ensure this condition, and has no adverse effect otherwise. The second transformation is useful although not clearly necessary. We can, however, argue that without a transformation of this nature, a good SDP solution is also possible for the graph in which edges connect vertices of Hamming distance at least (rather than exactly) \( m(1 - \gamma) \). The existence of solutions for such a dense graph seems intuitively questionable. Last, boosting the typical inner product can be shown to considerably boost the Taylor coefficients of a certain function which we need to show only has positive Taylor coefficients.
Comparison to Previous Work

There are more than half a dozen different integrality gap constructions for Vertex Cover in different Lift-and-Project systems known. Among these the most relevant to our work is the work of Charikar, Makarychev, and Makarychev [CMM09]. In [CMM09] a Sherali-Adams LP solution is presented which is based on embedding the vertices of the graph in the sphere. The similarity with our results is that Charikar et al. take a special case of caps, i.e. half-spheres, in order to determine probabilities. Consistency of these distribution is, just as in our case, guaranteed by the fact that these probabilities are intrinsic to the local distances of the point-set in question. However, the reason that these distributions behave differently than a global distribution (which is essential for an integrality gap construction) is completely different than ours. It is easy to see that when the caps in the construction are half spheres, the dimension does not play a role at all. However, in [CMM09] there is no global embedding of the points in the sphere but rather only a local one. In contrast, our distributions can be defined for all dimensions, however as we mentioned we must keep the dimension reasonably small in order to guarantee small objective value. Another big difference pertains to the different instances. While our construction may very well be (close to) the one that will give a Lasserre integrality gap bound, the instances of [CMM09] have no substantial integrality gap even for the standard SDP relaxation. Thus their result cannot be extended to the stronger Lasserre or Sherali-Adams SDP hierarchies.

It is also important to put the results of this chapter in context with the sequence of results dealing with SDP integrality gaps of Vertex Cover [GK98, Cha02, GMPT10, GMT08, GMT09a]. In these works the solution can be thought of as an approximation to a very simple set: a dimension cut, that is a face of the cube. This set is not a vertex cover, but in some geometric sense is close to one. The SDP solutions are (essentially) averagings of such dimension-cuts with some carefully crafted perturbations. Using the same language, the solution we present is based on Hamming balls of radius $m/2$ (i.e. translations of the majority function) rather than dimension-cuts (i.e. dictatorship functions.) The perturbation we apply to make such a solution valid is simply the small increase in the radius of the Hamming balls. Another distinction is that while all previous results use tensoring to construct their solutions, we mainly use it to certify its positive semidefiniteness. In other words, our solutions are defined geometrically and then tensoring is used to give an alternative view which helps to show they have the required positive semidefiniteness.5

6.2 Fooling LPs derived by the Sherali-Adams System

In this section we study the Sherali-Adams LP relaxation of Vertex Cover through the language of Definition 2.9 and Fact 2.1 for finite subgraphs of $B^m_n$ on $n$ vertices. In particular, for

---

5We will defer the definition of tensoring and how it can be used in establishing positive semidefiniteness of matrices to Subsection 6.3.1 on page 80
Chapter 6. Tight Gaps for Vertex Cover

77
every set $U \subseteq [n]$ we define a distribution of vertex covers that are locally consistent.

The family of distributions we are looking for arises from the following experiments. Fix a discrete subgraph $G = (V, E)$ of $B^m_\delta$ on $n$ vertices for which we want to construct a level-$t$ Sherali-Adams LP solution with small objective value. Given that $G = (V, E)$ is a subgraph of $B^m_\delta$ we can think of its vertices as points on $S^{m-1}$ and in particular talk about their Euclidean distances. The local distributions, $D(I)$, are defined by an experiment shown as Algorithm 5 that samples from them.

**Algorithm 5:** Local-Global Experiment

**Input:** A subgraph $G = (V, E)$ of $B^m_\delta$, $I \in \binom{V}{\leq t}$.

**Output:** $x \sim \{0, 1\}^I$ sampled according to $D(I)$.

**Parameter:** $1 \leq t \leq n$

1. Embed the $I$-induced subgraph of $G$ into $S^{t-1}$ preserving all Euclidean distances,
2. $C \leftarrow$ a random spherical cap of height $1 + \sqrt{\delta}$ in $S^{t-1}$,
3. for $v_i \in I$ do
   4. if $v_i \in C$ then $x_i \leftarrow 1$ else $x_i \leftarrow 0$
4. end

Notice that line 1 is possible because $|I| \leq t$.

**Lemma 6.3.** For every finite subgraph of $B^m_\delta$ on $n$ vertices, the family of distributions $\{D(I)\}_{I \in \binom{[n]}{\leq t}}$ is a family of consistent local distributions of vertex cover and by Fact 2.1 a solution of the level-$t-1$ Sherali-Adams LP relaxation of Vertex Cover.

**Proof.** The local consistency (Definition 2.9) follows from the following simple geometric fact: the probability distribution $D(I)$ only depends on the pairwise Euclidean distances of vertices in $I$ and the parameter $t$. Given this simple observation it is not hard to see that $D(U_1)$ is just the marginal of $D(U)$ when $U_1 \subseteq U$.

It therefore remains to argue that $D(I)$ is a distribution of vertex covers, i.e. they satisfy 2.19 for the edge constraints. This condition simply translates that if $ij \in E$ and $i, j \in I$ then one of $x_i$ and $x_j$ will always be assigned 1 by $D(I)$, i.e. one of $v_i, v_j$ is in $C$. This is true simply because the cap is big enough. In particular, for any two vertices $i, j$ outside the cap if $z_i, z_j$ are their vectors and $w$ is the vector corresponding to the tip of the cap, $w \cdot z_i, w \cdot z_j > \sqrt{\delta}$ which implies $\|z_i + z_j\| = \|w\| \|(z_i + z_j)\| \geq w \cdot (z_i + z_j) > 2\sqrt{\delta}$, where we have used Cauchy-Schwarz for the first inequality. Since $G$ is a subgraph of $B^m_\delta$, we conclude that $ij$ cannot be an edge.

All that remains is to show that the objective value of our solution is indeed small. In fact, we can show a stronger statement, not only is the objective value $n/2 + o(n)$ but each vertex contributes $1/2 + o(1)$ to the objective value. In particular we can show the following lemma.
Lemma 6.4. For any fixed $z \in S^{t-1}$ and $t \geq 8$, we have
\[ \Pr_{w \in S^{t-1}} [w \cdot z \leq \eta] \leq \frac{1}{2} + 8\eta\sqrt{t}, \]
when $w$ is distributed uniformly on $S^{t-1}$. Consequently, for any vertex $i \in I$ of the graph $G$ (subgraph of $B^n_{m\delta}$), we have
\[ \Pr_{S \sim D(I)} [i \in S] \leq \frac{1}{2} + 8\sqrt{\delta t}. \]

Proof. A vector $w \in S^{t-1}$ defines the complement of a spherical cap of height $1 - \eta$. We therefore need to determine the ratio between the measure of a cap over the measure of the $t$-dimensional sphere.

Denote by $\mu(S^{t-1})$ the surface area of a spherical cap of $S^{t-1}$, with height $1 - \eta$, i.e. the one defined by $w$ in the statement of the lemma. Let also
\[ \Upsilon_t(x, y) := \int_x^y \sin^{t-2} r \, dr. \]

The following equality is standard (see 3.621 in [GR65])
\[ \Upsilon_t(0, \pi) = \begin{cases} \pi(t-3)(t-5)\cdots 1 & t \text{ is even}, \\ \frac{2(t-3)(t-5)\cdots 2}{(t-2)(t-4)\cdots 3} & t \text{ is odd}. \end{cases} \]

We will need a lower bound on $\Upsilon_t(0, \pi)$. First assume that $t$ is even.
\[ \Upsilon_t(0, \pi) = \frac{\pi(t-3)(t-5)\cdots 1}{(t-2)(t-4)\cdots 2} = \pi(1 - 1/(t-2))(1 - 1/(t-4))\cdots (1 - 1/2) \]
using $1 - x \geq e^{-x-x^2}$ for $0 \leq x \leq 1/2$, $\sum_{i=1}^{n} 1/i \leq \ln(n) + \gamma + 1/2n$ [You91] and $\sum_{i=1}^{n} 1/i^2 < \pi^2/6$,
\[ \geq \pi \exp \left( -\frac{1}{2} \sum_{i=0}^{t/2-1} \frac{1}{i} - \frac{1}{2} \sum_{i=0}^{t/2-1} \frac{1}{i^2} \right) \geq \pi \exp \left( -\frac{\ln(t/2 - 1)}{2} - 1 - \frac{\pi^2}{12} \right) > \frac{1}{2\sqrt{t/2 - 1}}. \]

Doing the same calculation for odd $t$ it follows that for all $t \geq 8$,
\[ \Upsilon_t(0, \pi) \geq \frac{1}{5\sqrt{t}}. \] (6.1)

We will need the following fact
\[ \mu(S^{t-1}_\eta) = 2\pi \ U_{t-2}(0, \arccos \eta) \prod_{i=1}^{t-3} \Upsilon_i(0, \pi). \]
Note that in this language, the surface area of the hypersphere is \( 2 \mu(S_t^{t-1}) \). We therefore have
\[
\Pr_{w \in S_t^{t-1}}[w \cdot z \leq \eta] = \frac{2\mu(S_0^{t-1}) - \mu(S_t^{t-1})}{2\mu(S_0^{t-1})} = \frac{1}{2} + \frac{\mu(S_0^{t-1}) - \mu(S_t^{t-1})}{2\mu(S_0^{t-1})}
\]
\[
= \frac{1}{2} + \frac{\Upsilon_{t-2}(0, \pi/2) - \Upsilon_{t-2}(0, \arccos \eta)}{2\Upsilon_{t-2}(0, \pi/2)}
\]
\[
= \frac{1}{2} + \frac{\Upsilon_{t-2}(\arccos \eta, \pi/2)}{\Upsilon_{t-2}(0, \pi)} \leq \frac{1}{2} + 5\sqrt{t} \Upsilon_{t-2}(\arccos \eta, \pi/2) \quad \text{by \eqref{eq:4}}
\]
\[
\leq \frac{1}{2} + 5\sqrt{t}(\pi/2 - \arccos \eta).
\]

Now we need the Taylor expansion of \( \arccos \eta \), according to which
\[
\arccos \eta = \frac{\pi}{2} - \sum_{k=0}^{\infty} \frac{(2k)!}{2^{2k}(k!)^2(2k+1)} \eta^{2k+1} \leq \frac{\pi}{2} - \eta \sum_{k=0}^{\infty} \frac{(2k)!}{2^{2k}(k!)^2(2k+1)} = \frac{\pi}{2} - \eta \frac{\pi}{2}
\]

Therefore,
\[
\Pr_{w \in S_t^{t-1}}[w \cdot z \leq \eta] \leq \frac{1}{2} + \eta \frac{\pi}{2} 5\sqrt{t} < \frac{1}{2} + 8\eta \sqrt{t}. \]

The following theorems follow from Lemma 6.4.

**Theorem 6.5.** Let \( G \) be a finite subgraph of \( B_\delta^m \) on \( n \) vertices. Then the level-\( \left( \frac{c^2}{649} - 1 \right) \) Sherali-Adams LP relaxation of Vertex Cover has objective value at most \( (1/2 + \epsilon)n \) for \( G \).

**Proof.** We use Experiment Local-Global with parameter \( t = \frac{c^2}{649} \), to define a family of consistent local distributions. From Lemma 6.3 these are locally consistent distributions of vertex covers. Then Fact 2.1 gives us a level-\( t - 1 \) Sherali-Adams solution. By Lemma 6.4, setting \( \eta = \sqrt{\delta} \), the contribution of each vertex in the objective function is exactly \( 1/2 + \epsilon \).

**Theorem 6.6.** For every \( \epsilon \), there are graphs on \( n \) vertices such that the level-\( \Omega(\frac{\log n}{\log \log n}) \) Sherali-Adams LP relaxation of Vertex Cover has integrality gap \( 2 - \epsilon \).

**Proof.** We start with the \( n \)-vertex Frankl-Rödl graphs \( G_\delta^n \), \( n = 2^m \), which are finite subsets of the Borsuk graph \( B_\delta^m \). We set \( \delta = \Theta(\sqrt{\log m/m}) \) so that the conditions of Theorem 2.3 are satisfied, i.e. all vertex covers of \( G_\delta^n \) have size \( n - o(n) \). Theorem 6.5 then implies that the level-\( \left( \frac{c^2}{649} - 1 \right) \) Sherali-Adams relaxation of the vertex cover polytope has integrality gap at least
\[
\frac{n - o(n)}{(1/2 + \epsilon)n} = 2 - 2\epsilon - o(1). \]

### 6.3 Fooling SDPs derived by the Sherali-Adams System

#### 6.3.1 Preliminary Observations for the Sherali-Adams SDP Solution

Let \( y \) be a solution of the Sherali-Adams LP relaxation defined from the local distributions \( \{D(I)\} \), namely \( y_I = \Pr_{S \sim D(I)}[I \subseteq S] \). Then \( y \) uniquely determines the matrix \( M_1 = M_1(y) \) in
(2.13). In order to establish a Sherali-Adams SDP integrality gap, we need to show that $M_1(y)$ is positive-semidefinite for an appropriately chosen $y$.

It is convenient to denote by $M_1'(y)$ the principal submatrix $M_1(y)$ indexed by nonempty sets. Note that for the solution we introduced in the previous section, all $y_{ij}$ attain the same value, say $y_R$. In other words, $M_1(y) = \left( \begin{array}{cc} 1 & y_R \\ y_R^T & M_1'(y) \end{array} \right)$, where 1 denotes the all 1 vector of appropriate size.

**Lemma 6.7.** Suppose that $1$ is an eigenvector of $M_1'(y)$. Then $M_1(y) \succeq 0$ iff $M_1'(y) \succeq 0$ and for some $j \in V$, $\text{avg}_{i \in V} y_{ij} \geq y_R^2$.

**Proof.** We will first show that $M_1 = M_1(y) \succeq 0$ if and only if $M_1'(y) - y_R^2 \cdot 1 \succeq 0$, where $J$ is the all 1 matrix. We remind the reader that $M_1 \succeq 0$ if and only if

$$0 \leq \min_{v} v \cdot M_1 v = \min_{w} \min_{\alpha} [\alpha | w] \cdot M_1 [\alpha | w] = \min_{w} \left( \alpha^2 + 2 \alpha y_R w \cdot 1 + w \cdot M_1' w \right). \quad (6.2)$$

We can write any vector $v$ as $v = [\alpha | w]$ where $\alpha$ is the first coordinate of $v$ and $w$ is the rest. We will fix $w$ and find the $\alpha$ that minimizes $[\alpha | w] \cdot M_1 [\alpha | w]$ by taking the derivative.

$$0 = \frac{d}{d\alpha} ([\alpha | w] \cdot M_1 [\alpha | w]) = 2\alpha + 2 y_R w \cdot 1 \Rightarrow \alpha = -y_R w \cdot 1.$$ 

So the minimum in (6.2) can be rewritten as,

$$\min_{v} v \cdot M_1 v = \min_{w} (w \cdot M_1' w - y_R^2 (w \cdot 1)^2) = \min_{w} (w \cdot M_1' w - y_R^2 w \cdot Jw)$$

$$= \min_{w} w \cdot (M_1 - y_R^2 J)w.$$ 

So $M_1 \succeq 0$ if and only if $M_1' - y_R^2 \succeq 0$.

Now observe that the 1 is an eigenvector of both $M_1'(y)$ and $J$ and all the other eigen values of $J$ are zero. So the eigen vectors of $M_1(y)$ are also eigen vectors of $M_1'(y)$ and their corresponding eigen values are all the same except for the eigen value corresponding to 1. The eigen value of $M_1'(y)$ corresponding to 1 is the eigen value of $M_1(y)$ corresponding to the same vector minus $ny_R^2$. So $M_1'(y) \succeq 0$ if and only if $M_1(y) \succeq 0$ and the eigen value of 1 $\cdot M_1(y)1 \geq n y_R^2$. The second condition is equivalent to the second condition of the Lemma because the value $\text{avg}_{i \in V} y_{ij}$ does not depend on $j$. \hfill \Box

The next Lemma establishes a sufficient condition for solutions fooling SDP relaxations for Borsuk graphs. The proof uses the standard tool of tensoring. Recall that for $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^m$ their **tensor product** $u \otimes v \in \mathbb{R}^{nm}$ is a vector indexed by ordered pairs from $[n] \times [m]$ taking value $u_i v_j$ at coordinate $(i, j)$. For any polynomial $P(x) = c_0 + c_1 x + \ldots + c_d x^d$ with nonnegative coefficients consider the function $\otimes P(\cdot)$ mapping a vector $u \in \mathbb{R}^n$ to the vector $\otimes P(u) = (\sqrt{c_0} u_0^{\otimes 1}, \ldots, \sqrt{c_d} u_0^{\otimes d}) \in \mathbb{R}^{\sum_{i=0}^d n^i}$, where $u_i^{\otimes i}$ is the vector obtained by tensoring $u$ with itself $i$ times. Polynomial tensoring can be used to manipulate inner products in the
sense that $\otimes_P(u) \cdot \otimes_P(v) = P(u \cdot v)$; it was used as an ingredient in many integrality gap results such as [GK98, Cha02, GMPT10, GMT08].

**Lemma 6.8.** Let $y$ be a solution to the level-$t$ Sherali-Adams LP relaxation for Vertex Cover for a (subgraph of) Borsuk graph with vector representation $u_1, \ldots, u_n \in S^{m-1}$ and suppose that the value $y_{\{i,j\}}$ can be expressed as $f(u_i \cdot u_j)$. If the Taylor expansion of $f(x)$ has no negative coefficients, then $M'_1(y) \succeq 0$.

**Proof.** Consider the Taylor expansion of $f(x) = \sum_{i=0}^{\infty} a_i x^i$, where $a_i \geq 0$. We map $u_i \in S^{m-1}$ to an infinite dimensional space as follows $u_i \mapsto \otimes f(u_i)$. Then the vectors $\otimes f(u_i)$ constitute the Cholesky decomposition of $M'_1(y)$, and therefore $M'_1(y) \succeq 0$. \hfill \blacksquare

We now examine the Sherali-Adams solution of some special case that will be instructive for our general argument. Consider some $n$ vertex subgraph $G = (V,E)$ of $B^m_{\rho^2}$ with vector representation $z_i \in S^{m-1}$. Suppose also that edges $ij \in E$ appear exactly when $z_i \cdot z_j = -1 + 2\rho^2$, and that for all other pairs $i, j \in V$ we have $z_i \cdot z_j \geq -1 + 2\rho^2$. Run the Local-Global experiment in Algorithm 5 with parameter $t = 2$ to define the level-2 Sherali-Adams solution $y$

$$y_I = \Pr_{w \in S^1} [w \cdot z_i \leq \rho, \forall i \in I] \quad (6.3)$$

for all $I$ of size at most 2, where $w$ is distributed uniformly on the circle.

**Claim 6.9.** For all $i, j$, $y_{\{i,j\}} = f(z_i \cdot z_j, \rho)$ where,

$$f(x, \rho) = \begin{cases} 1 - \frac{2\rho}{\pi} & \text{if } x \leq 2\rho^2 - 1, \\ 1 - \frac{\rho}{\pi} - \frac{\theta_x}{2\pi} & \text{otherwise}, \end{cases}$$

where $\theta_x = \arccos(x)$. In particular, when $z_i \cdot z_j = 1$, $y_{\{i,j\}} = 1 - \frac{\theta}{\pi}$.

Notice that the function $f(x, \rho)$ above is continuous and $f(x, 1) = 1$.

**Proof.** We have,

$$y_{i,j} = \Pr_{w \in S^1} [w \cdot z_i \leq \rho \& w \cdot z_j \leq \rho, \text{ with } z_i \cdot z_j = x].$$

Note that the above is the probability that two points at angle $\theta_x = \arccos(z_i \cdot z_j)$ remain on the unit circle after removing a random cap of measure $\arccos(\rho)/\pi$. It is not hard to see that the first point is in this removed cap with probability $\theta/\pi$. Furthermore, the probability that the first point is outside the cap while the second point is inside depends on $\theta_x$ versus $2\theta$; it is $\frac{\theta}{2\pi}$ if $\theta_x \leq 2\theta$ (i.e. $x \geq 2\rho^2 - 1$) and $\frac{\theta}{\pi}$ otherwise. \hfill \blacksquare

The next fact is motivated by the condition of Lemma 6.8.
Fact 6.10. If $\rho \in [0,1]$, the Taylor expansion the function $g(x) = 1 - \frac{\theta_\rho}{\pi} - \frac{\theta_\rho}{2\pi}$ around $x = 0$ has no negative coefficient.

Proof. Notice that in the Taylor expansion of $\arccos(x)$ all the terms except the constant term are non-positive so it is enough to show that the $g(0) \geq 0$. Since $\rho > 0$, $\theta_\rho < \pi/2$ so $g(0) = 1 - \frac{\theta_\rho}{\pi} - 1/4 > 1/4$.

Note that if we start with a configuration of vectors $z_i$ for which $z_i \cdot z_j \geq -1 + 2\rho^2$ for all pairs $i, j \in V$, then the value $y_{(i,j)}$ will be described as a function on the inner product $z_i \cdot z_j = x$, and this function on $x$ will have Taylor expansion with nonnegative coefficients. Unfortunately, for our Sherali-Adams solution of the previous sections this is not the case. We establish this extra condition in Section 6.3.2, making sure that $M_1(y)$ is positive semidefinite. Proving that the matrix $M_1(y)$ is positive semidefinite will require one extra simple argument, which is self evident from Lemma 6.7.

6.3.2 An Easy level-2 Sherali-Adams SDP Solution

In this section we apply the techniques developed in Section 6.3.1 to show a tight integrality gap for VERTEX COVER in the level-2 Sherali-Adams SDP system. This serves as an instructive example for higher levels whose proof are a smooth generalization of the arguments below. We will show,

Theorem 6.11. For any $\epsilon > 0$, there exist $\delta > 0$ and sufficiently big $m$, such that the level-2 Sherali-Adams SDP system for VERTEX COVER on $G_m^m$ has objective value at most $2^m(1/2 + \epsilon)$.

As the theorem states, we start with the Frankl-Rödl graph $G_m^m = (V,E)$, which is a subset of $B_m^m$, with vector representation $u_i$. Our goal is to define $y$ in the context of Theorem 6.5, so as the matrix $M_1(y)$ to be positive semidefinite. Our Sherali-Adams solution as it appears in Theorem 6.5 does not satisfy the constraint $M_1(y) \succeq 0$, for reasons that will be clear shortly. For this, we need to apply the transformation $u_i \mapsto z_i := (\sqrt{\zeta},\sqrt{1-\zeta} \otimes_P(u_i))$, for some appropriate tensoring polynomial $P(x)$, and some $\zeta > 0$ (that is allowed to be a function of $(m,\delta)$). We will use the following lemma first proved by Charikar [Cha02].

Lemma 6.12 ([Cha02]). There exist a polynomial $P(x)$, with nonnegative coefficients and $P(1) = 1$, such that for all $x \in [-1,1]$, we have $P(x) \geq P(-1 + 2\delta) = -1 + 2\delta_0$, for some $\delta_0 = \Theta(\delta)$. Moreover, for every constant $c > 0$ and for every $x \in (-c/\sqrt{m},c/\sqrt{m})$, we have $|P(x)| = O(\sqrt{1/m})$.

We use the polynomial $P$ of Lemma 6.12 to map the vectors $u_i$ to the new vectors $z_i$. Note that with this transformation, for an edge $ij \in E$ we have $z_i \cdot z_j = \zeta + (1-\zeta)P(-1 + 2\delta) = \zeta + (1-\zeta)(-1 + 2\delta_0) = -1 + 2(\zeta(1-\delta_0) + \delta_0)$. If we denote $\sqrt{\zeta(1-\delta_0) + \delta_0}$ by $\rho$, then the above transformation maps $G_m^m$ to (a subgraph of) $G_m^m$, where $m'$ is the degree of the polynomial $P$. We can therefore run Algorithm 5 with parameter $t = 2$ (and $\delta = \rho^2$) on the
vectors \( z_i = (\sqrt{\zeta}, \sqrt{1 - \zeta} \otimes \rho(u_i)) \). Then Lemma 6.3 implies that \( y \) as defined in (6.3) is a level-2 Sherali-Adams solution (the parameters \( \delta, \zeta \) will be fixed later.) Next we show that for a slightly perturbed \( y \) we have that \( M_1(y) \) is positive semidefinite.

Notice that among the vectors \( z_i \)'s the ones with the smallest inner product are the edges. In fact the minimum inner product \( z_i \cdot z_j \) is by definition \(-1 + 2\rho^2\). So one we can use Fact 6.10 and Claim 6.10 to establish the positive definiteness of \( M_1(y) \). In particular, since for every \( i, j \), \( z_i \cdot z_j \geq -1 + 2\rho^2 \) the value of \( y_{[i,j]} \) is exactly \( g(\zeta + (1 - \zeta)P(z_i \cdot z_j)) \), where \( g(x) = 1 - \frac{\theta}{\pi} - \frac{\arccos(x)}{2\pi} \). By Fact 6.10 we know that the function \( g(x) \) has Taylor expansion with nonnegative coefficients. Since \( \zeta + (1 - \zeta)P(x) \) is a polynomial with nonnegative coefficients, it follows that \( g(\zeta + (1 - \zeta)P(x)) \) has Taylor Expansion with nonnegative coefficients. Hence, we can apply Lemma 6.8 to obtain that

**Lemma 6.13.** The matrix \( M'_1(y) \) is positive semidefinite.

In what follows we describe a way to extend the positive semidefiniteness of \( M'_1(y) \) to that of \( M_1(y) \). In fact what we will show is general and holds for any level \( t \) (where \( t \) is the Sherali-Adams level which solution \( y \) was engineered for). Since the entries of \( M'_1(y) \) are a function of the inner product of the corresponding vectors of the hypercube, it follows that the all 1 vector is an eigenvector for \( M'_1(y) \).

By Lemma 6.7 it follows that we need to show that \( \text{avg}_{i \in V} y_{[i,j]} - y^2_{[i]} \geq 0 \). It turns out that this is not the case, but we can establish a weaker condition.

**Lemma 6.14.** There exist \( c > 0 \) (not depending on \( m, \rho \)), such that \( \text{avg}_{i \in V} y_{[i,j]} - y^2_{[i]} \geq -cD\rho \).

We omit the proof of this lemma but the high level argument goes as follows. Whenever two points have positive inner product, the probability that both are in a random cap is at least 1/4. It can be shown that due to the affine transformation, all but exponentially small fraction of the pairs will have positive inner products, hence we get that the average of \( y_{[i,j]} \) is at least 1/4 - \( o(1) \). On the other hand, from Section 6.2 we know that \( y_{[i]} \leq 1/2 + O(D\rho) \).

Boosting: It remains to show how to “boost” the solution to move from the relaxed condition to the exact, and necessary one. The idea is simple. Consider a very wasteful integral solution to Vertex Cover, namely the solution that takes all vertices. Clearly, if we take a convex combination of this solution with the existing one we still get a Sherali-Adams solution. If the weight of the integral solution is some small number \( \xi > 0 \) then the objective value increases by no more than \( \xi/2 \) which can be absorbed for arguments to go through as long as \( \xi = O(\epsilon) \). The strict convexity of the quadratic function, however, lets this simple perturbation improve the bound on averages as required by Lemma 6.7. This observation is made precise in the following Lemma.

**Lemma 6.15.** Let \( y' \) be the matrix \( y' = (1 - \xi)y + \xi J \) where \( J \) represents the all 1 solution. Also let \( s = y'_{[i]} \) and \( s' = y'_{[i]} \). Then \( \text{avg}_{i,j} y'_{[i,j]} - s'^2 = \Omega(\xi) \).

---

\(^6\)This is because the matrix \( M'_1(y) \) is transitive.
Proof.

\[
\text{avg}_{i,j} y'_{(i,j)} - s^2 = (1 - \xi) \text{avg}_{i,j} y_{(i,j)} + \xi - (1 - \xi)s + \xi^2
\]
\[
= (1 - \xi)\text{avg}_{i,j} A_{(i,j)} + \xi - (1 - \xi)^2 s^2 - 2(1 - \xi)\xi s - \xi^2
\]
\[
\geq (1 - \xi)\text{avg}_{i,j} A_{(i,j)} - s^2 + \xi + \xi s - \xi^2 s^2 - 2\xi s - \xi^2
\]
\[
\geq -O(D\rho + \xi^2) + (1 - s)^2 \xi = \Omega(\xi)
\]

In the last inequality we use the fact that we only take \(D\) so as to make \(D\rho = O(\epsilon)\), that \(s\) is bounded away from 1 and by taking \(\xi = O(\epsilon)\).

We are now ready to formally prove Theorem 6.11.

Proof of Theorem 6.11. We start with the \(n\)-vertex Frankl-Rödl graph \(G^m_\delta\), with \(\delta = \Theta\left(\frac{\log n}{\log \log n}\right)\) so as to satisfy the conditions of Theorem 2.3. We use the polynomial of Lemma 6.12 to obtain the vectors \(z_i = (\sqrt{\zeta}, \sqrt{1 - \zeta} \otimes p(u_i))\), with \(\zeta = \delta_0\) (where \(\delta_0 = \Theta(\delta)\) by Lemma 6.12). We set \(\rho = \sqrt{\zeta(1 - \delta_0) + \delta_0} = \sqrt{\Theta(\zeta)}\), and we run Algorithm 5 on the vectors \(z_i\) with parameter \(t = 2\) (and \(\delta = \rho^2\)), to obtain the vector \(y\). By Lemma 6.3, we have that \(y\) as defined in (6.3) is a level-2 Sherali-Adams solution. Note that since \(\delta = o(1)\) we conclude from Lemma 6.4 that \(y_{(i)} = 1/2 + \Theta(\delta)\).

Next we define \(y'\) as \((1 - \xi)y + \xi J\). We already argued that \(M'(y')\) is positive semidefinite. By the above discussion (and Lemma 6.15) we conclude that \(\text{avg}_{i,j} y'_{(i,j)} - y'^2_{(i)} \geq 0\). We can therefore use Lemma 6.7 to conclude that \(M(y') \succeq 0\). The last thing to note is that the contribution of every vertex in the objective value is \(1/2 + O(\delta)\).

6.3.3 Extending to the Level-\((t + 1)\) Sherali-Adams SDP Relaxation

For the level-\((t + 1)\) Sherali-Adams SDP relaxation we start with the \(n\)-vertex Frankl-Rödl graphs \(G^m_\delta\), \(n = 2^m\) with vector representation \(u_i\). The value of \(\delta\) is chosen so as to satisfy Theorem 2.3, namely \(\delta = \Theta(\sqrt{\log m/m})\). As in Section 6.3.2 we apply two transformations to \(u_i\); one using the tensoring polynomial of Lemma 6.12 and one affine transformation. Then we use the resulting vectors \(z_i = (\sqrt{\zeta}, \sqrt{1 - \zeta} \otimes p(u_i))\) to define a level-\((t + 2)\) Sherali-Adams solution that we denote by \(y\). Our construction of \(y\) will have a parameter \(\rho\) to be set later.

Our goal is to meet the conditions of Lemma 6.7. Namely, the first thing to ensure is that \(M'(y)\) is positive semidefinite. In this direction, from Lemma 6.8 it suffices to show that the Taylor expansion of the function that describes the value of \(y_{(i,j)}\) as \(y_{(i,j)} = f(u_i \cdot u_j)\), has Taylor expansion with nonnegative coefficients. Given that this function at 0 will always represent some probability, the problem is equivalent to showing that the first derivative of this function has such a good Taylor expansion. Our transformation on the vectors \(u_i\) can be thought as mapping their inner product \(u\) first to \(x = P(u)\), and second \(x\) to \(\kappa\zeta(x) = \zeta + (1 - \zeta)x\). Under this notation, we can show the following lemma. The proof is a rather technical calculation and is left for Section 6.3.4.
Lemma 6.16. The derivative of the functional description of \(y_{i,j}\) is

\[
D_\zeta(x) := -(\arccos(\kappa_\zeta(x)))'(1 - \frac{2\rho^2}{1 + \kappa_\zeta(x)})^{1/2}.
\]

(6.4)

Therefore, to conclude that \(M'_1(y) \geq 0\) it suffices to show the next technical lemma. The proof is an arguments along the lines of the proof of Claim 6.13 and will be left for Section 6.3.5.

**Lemma 6.17.** Set \(t = 4\) and \(\rho^2 \in [\zeta, \zeta + \zeta^3]\). Then for sufficiently small \(\zeta\), the function \(D_\zeta(x)\) as it reads in Lemma 6.16 has Taylor expansion with nonnegative coefficients.

Now we are ready to prove Theorem 6.1. First we set \(t = 4\) and obtain a level-\((t + 1)\) Sherali-Adams solution from the vectors \(z_i = (\sqrt{\zeta}, \sqrt{1 - \zeta} \otimes P(u_i))\). We need to set \(\zeta = \sqrt[3]{\delta_0}\), where \(\delta_0 = (1 + \min(P(x)))/2\). Since the rounding parameter we need is \(\rho = \sqrt{\zeta(1 - \delta_0)} + \delta_0\), it is easy to see that \(\rho^2 = \zeta + \zeta^3 - \zeta^4\). It follows by Lemma 6.17 that the matrix \(M'_1(y)\) is positive semidefinite.

Now call \(c\) the constant for which \(\text{avg}_{i \in V} y_{i,j} - y_{i,j}^2 \geq -ct\rho^2\). We also know that if \(t\rho^2\) is no more than a small constant \(\epsilon/10\), then \(y_{i,j} \leq 1/2 + \epsilon\). Then define \(y' = (1 - 4c\epsilon)y + (4c\epsilon)1\).

As we did for the level-2 Sherali-Adams SDP solution, the vector \(y'\) is a level-\((t + 1)\) Sherali-Adams solution. Moreover, the matrix \(M'_1(y')\) is positive semidefinite, and \(\text{avg}_{i \in V} y'_{i,j} - y'_{i,j}^2 \geq 0\). All conditions of Lemma 6.7 are satisfied implying that \(M'_1(y')\) is positive semidefinite. Finally, note that the contribution of the singletons is no more than \(1/2 + \Theta(t\epsilon\rho^2)\). Hence, if we start with \(t\rho^2 = o(1)\), the contribution of the singletons remains \(1/2 + o(1)\). On the other hand, choosing \(\delta = \Theta(\sqrt{\log m/m})\) results in graphs \(G^m_\delta\) with no vertex cover smaller than \(n - o(n)\).

The maximum value of \(t\) in Lemma 6.17 dictates the limitation on the level of our integrality gap. In particular we have the following conjecture and the proof of Theorem 6.2 is straightforward.

**Conjecture 6.18.** Set \(t\) be any even integer and \(\rho^2 \in [\zeta, \zeta + \zeta^3]\). Then for sufficiently small \(\zeta\), the function \(D_\zeta(x)\) as it reads in Lemma 6.16 has Taylor expansion with nonnegative coefficients.

**Theorem 6.19.** Assuming Conjecture 6.18, for every constants \(\epsilon > 0\) and \(t\), the level-\(t\) SDP derived by the Sherali-Adams SDP system for VERTEX COVER has integrality gap \(2 - \epsilon\).

**Remark.** [On the validity of Conjecture 6.18] Evidence for the validity of Conjecture 6.18 is both experimental and theoretical. In particular, some relatively simple arguments can show the following two statements: (a) For every \(N_0 > 0\) there exist small enough \(\zeta > 0\), such that the first \(N_0\) Taylor coefficients of \(D_\zeta(x)\) are positive, (b) For every \(\zeta > 0\), there exist \(N_0 > 0\) such that all but the first \(N_0\) Taylor coefficients of \(D_\zeta(x)\) are positive. While these partial results are not enough to imply Sherali-Adams SDP lower bounds, they do seem to indicate that Conjecture 6.18 is true.
6.3.4 A Functional Description of the Sherali-Adams Solution
(proof of Lemma 6.16)

Our goal is to prove the analog of Theorem 6.11 for a higher level of the Sherali-Adams system, satisfying the positive semidefinite constraint as well. We start with the configuration of Section 6.3.1, namely two vectors $z_i, z_j$ having inner product $\langle x \rangle \geq -1 + 2\rho^2$. We now revisit (6.3) taking into consideration that our experiment takes place in $S^{t+1}$, $t \geq 0$ (this will give rise to a level-$(t + 2)$ Sherali-Adams solution). As in the construction of the level-2 Sherali-Adams SDP solution, we give a functional description of the value of the doubletons $y_{\{i,j\}}$. Then we only need to show that the Taylor expansion of this function has nonnegative coefficients.

Remember that the experiment can be realized by considering $t + 2$ independent random variables $W_i$ chosen from the normal distribution. Note that the random variables $\sum_{i=3}^t W_i^2 := U_1$ and $W_1^2 + W_2^2 := U_2$ follow the distributions $\chi^2(t)$ and $\chi^2(2)$ respectively. We will need the following fact.

Fact 6.20. Let $U_1, U_2$ be two stochastic independent random variables of the distributions $\chi^2(t)$ and $\chi^2(2)$ respectively. Then the random variable $\frac{U_1/t}{U_2/2}$ has probability density function

$$g_{t,2}(u) = \frac{t}{u} \sqrt{\frac{(tu)^t}{(tu + 2)^{t+2}}}.$$ 

Moreover, the associated distribution is known as the $F_{t,2}$-distribution.

Conditioning on $U_1 = u_1$, $U_2 = u_2$, as in Section 6.3.1, we can normalize the events by $u_2$, in which case we can interpret the probability of the event associated with $y_{\{i,j\}}$ again in terms of an experiment that takes place in $S^1$. The difference is that this time, the rounding parameter becomes $\rho / \sqrt{1 - tu/2}$. We just showed the following fact.

Fact 6.21. Let $y_{\{i_1, i_2\}} = \Pr_{w \in S^{t+1}}[w \cdot z_i \leq \rho, \ i = 1, 2]$, namely the values $y_{\{i_1, i_2\}}$ are the result of a random cut that takes place in $S^{t+1}$. The same cut is a distribution of cuts in $S^1$ with rounding parameter $\eta = \rho / \sqrt{1 - tu/2}$. The probability density function of $u$ is the $F_{t,2}$-distribution of Fact 6.20.

For two vectors with $z_i \cdot z_j = x$, we can adjust the notation of (6.3) so as to define $F_t(x, \rho) = y_{\{i,j\}}$. Applying Claim 6.9 we have

$$F_t(x, \rho) = \int_0^\infty g_{t,2}(u) f(x, \rho \sqrt{1 + tu/2}) du.$$  

(6.5)

Now define

$$C := 2 - \rho^2 / t \rho^2, \quad h(x) := \frac{1 - 2\rho^2 + x}{t \rho^2}$$
emphasizing that \( C \) does not depend on \( x \). Applying Claim 6.9 we have
\[
f(x, \rho \sqrt{1 + tu/2}) = \begin{cases} 
1, & u \geq C \\
1 - \frac{2\theta(\rho \sqrt{1 + tu/2})}{\pi}, & h(x) \leq u \leq C, \\
1 - \frac{\theta(x \sqrt{1 + tu/2})}{\pi} - \frac{\theta}{2\pi}, & 0 \leq u \leq h(x)
\end{cases}
\] (6.6)

Note that this is well defined, since \( \rho > 0 \) and for all \( x \in [-1 + 2\rho^2, 1] \) we have \( 0 \leq h(x) \leq C \) (recall that \( x \) represents the inner product of unit vectors, that have always angle at most \( 2\arccos(\rho) \)).

This allows us to find a nicer expression for (6.5). For notational simplicity, let \( \eta = \rho \sqrt{1 + tu/2} \). Then,
\[
F_t(x, \rho) = \int_{u=C}^{\infty} g_{t,2}(u) du + \int_{u=0}^{h(x)} g_{t,2}(u) \left( 1 - \frac{\theta_u}{\pi} - \frac{\theta_x}{2\pi} \right) du + \int_{u=h(x)}^{C} g_{t,2}(u) \left( 1 - \frac{2\theta_u}{\pi} \right) du
\]
\[
= 1 - \frac{1}{\pi} \int_{u=0}^{C} g_{t,2}(u)\theta_u du + \frac{1}{2\pi} \left( -\theta_x \int_{u=0}^{h(x)} g_{t,2}(u) du - 2 \int_{u=h(x)}^{C} g_{t,2}(u)\theta_u du \right)
\]
where the component of \( F_t(x, \rho) \) that depends on \( x \) now becomes clear.

Next we find the first partial derivative of \( F_t(x, \rho) \), with respect to \( x \). For this define
\[
H_1(x) := \theta_x \int_{u=0}^{h(x)} g_{t,2}(u) du, \quad H_2(x) := \int_{u=h(x)}^{C} g_{t,2}(u)\theta_u du,
\]
Now observe that
\[
\frac{\partial}{\partial x} H_1(x) = (\arccos(x))' \int_{u=0}^{h(x)} g_{t,2}(u) du + \arccos(x) \left( \frac{\partial h(x)}{\partial x} \right) \frac{\partial h(x)}{\partial x}
\]
and that
\[
\frac{\partial}{\partial x} H_2(x) = -g_{t,2}(h(x)) \arccos(\frac{\rho \sqrt{1 + t h(x)/2}}{2}) \frac{\partial h(x)}{\partial x}
\]
\[
= -g_{t,2}(h(x)) \arccos(\frac{1 + x}{2}) \frac{\partial h(x)}{\partial x} = -\frac{1}{2} g_{t,2}(h(x)) \arccos(x) \frac{\partial h(x)}{\partial x}
\]
where the last equality follows from the identity \( \arccos(\sqrt{\frac{1 + x}{2}}) = \arccos(x)/2 \). Combining the above, we conclude that
\[
\frac{\partial}{\partial x} F_t(x, \rho) = -(\arccos(x))' \int_{u=0}^{h(x)} g_{t,2}(u) du = -(\arccos(x))' \left[ \left( \frac{tu}{tu + 2} \right)^{t/2} \right]_{0}^{h(x)}
\]
\[
= -(\arccos(x))' \left( \frac{t h(x)}{t h(x) + 2} \right)^{t/2} = -(\arccos(x))' \left( \frac{1 - 2\rho^2 + x^{t/2}}{1 + x}^{t/2} \right).
\] (6.7)
The proof of Lemma 6.16 is now complete.

6.3.5 Positive coefficients of the Taylor Expansion
(Proving Lemma 6.17 in Three Steps)

The goal of this subsection is to prove Lemma 6.17. The rest of the subsection is organized as follows. Subsection 6.3.5.1 deals with some preliminary technicalities needed in the proof. Subsection 6.3.5.2 contains the proofs of some claims that appear in the first part; these are mostly exhaustive calculations. Finally Subsection 6.3.5.3 concludes with the proof of Lemma 6.17.

6.3.5.1 Preliminary Technicalities

We study the Taylor expansion of the function (6.4), namely

\[-(\arccos(\kappa(x)))'(\frac{1 - 2\rho^2 + \kappa(x)}{1 + \kappa(x)})^{t/2},\]

where \(\kappa(x) = \zeta + (1 - \zeta)x\). First note that \((\arccos(\kappa(x)))' = (1 - \zeta)/\sqrt{1 - \kappa^2(x)}\), that is we need to study the Taylor expansion of the function

\[H_{\zeta,\rho}^{(t)}(x) := \frac{1}{\sqrt{1 - \kappa^2(x)}} \left(\frac{1 - 2\rho^2 + \kappa(x)}{1 + \kappa(x)}\right)^{t/2}.\]

In what follows we set \(T = t/2\). First we note that it suffices to show that

\[\frac{\partial}{\partial x} \cdot \frac{H_{\zeta,\rho}^{(t)}(x)}{H_{\zeta,\rho}^{(0)}(x)}\]  

has Taylor expansion with positive coefficients, as the function \(H_{\zeta,\rho}^{(0)}(x)\) clearly exhibits the same property.

The reader can first verify that (6.8) equals

\[\frac{(1 - 2\rho^2 + \zeta + (1 - \zeta)x)^{T-1} p_{\zeta,\rho}(x)}{(1 + \zeta + (1 - \zeta)x)^{T+1} (1 - x)}\]

where

\[p_{\zeta,\rho}(x) := \zeta + 2T\rho^2(1 - \zeta) - \zeta(2\rho^2 - \zeta) + (1 - \zeta)(2\zeta - 2T\rho^2 - 2\rho^2 + 1)x + (1 - \zeta)^2x^2.\]

In order to find the Taylor expansion of (6.9) we proceed in two steps. First, we find a closed formula for the coefficients of (6.9) in the absence of the numerator, which is a bounded degree polynomial. Second, we would like to convolute the series (for which we know explicitly the coefficients) with the polynomial \((1 - 2\rho^2 + \zeta + (1 - \zeta)x)^{T-1} p_{\zeta,\rho}(x)\). The presence of many
different parameters complicates the analysis. We observe however that since parameter $\rho^2$ only appears in the numerator of (6.9), and since $\rho^2 \leq \zeta + \zeta^3$ and $T$ is a constant, we may assume that $\rho^2 = \zeta$. This is because, if we show that all coefficients of this “perturbed” rational function are at least $\Omega(\zeta)$, and that in the absence of the bounded degree polynomial of the numerator, the Taylor coefficients of the function are bounded in absolute value by $O(1/\zeta^2)$, then this will imply positivity for the coefficients of the original function $H_{\zeta, \rho}^{(t)}(x)$ as well. We will make this argument precise in Subsection 6.3.5.3, where we actually prove Lemma 6.17.

To summarize, we set $\rho^2 = \zeta$ in (6.9). After we factor out the multiplicative constant $(1 - \zeta)/(1 + \zeta)$, the function we need to study becomes

$$
\frac{(1 + x)^{T-1}}{(1 + (1 - x)x)^{T+1}} \frac{\zeta(2T + 1) - 2T\zeta + 1)x + (1 - \zeta)x^2}{(1 - x)} \tag{6.10}
$$

where $x = 2\zeta/(1 + \zeta)$.

Lemma 6.22 below is the main technical argument for showing non-negativity of the Taylor coefficients of (6.10). This will allow us to conclude with the formal proof of Lemma 6.17 which appears in Subsection 6.3.5.3. For the sake of exposition, we extract some exhaustive calculations and we turn them into claims which we prove in Subsection 6.3.5.2.

**Lemma 6.22.** Fix a sufficiently small $\zeta > 0$, and set $T = 2$. Then for all $N \in \mathbb{N}$, the degree-$N$ Taylor coefficient of the rational function (6.10) is at least $25\zeta - o(\zeta)$. Here the $o(\zeta)$ notation is used for a function independent of $N$ that goes to 0 faster than linear as $\zeta$ goes to zero.

**Proof.** First we note that

$$
\frac{1}{1 - x} \cdot \frac{1}{(1 + (1 - x)x)^3} = \left(\sum_{j=0}^\infty x^j\right) \cdot \left(\sum_{j=0}^\infty (-1)^j \left(\frac{j + 2}{2}\right)(1 - x)^j x^j\right)
$$

$$
= \sum_{N \geq 0} \left(\sum_{j=0}^N (-1)^j \left(\frac{j + 2}{2}\right)(1 - x)^j\right) x^N = \sum_{N \geq 0} L_N(x) x^N,
$$

where

$$
L_N(x) := \frac{1}{(2 - x)^3} \left[1 + \frac{(-1)^N(1 - x)^{N+1}}{2} \frac{(2 - x)^2(N + 1)^2 + (2 - x)(4 - x)(N + 1) + 2}{(N + 1) + 2}\right].
$$

Next, we need to convolute with the bounded polynomial

$$(1 + x)(5\zeta + (1 - 4\zeta)x + (1 - \zeta)x^2) = 5\zeta + (1 + \zeta)x + (2 - 5\zeta)x^2 + (1 - \zeta)x^3.
$$

**Claim 6.23.** The first four terms of the Taylor expansion of (6.10) read as

$$
5\zeta + \left(1 - \Theta(\zeta)\right)x + \left(19\zeta - \Theta(\zeta^2)\right)x^2 + \left(1 - \Theta(\zeta)\right)x^3 + \cdots
$$
Claim 6.23, which is shown in Subsection 6.3.5.2, says that the first 4 terms of the expansion of (6.10) satisfy what Lemma 6.22 claims to be true. Hence, it remains to focus on higher degree coefficients.

The degree $N + 1$ coefficient of (6.10), for $N \geq 2$, is

$$(1 - \zeta)L_N(N - 2) + (2 - 5\zeta)L_N(N - 1) + (1 + \zeta)L_N(N) + 5\zeta L_N(N + 1). \quad (6.12)$$

Thus, we now need to show that expression (6.12) is at least $\Omega(\zeta)$.

In Subsection 6.3.5.2, the reader can verify that

**Claim 6.24.** Expanding (6.12), and factoring $\frac{1}{(2-\zeta)^3(1+\zeta)^4}$ out, results into

$$(1 + \zeta)^4 + (-1)^N \left(\frac{1 - \zeta}{1 + \zeta}\right)^N (10\zeta^2 N^2 - 2(6 - 18\zeta + \zeta^2)\zeta N + 1 - 20\zeta + 36\zeta^2 - 4\zeta^3 - \zeta^4). \quad (6.13)$$

Next we study the function

$$g_\zeta(N) := \left(\frac{1 - \zeta}{1 + \zeta}\right)^N (10\zeta^2 N^2 - 2(6 - 18\zeta + \zeta^2)\zeta N + 1 - 20\zeta + 36\zeta^2 - 4\zeta^3 - \zeta^4)$$

and we show that $|g_\zeta(N)| < 1 - 48\zeta + o(\zeta)$; this suffices for our purposes. To that end, we compute the extreme points of $g_\zeta(N)$. Note that checking the condition for these extreme points as well the boundary cases $N = 2$ and $N \to \infty$ is enough. Luckily, the boundary cases are easy to deal with. Note that $g_\zeta(2) = 1 - 48\zeta + o(\zeta)$, while $g_\zeta(N)$ clearly goes to 0 as $N$ tends to infinity. Thus, it remains to study the extreme points of $g_\zeta(N)$.

**Claim 6.25.** $g_\zeta(N)$ attains two local extreme points at $N_{1,2}$ for which

$$\lim_{\zeta \to 0} g_\zeta(N_1) \simeq 0.32, \quad \lim_{\zeta \to 0} g_\zeta(N_2) \simeq -0.9898.$$

Claim 6.25, whose proof can be found in Subsection 6.3.5.2, implies that (6.13) is at least $52\zeta - o(\zeta)$. Recall from Claim 6.24 that we have scaled all coefficients by $\frac{1}{4}(2-s)^3(1+\zeta)^4$, and thus $25\zeta - o(\zeta)$ is a lower bound for all Taylor coefficients we are looking for.

### 6.3.5.2 Proving Claims that Appear in the Proof of Lemma 6.22

This section is devoted into showing all claims that appear in Lemma 6.22 which mostly involve brute-force calculations.

**Proof of Claim 6.23.** Recalling (6.11), note that

$$L_N(0) = 1 \quad L_N(1) = 3x - 2 \quad L_N(2) = 6x^2 - 9x + 4 \quad L_N(3) = 10x^3 - 24x^2 + 21x - 6$$
This implies that the first 4 coefficients of (6.10) are,

\[(6.10) = 5\zeta + \left(1 + \zeta + (3\zeta - 2)5\zeta\right)x + \left(2 - 5\zeta + (3\zeta - 2)(1 + \zeta) + (6\zeta^2 - 9\zeta + 4)5\zeta\right)x^2 + \]
\[\left(1 - \zeta + (3\zeta - 2)(2 - 5\zeta) + (6\zeta^2 - 9\zeta + 4)(1 + \zeta) + (10\zeta^3 - 24\zeta^2 + 21\zeta - 6)5\zeta\right)x^3 + \cdots \]
\[= 5\zeta + \left(1 - \Theta(\zeta)\right)x + \left(2 - 5\zeta + 3\zeta - 2 + 3\zeta - 2\zeta + 30\zeta^2\zeta - 45\zeta + 20\zeta\right)x^2 + \]
\[\left(1 - \zeta - 4 + 10\zeta + 6\zeta + 4 - 9\zeta + 4\zeta - 30\zeta - \Theta(\zeta^2)\right)x^3 + \cdots \]
\[= 5\zeta + \left(1 - \Theta(\zeta)\right)x + \left(19\zeta - \Theta(\zeta^2)\right)x^2 + \left(1 - \Theta(\zeta)\right)x^3 + \cdots \]

\[\square \]

**Proof of Claim 6.24.** If we expand (6.12) and after we factor \(\frac{1}{(2 - 8\zeta)(1 + \zeta)^2}\) out we obtain

\[
(1 + \zeta)^4 + \frac{(1 + \zeta)^4}{8}(-1)^N\left(1 - \zeta\right)\left(\frac{1 + \zeta}{1 + \zeta}\right)^N\left[\left(1 - \zeta\right)\left(1 + \zeta\right)\left(\frac{4}{(1 + \zeta)^2}(N - 1)^2 + \frac{2(4 + 2\zeta)}{(1 + \zeta)^2}(N - 1) + 2\right)\right. \\
\left. - (2 - 5\zeta)\left(\frac{4}{(1 + \zeta)^2}N^2 + \frac{2(4 + 2\zeta)}{(1 + \zeta)^2}N + 2\right)\right. \\
\left. + (1 + \zeta)\left(\frac{4}{(1 + \zeta)^2}(N + 1)^2 + \frac{2(4 + 2\zeta)}{(1 + \zeta)^2}(N + 1) + 2\right)\right] \\
\left. - 5\zeta\left(\frac{1 - \zeta}{2}\right)^2\left(\frac{4}{(1 + \zeta)^2}(N + 2)^2 + \frac{2(4 + 2\zeta)}{(1 + \zeta)^2}(N + 2) + 2\right)\right] \\
= (1 + \zeta)^4 + \frac{1}{4}(-1)^N\left(1 - \zeta\right)\left(\frac{1 + \zeta}{1 + \zeta}\right)^N\left[\left(1 + \zeta\right)^3(2(N - 1)^2 + (4 + 2\zeta)(N - 1) + (1 + \zeta)^2)\right. \\
\left. - (1 + \zeta)^2(2 - 5\zeta)(2N^2 + (4 + 2\zeta)N + (1 + \zeta)^2)\right. \\
\left. + (1 + \zeta)^2(1 - \zeta)(2(N + 1)^2 + (4 + 2\zeta)(N + 1) + (1 + \zeta)^3)\right. \\
\left. - 5\zeta(1 - \zeta)^2(2(N + 2)^2 + (4 + 2\zeta)(N + 2) + (1 + \zeta)^3)\right] \\
= (1 + \zeta)^4 + \frac{1}{4}(-1)^N\left(1 - \zeta\right)\left(\frac{1 + \zeta}{1 + \zeta}\right)^N\left[2N^2\left((1 + \zeta)^3 - (1 + \zeta)^2(2 - 5\zeta) + (1 + \zeta)^2(1 - \zeta) - 5\zeta(1 - \zeta)^2\right)\right. \\
\left. + N\left((1 + \zeta)^3(-4 + 4 + 2\zeta) - (1 + \zeta)^2(2 - 5\zeta)(4 + 2\zeta)\right.\right. \\
\left. + (1 + \zeta)^2(1 - \zeta)(4 + 4 + 2\zeta) - 5\zeta(1 - \zeta)^2(8 + 4 + 2\zeta)\right.\right. \\
\left. + (1 + \zeta)^3(2 - 4 - 2\zeta + (1 + \zeta)^2) - (1 + \zeta)^2(2 - 5\zeta)(1 + \zeta)^2\right]
\begin{align*}
+ (1 + \zeta)^2(1 - \zeta)(2 + 4 + 2\zeta + (1 + \zeta)^2) \\
- 5\zeta(1 - \zeta)^2(8 + 8 + 4\zeta + (1 + \zeta)^2) \\
= (1 + \zeta)^4 + \frac{1}{4}(-1)^N \left( \frac{1 - \zeta}{1 + \zeta} \right)^N \left[ 40\zeta^2 N^2 + N \left( -8\zeta^3 + 144\zeta^2 - 48\zeta \right) \\
- 4\zeta^4 - 16\zeta^3 + 144\zeta^2 - 80\zeta + 4 \right].
\end{align*}

By rearranging the terms in the brackets we obtain (6.13) as claimed. \hfill \blacksquare

Proof of Claim 6.25. The function \( g_\zeta(N) \) attains its extreme points at the roots of the equation \( \frac{\partial g_\zeta(N)}{\partial N} = 0 \). It is easy to check that,

\[
\frac{\partial g_\zeta(N)}{\partial N} = \left( \frac{1 - \zeta}{1 + \zeta} \right)^N \left( -12\zeta + 36\zeta^2 - 2\zeta^3 + (1 - 20\zeta + 36\zeta^2 - 4\zeta^3 - \zeta^4) \ln \frac{1 - \zeta}{1 + \zeta} \\
+ N \left( 20\zeta^2 + (-12\zeta + 36\zeta^2 - 2\zeta^3) \ln \frac{1 - \zeta}{1 + \zeta} + 10\zeta^2 N^2 \ln \frac{1 - \zeta}{1 + \zeta} \right) \right).
\]

The roots of \( \frac{\partial g_\zeta(N)}{\partial N} = 0 \) are then,

\[
N_{1,2} := \frac{1}{10\zeta} \left( 6 - \frac{10\zeta}{\ln \left( \frac{1 - \zeta}{1 + \zeta} \right)} + \zeta^2 - 18\zeta \pm \sqrt{26 - 16\zeta - 24\zeta^2 + 4\zeta^3 + 11\zeta^4 + \frac{100\zeta^2}{\ln^2 \left( \frac{1 - \zeta}{1 + \zeta} \right)}} \right).
\]

It remains to evaluate \( g_\zeta(N) \) at the roots \( N_{1,2} \). Notice that both roots are \( \Theta(1/\zeta) \) so we have,

\[
\lim_{\zeta \to 0} g_\zeta(N_{1,2}) = \lim_{\zeta \to 0} \left( \frac{1 - \zeta}{1 + \zeta} \right)^{N_{1,2}} \left( 10\zeta^2 N_{1,2}^2 - 2(6 - 18\zeta + \zeta^2)\zeta N_{1,2} + 1 - 20\zeta + 36\zeta^2 - 4\zeta^3 - \zeta^4 \right) \\
= \lim_{\zeta \to 0} \exp(-2\zeta N_{1,2}/(1 + \zeta)) \left( 10(\zeta N_{1,2})^2 - 12\zeta N_{1,2} + 1 \right) \\
= \exp(-2 \lim_{\zeta \to 0} \zeta N_{1,2}) \left( 10 \left( \lim_{\zeta \to 0} \zeta N_{1,2} \right)^2 - 12 \left( \lim_{\zeta \to 0} \zeta N_{1,2} \right) + 1 \right). \tag{6.14}
\]

But we have,

\[
\lim_{\zeta \to 0} \zeta N_{1,2} = \lim_{\zeta \to 0} \frac{1}{10} \left( 6 - \frac{10\zeta}{\ln \left( \frac{1 - \zeta}{1 + \zeta} \right)} + \zeta^2 - 18\zeta \pm \sqrt{26 - 16\zeta - 24\zeta^2 + 4\zeta^3 + 11\zeta^4 + \frac{100\zeta^2}{\ln^2 \left( \frac{1 - \zeta}{1 + \zeta} \right)}} \right) \\
= \frac{1}{10} \left( 6 - \lim_{\zeta \to 0} \frac{10\zeta}{\ln \left( -2\zeta/(1 + \zeta) \right)} \right) \pm \sqrt{26 + \lim_{\zeta \to 0} \frac{100\zeta^2}{\ln^2 \left( -2\zeta/(1 + \zeta) \right)^2}} \\
= \frac{1}{10} (6 + 5 \pm \sqrt{26 + 25}) \\
= (11 \pm \sqrt{51})/10.
\]
Plugging this into (6.14) yields
\[
\lim_{\zeta \to 0} g_{\zeta}(N_1) = \exp(- (11 + \sqrt{51})/5) \left( (11 + \sqrt{51})^2/10 - 12(11 + \sqrt{51})/10 + 1 \right) \simeq 0.32,
\]
\[
\lim_{\zeta \to 0} g_{\zeta}(N_2) = \exp(- (11 - \sqrt{51})/5) \left( (11 - \sqrt{51})^2/10 - 12(11 - \sqrt{51})/10 + 1 \right) \simeq -0.9898. \quad \blacksquare
\]

6.3.5.3 Sum Up - The proof of Lemma 6.17

We are now ready to finish the proof of Lemma 6.17.

Proof of Lemma 6.17. As stated before it is enough to show positivity of all coefficients of the expansion of (6.9) for \( \rho^2 \in [\zeta, \zeta + \zeta^3] \) and \( T = 2 \). Notice that \( \rho \) appears only in the denominator of (6.9) which is a polynomial of degree 3 in \( x \). First consider the function

\[
\frac{1}{(1 + \zeta + (1 - \zeta)x)^3} (1 - x) = \sum_{N \geq 0} L_{\zeta}(N) x^N / (1 + \zeta)^3
\]

\[
= \sum_{N \geq 0} \frac{1}{(2 - x)^3} \left[ 1 + \frac{(-1)^N (1 - x)^{N+1}}{2} \left( (2 - x)^2(N + 1)^2 + (2 - x)(4 - x)(N + 1) + 2 \right) \right] x^N / (1 + \zeta)^3,
\]

where \( x = 2\zeta/(1 + \zeta) \). We argue that for small enough \( \zeta \), \( |L_{\zeta}(N)| \leq 8/\zeta^2 \). Indeed,

\[
|L_{\zeta}(N)| = \frac{1}{(2 - x)^3} \left[ 1 + \frac{(-1)^N (1 - x)^{N+1}}{2} \left( (2 - x)^2(N + 1)^2 + (2 - x)(4 - x)(N + 1) + 2 \right) \right]
< 2 + (1 - x)^{N+1}(2(N + 1)^2 + 4(N + 1)) \leq 2 + 6(1 - x)^{N+1}(N + 1)^2.
\]

The second term has only one extreme point at \( N = 2 \ln(1/(1 - x)) \) for which it takes the value \( 24/(e^2 \ln^2(1 - x)) \). Thus,

\[
|L_{\zeta}(N)| < 2 + \max(6, 4/ \ln^2(1 - x)) \leq 2 + \max(6, 4/ \ln^2(e^{-x})) \leq 2 + \max(6, 4/ x^2) \leq 8/\zeta^2.
\quad (6.15)
\]

In the numerator of (6.9) we have the following polynomial,

\[
(1 - 2\rho^2 + \zeta + (1 - \zeta)x)^{T-1} p_{\zeta, \rho}(x)
\]

\[
= (1 - 2\rho^2 + \zeta + (1 - \zeta)x) ((\zeta + 4\rho^2(1 - \zeta) - \zeta(2\rho^2 - \zeta) + (1 - \zeta)(2\zeta - 6\rho^2 + 1)x + (1 - \zeta)^2x^2)
\]

\[
= h_{\zeta}(x) + 2\rho^2(2 - 3\zeta) - 6\rho^2(1 - \zeta)x - 2\rho^2 p_{\zeta, \rho}(x),
\quad (6.16)
\]

for some function \( h_{\zeta}(x) \). Note that all the coefficients of \( p_{\zeta, \rho}(x) \) are at most 1 in absolute value, so each of the three coefficients of the above polynomial in which \( \rho \) appear are at most \( 8\rho^2 \) in absolute value. One can now combine (6.15) and (6.16) to conclude that for all \( \zeta > 0, \delta > 0 \), and \( n \in \mathbb{N} \), the \( n \)th coefficient of (6.9) changes by at most \( 24\delta/\zeta^2 \) when moving from \( \rho^2 = \zeta \) to
\[ \rho^2 = \zeta + \delta. \] Given that by Lemma 6.22 these coefficients are at least \(25\zeta - o(\zeta)\) when \(\rho^2 = \zeta\), they are all strictly positive when \(\zeta\) is small enough and \(\rho^2 \in [\zeta, \zeta + \zeta^3]\). This completes the proof.

For completeness, we conclude with a short discussion on the convergence of the power series we are considering. Recall that much of our technical work was devoted into analytic properties of the function

\[ f_t(x, \rho) := \Pr_{w \in S^t}[w \cdot z_i \leq \rho \& w \cdot z_j \leq \rho, \text{ with } z_i \cdot z_j = x], \quad (6.17) \]

In particular, we have shown that its derivative \(\partial_x f_t(\kappa(x), \rho)\) agrees with a nonnegative power series \(\sum_{n \geq 0} c_n x^n\) in the interval \((-1 + 2\rho^2, 1)\). It remains to show that \(f_t(\kappa(x), \rho)\) itself agrees with the power series

\[ f_t(0, \rho) + \sum_{n \geq 0} \frac{c_n}{n + 1} x^{n+1} \quad (6.18) \]

on the closed interval \([-1 + 2\rho^2, 1]\). This is the case as long as the sum (6.18) is finite on that interval.

For this, we bound the coefficients \(c_n\) by \(O(1/\sqrt{n})\). This bound follows from the so-called Transfer theorem in singularity analysis.

**Theorem 6.26** (Transfer theorem, special case of Theorem VI.3(i) in [FS09]). Let \(g: D \setminus \{1\} \to \mathbb{C}\) be a function of complex number. Suppose \(g\) is analytic on a disk \(D\) around 0 of radius \(R > 1\), except for \(\{x \in \mathbb{R} \mid x \geq 1\}\), and suppose \(z = 1\) is a singularity of \(g\). Further assume \(g(z) = O(1/\sqrt{z-1})\) near \(z = 1\). Then the \(n\)-th power series coefficient \(c_n\) of \(g\) is at most \(O(1/\sqrt{n})\).

For \(\zeta > 0\), the transfer theorem is applicable to \(g(z) := \partial_z f_t(\kappa(z), \rho) = H_{t, \zeta, \rho}^t(z)\), because the singularities of \(H_{t, \zeta, \rho}^t\) are 1 and \(-(1 + \zeta)/(1 - \zeta) < -1\). Further, we have the bounds

\[ \frac{1}{\sqrt{1 - y^2}} = O\left(\frac{1}{\sqrt{z - 1}}\right), \quad \left(1 - \frac{2\rho^2}{y + 1}\right)^{t/2} = O(1) \]

near \(z = 1\). Hence \(H_{t, \zeta, \rho}^t(z) = O(1/\sqrt{z-1})\), as required.
Chapter 7

An Isoperimetric Inequality for the Hamming Cube with Applications for Integrality Gaps in Degree-bounded Graphs

Let \( n \) be a large positive integer, and \( d \) be an even integer so that \( d/n \) is a constant bounded away from 0 and 1. Theorem 2.3 ([FR87]) says that if a subset of \( \{0,1\}^n \) does not contain any pair of elements in Hamming distance exactly \( d \), then it has exponentially small density.

We study the following “density” variant of the above problem: Consider a “large” subset of \( \{0,1\}^n \). By Theorem 2.3 there exists at least one pair of elements in the set that are in Hamming distance exactly \( d \). But is it true that there are “many” such pairs? We answer this question in the positive using some recent results in discrete harmonic analysis. As an application we show that certain strong linear and semidefinite programming relaxations of the Vertex Cover and Independent Set problems in bounded-degree graphs have large Integrality Gaps.

Frankl and Rödl studied this problem, and obtained results comparing the number of such pairs to the total number of pairs of elements of the subset, but only for the case where \( d/n \approx 1/2 \). In contrast, we study this problem for the following setting of parameters: Let \( F \subseteq \{0,1\}^n \) be a set of constant density \( \alpha > 0 \), i.e. \( \alpha = |F|/2^n \), and let \( d \) be an even integer with \( d/n \) uniformly bounded away from both 0 and 1. We prove strong lowerbounds on the number of pairs \( x, y \in F \) with \( x \) and \( y \) being different in exactly \( d \) coordinates. In fact we can show that among all \( x, y \) which are different in precisely \( d \) coordinates, a constant fraction (depending on \( d/n \) and \( \alpha \)) are in \( F \). Our techniques also yield non-trivial lowerbounds for (slightly) subconstant \( \alpha \) and sublinear \( d \).

\(^1\)Results in this chapter appear in [BHM11]
Our Results

In this chapter we prove two isoperimetric inequalities on the discrete cube.

**Theorem 7.1.** For all $0 < \delta < 1$ and $0 < \alpha \leq 1$, there exists $\epsilon = \epsilon(\delta, \alpha) > 0$ such that for any subset $U \subseteq \{0,1\}^n$ of density at least $\alpha$, we have

$$\Pr_{x,y} [x \in U, y \in U] \geq \epsilon - o(1),$$

where $x$ is chosen uniformly at random from $\{0,1\}^n$ and $y$ is chosen at random so that $d_H(x, y) = 2\lfloor \delta n/2 \rfloor$, i.e. $\delta n$ rounded down to the closest even number.

Furthermore, one can take $\epsilon = 2(\alpha/2)^{1-\frac{2}{|\delta - 1|}}$. If all elements of $U$ have the same parity, then it is possible to take $\epsilon = 2\alpha^{1-\frac{2}{|\delta - 1|}}$.

**Theorem 7.2.** For all $0 < \delta < 1$ and $0 < \alpha_1, \alpha_2 \leq 1$, there exists $\epsilon' = \epsilon'(\delta, \alpha_1, \alpha_2) > 0$ such that for any two subsets $U, W \subseteq \{0,1\}^n$ of densities at least $\alpha_1$ and $\alpha_2$, we have

$$\Pr_{x,y} [x \in U, y \in W] \geq \epsilon' - o(1),$$

where $x$ is chosen uniformly at random from $\{0,1\}^n$ and $y$ is chosen at random so that $d_H(x, y)$ is one of $\lfloor \delta n \rfloor$ and $\lfloor \delta n \rfloor + 1$ with equal probability.

Furthermore, one can take,

$$\epsilon' = \exp \left( \ln \alpha_1 + 2|\rho|\sqrt{\ln \alpha_1 \ln \alpha_2 + \ln \alpha_2} \over (1 - \rho^2) \right)$$

for $\rho = 1 - 2\delta$.

**Remark.** In Theorem 7.1 the fact that $d_H(x, y)$ is always even is required: the set $U = \{x \in \{0,1\}^n : w_H(x) \text{ is even}\}$ does not have any pair of points with odd Hamming distance. Similarly, it is required that the probability in Theorem 7.2 is taken over two cases with different parity of $d_H(x, y)$ as is shown by the following two examples: $U = W = \{x \in \{0,1\}^n : w_H(x) \text{ is even}\}$ and $U = \{x \in \{0,1\}^n : w_H(x) \text{ is even}\}$, $W = \{x \in \{0,1\}^n : w_H(x) \text{ is odd}\}$.

As an application we prove the following two integrality gap lowerbounds for VERTEX COVER and INDEPENDENT SET problems in graphs of bounded degree. The proof combines a simple sampling technique with integrality gaps for these problems in general graphs using Theorem 7.1.

**Theorem 7.3.** For any $\ell$, the integrality gap of the level-$\ell$ Lovász-Schrijver SDP ($LS^\ell$) relaxation for VERTEX COVER in graphs of bounded degree $d$ is at least $2(1 - 12\ell \ln \ln d/d + o_d(\ln \ln d))$.

**Theorem 7.4.** For any fixed $\ell$, the integrality gap of the level-$\ell$ Sherali-Adams LP for INDEPENDENT SET problem on graphs of bounded degree $d$ is at least $\Omega_d(\frac{d}{\log d})$, where the $\Omega_d(\cdot)$ hides constants dependent on $\ell$ but not $d$. 
Previous works

The following theorem of [FR87] is a slightly weaker version of Theorem 2.3 and of particular interest to us.

**Theorem 7.5** (Theorem 1.11 in [FR87]). Let $Q$ be a finite set of size $q$, $F \subseteq Q^n$ and $k$ be an integer. Assume that $\delta n < k < (1 - \delta)n$ for some $0 < \delta < 1/2$ and that if $q = 2$, then $k$ is even. If no $x, y \in F$ are different in precisely $k$ coordinates, then $|F| < (q - \epsilon)n$ for some $\epsilon = \epsilon(\delta, q) > 0$ which only depends on $\delta$ and $q$.

Identifying the elements of $\{0, 1\}^n$ with subsets of $[n]$ in the natural way, we note that the study of subsets of $\{0, 1\}^n$ with no pair of elements at a particular distance is closely related to the study of families of subsets of $\{1, \ldots, n\}$ with no pair of elements of a particular intersection size. In fact if all members of the family are of the same size, the two problems are equivalent. Given this connection it is worthwhile to mention the following theorem of [FR87] from which most of their other theorems follow.

**Theorem 7.6** (Theorem 1.4 in [FR87]). Let $F, G \subset \wp([n])$ be two families and $k$ an integer satisfying $\delta n < k < (1/2 - \delta)n$. If any pair elements $F \in F$ and $G \in G$ satisfy $|F \cap G| \neq k$, then $|F||G| < (4 - \epsilon)n$ for $\epsilon = \epsilon(\delta) > 0$ only depending on $\delta$.

We note that the above theorem settled a conjecture of Paul Erdös [Erd76] (See (3) under Problem 22). In fact the above theorems are just two of a wealth of knowledge about size of set systems (or subsets of $\{0, 1\}^n$) that avoid certain intersection (distance) patterns. The most famous is perhaps a theorem of Erdös, Ko and Rado [EKR61] stating that if a family $F$ of subsets of size $k$ of $[n]$ with $n \geq 2k$ has the property that every two of its members intersect, then its size is at most $\binom{n-1}{k-1}$. We refer the reader to the introduction of [FR87], Chapter 4 in [BF92], [Fra95] and [Sga99] for many beautiful theorems and questions of this nature.

There are two results which lowerbound the number of elements of certain intersections or distance which are particularly relevant us. The first is the following theorem of [FR87].

**Theorem 7.7** (Theorem 1.7 in [FR87]). For every $0 < \sigma < 1/2$, there exist $\epsilon = \epsilon(\sigma) > 0$ and $\delta = \delta(\sigma) > 0$ such that for every family $F \subset \wp([n])$ and any $\ell$ satisfying $|\ell - n/4| < \delta n$ we have,

$$\left| \left\{(F_1, F_2) : F_1, F_2 \in F, |F_1 \cap F_2| = \ell \right\} \right| > (1 - \delta)^n|F|^2.$$  

In other words the ratio of the number of pairs of elements of $F$ with intersection size $\ell$ over the total number of pairs of elements of $F$ is lowerbounded by $(1 - \delta)^n$. This is exponentially small, but we can make the base of the exponent arbitrarily close to 1 at the expense of the theorem being applicable only to $\ell$ which is very close to $n/4$. The following easy corollary of Theorem 7.7 is comparable to our Theorem 7.1.
Chapter 7. An Isoperimetric Inequality for the Hamming Cube

Corollary 7.8. For any $0 < \sigma < 1/2$, there exist $\epsilon = \epsilon(\sigma) > 0$ and $\delta = \delta(\sigma) > 0$ such that for every subset $S \subset \{0,1\}^n$, and any even $d$ satisfying $|d - n/2| < \delta n$, we have

$$\left| \{ (x, y) : x, y \in S, d_H(x, y) = d \} \right| > (1 - \delta)^n |S|^2.$$ 

As mentioned earlier the differences are in the setting of various parameters: (a) Corollary 7.8 is applicable to sets which are exponentially smaller than $2^n$ but Theorem 7.1 is applicable only to sets of constant density; (b) Corollary 7.8 is only applicable to $d$ which is very close to $n/2$ whereas Theorem 7.1 is applicable to any $d$ where $d/n$ is a constant bounded away from 0 and 1; (c) even when $d = n/2$ and $S$ is a constant density set the bound of Corollary 7.8 is (sub)exponentially weaker than the total number of pairs of elements of $\{0,1\}^n$ of distance exactly $d$, whereas Theorem 7.1 shows that a constant fraction of all pairs of distance $d$ are in $S$ in this setting.

The second result comparable to Theorems 7.1 and 7.2 is the following theorem of Mossel et al. [MOR+06].

Theorem 7.9 ([MOR+06]). Let $-1 < \rho < 1$. For any two subsets $U, W \subseteq \{0,1\}^n$ of densities $|U|/2^n = \exp(-u^2/2)$ and $|W|/2^n = \exp(-w^2/2)$, we have

$$\Pr_{x, y}[x \in U, y \in W] \geq \exp\left(-\frac{u^2 + 2|\rho|uw + w^2}{2(1 - \rho^2)}\right),$$

where $x$ is chosen uniformly at random from $\{0,1\}^n$ and $y$ is chosen at random so that $\Pr[x_i = y_i] = (1 + 2\rho)/2$ and $y_i$’s are independent.

Theorem 7.9 is a direct analogue of Theorem 7.2 when we are interested in the number of pairs of elements at expected distance $d$, i.e. when we consider random $x, y \in \{0,1\}^n$ from a product distribution with $x$ being a uniformly random element and $y_i$ chosen independently for each $i$ such that $y_i$ is different from $x_i$ with probability $d/n$. Since the distribution of $x$ and $y$ is a product distribution, tools from discrete Fourier analysis are useful in the proof of Theorem 7.9.

Techniques

It is not hard to see that Theorem 7.1 follows from Theorem 7.2. The bulk of the present chapter is devoted to proving Theorem 7.2. The proof uses tools from discrete Fourier analysis to reduce Theorem 7.2 to Theorem 7.9. The probability in Theorem 7.9 can be rewritten in terms of the Bonami-Beckner [Bon68, Bec75] operator $T_\rho$. To obtain this reduction we define a similar operator which we call $S_d$, and then express the probability in Theorem 7.2 as an expectation involving $S_d$. Both $T_\rho$ and $S_d$ are linear averaging operators and the main idea is to try to show that they are close in ($\ell_2$ to $\ell_2$) operator norm, i.e. for all functions $f : \{0,1\}^n \to \mathbb{R}$, $\|T_\rho f - S_d f\|_2 \leq \epsilon \|f\|_2$. 
However, it turns out that $T_{\rho}$ and $S_d$ are not close in the operator norm. Instead the main technical part of the chapter is to show that $T_{\rho}$ and $\frac{S_d + S_{d+1}}{2}$ are close in this norm. To show this we observe that the two operators have the same eigenfunctions and their eigenvalues are close (See Lemma 7.11 and Theorem 7.17.) While the eigenvalues of $T_{\rho}$ are very simple (they are $\rho^i$ for $0 \leq i \leq n$) the eigenvalues of $S_d$ are (up to normalization) the value of the well-studied “Krawtchouk Polynomials” $\kappa_d^{(n)}(x)$ (see Definition 7.3.) To prove Theorem 7.17 we estimate the value of these polynomials.

Once it is shown that $T_{\rho}$ and $\frac{S_d + S_{d+1}}{2}$ are close in operator norm, it is not hard to show that Theorem 7.9 implies Theorem 7.2. In fact the $o(1)$ term in Theorem 7.2 comes from the difference in operator norm between $T_{\rho}$ and $\frac{S_d + S_{d+1}}{2}$. The restriction that we can only count pairs at distances $d$ or $d+1$ together comes from the fact that we can show $\frac{S_d + S_{d+1}}{2}$ (but not $S_d$) is close to $T_{\rho}$. For a more detailed overview of the proof we refer the reader to the beginning of Section 7.2.

**Organization**

The rest of the chapter is organized as follows. In Section 7.1 we introduce the necessary notation including the definition of the $S_d$ operator and some simple facts about its eigenvalues. The proof of Theorems 7.1 and 7.2 appear in Section 7.2. Finally, we prove Theorems 7.4 and 7.3 in Section 7.3.

**7.1 Preliminaries**

In this section we review some definitions and facts that are only used in this chapter.

In this chapter we identify the cube $\{0, 1\}^n$ with $\mathbb{Z}_2^n$, and thus for $x, y \in \{0, 1\}^n$ their sum $x + y$ is defined.

The Bonami-Beckner [Bon68, Bec75] operator $T_{\rho}$ is a linear operator defined as follows.

**Definition 7.1.** For $-1 \leq \rho \leq 1$ define $T_{\rho}$ as a linear operator acting on the space of real functions on the cube as follows. Let $f : \{0, 1\}^n \to \mathbb{R}$ be a real function on the cube. Then $T_{\rho}f : \{0, 1\}^n \to \mathbb{R}$ is

$$T_{\rho}f : x \mapsto \mathbb{E}[f(y)],$$

where $y_i$'s are chosen independently such that $Pr[y_i = 1] = (1 + \rho)/2$.

It is easy to see that the probability in Theorem 7.9 can be rewritten as

$$Pr_{x,y}[x \in U, y \in W] = 1_U \cdot T_{\rho}1_W,$$

where $1_U$ (similarly $1_W$) is defined as $1_U(x) = 1$ if and only if $x \in U$. This motivates us to define the following averaging operator.
Definition 7.2. For $0 \leq d \leq n$ define $S_d$ as follows. For $f : \{0,1\}^n \to \mathbb{R}$, $S_d f : \{0,1\}^n \to \mathbb{R}$ is

$$S_d f : x \mapsto \mathbb{E}_y [f(y)],$$

where $y$ is chosen at random such that $d_{H}(x,y) = d$.

To prove Theorem 7.2 we show that the $S_d$ is very similar to $T_\rho$ in a precise sense. It is easy to check that

$$T_\rho \chi U = \left( \sum_i (-1)^i \binom{|U|}{i} \left( \frac{1-\rho}{2} \right)^i \left( \frac{1+\rho}{2} \right)^{|U|-i} \right) \chi U = \rho^{|U|} \chi U,$$

i.e. the Fourier characters are the eigenfunctions of $T_\rho$ with eigenvalues $\rho^k$ for $0 \leq k \leq n$. The motivation of the following definition comes from studying the eigenvalues of the $S_d$ operator.

Definition 7.3. for $0 \leq d \leq n$ we use $\kappa_d^{(n)}(x)$ for the Krawtchouk Polynomial of degree $d$ defined as,

$$\kappa_d^{(n)}(x) = (-1)^d n^d \sum_{i=0}^{d} (-1)^i \binom{x}{i} \binom{n-x}{d-i}.$$

When the parameter $n$ is clear from context we abbreviate this to $\kappa_d(x)$.

It is well known that these polynomials have $d$ distinct real roots all between 0 and $n$. Furthermore, the location of the first root is well understood.

Lemma 7.10. $\kappa_d(x)$ is a degree $d$ polynomial in $x$ with $d$ distinct roots $0 < x_0^{(d)} < \cdots < x_{d-1}^{(d)} < n$. Furthermore,

$$x_0^{(d)} = \left( \frac{1}{2} - \sqrt{\frac{d}{n}} \frac{1-d}{n} \right) n \pm o(n).$$

We note that the first statement in the above lemma is standard; see 1.2.13 in [vL92]. The second statement seems to be due to [MRRW77]; see A.20 there and the paragraph after that or (17) in [Sam01].

The following lemma establishes a few facts about the eigenvalues of $S_d$.

Lemma 7.11. $S_d$ is a linear operator with the following properties for all $U \subseteq [n]$.

$$S_d \chi U = \binom{n}{d} \sum_i (-1)^i \binom{|U|}{i} \binom{n-|U|}{d-i} \chi U = \frac{(-1)^n 2^d \kappa_d(|U|)}{\binom{n}{d}} \chi U$$

$$\chi U \cdot S_d \chi U = (-1)^d \chi_U \cdot S_d \chi_U$$

$$\chi_U \cdot S_{n-d} \chi U = (-1)^{|U|} \chi_U \cdot S_d \chi U$$

Proof. Let $k = |U|$. It follows from the definition that

$$(S_d \chi U)(x) = \frac{1}{\binom{n}{d}} \sum_{y : d_{H}(x,y) = d} \chi_U (y) = \frac{\chi_U(x)}{\binom{n}{d}} \sum_{z : d_{H}(z) = d} \chi_U(z) = \frac{\chi_U(x)}{\binom{n}{d}} \sum_i (-1)^i \binom{k}{i} \binom{n-k}{d-i}$$
where in the second step we take $z = x + y$ and use $\chi_U(y) = \chi_U(z)\chi_U(x)$ and in the last step we group the different possible $z$’s based on the intersection of $U$ with their set of $-1$’s. We have established (7.4). As for (7.5) it is simple to check from (7.4). To establish (7.6) notice that $S_{n-d}f = S_dS_nf$ and $S_n\chi_U = (-1)^{|U|}\chi_U$.

The $\ell_2 \to \ell_2$ operator norm of a linear operator $T$ from the set of function $f : \{0, 1\}^n \to \mathbb{R}$ to itself is defined as

$$\|T\|_{2 \to 2} \overset{\text{def}}{=} \sup_{f \in \{0, 1\}^n \to \mathbb{R}} \frac{\|Tf\|_2}{\|f\|_2} = \sup_{f \in \{0, 1\}^n \to \mathbb{R}, \|f\|_2 = 1} \|Tf\|_2. \quad (7.7)$$

It is not hard to see that $\|T\|_{2 \to 2}$ is equal to $\max_i (|\lambda_i|)$ where $\lambda_i$’s are the eigenvalues of $T$. This norm naturally gives rise to a distance metric between operators defined as $d(T, T') \overset{\text{def}}{=} \|T - T'\|_{2 \to 2}$.

Let us remind the reader of the entropy function $H(\cdot)$ defined as

$$H(x) = -x \ln x - (1 - x) \ln(1 - x). \quad (7.8)$$

Lemma 7.12. The entropy function is concave, i.e.,

$$H(x)\alpha + H(y)(1 - \alpha) \leq H(\alpha x + (1 - \alpha)y), \quad \text{for all } 0 \leq x, y, \alpha \leq 1. \quad (7.9)$$

Proof. It is enough to show that the second derivative of $H(x)$ is non-positive between 0 and 1.

$$\frac{d^2}{dx^2}H(x) = \frac{d}{dx}(-\ln x + \ln(1 - x)) = -1/x - 1/(1 - x) < 0. \quad \blacksquare$$

Fact 7.13 (Stirling’s approximation. See [Rob55] or §2.9 in [Fel68]).

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \exp(\lambda_n) \frac{1}{12n + 1} < \lambda_n < \frac{1}{12n}. \quad (7.10)$$

The next lemma is easily obtainable from Stirling’s approximation.

Lemma 7.14. For any $0 < \alpha \leq 1/2$ we have,

$$\left(\frac{n}{\alpha n}\right) = \frac{1}{\sqrt{2\pi\alpha(1 - \alpha)n}} \exp(H(\alpha)n + \lambda_n - \lambda_{\alpha n} - \lambda_{(1-\alpha)n}) < \frac{1}{\sqrt{\pi n}} \exp(H(\alpha)n)(1 + o(1)), \quad (7.11)$$

where $H(\cdot)$ is the entropy function and $(\lambda_i)$ is the sequence in (7.10) which goes to zeros as $n \to \infty$. 

Proof. By Fact 7.13,
\[
\binom{n}{\alpha n} = \frac{n!}{(\alpha n)!(1-\alpha)n)!} = \frac{\sqrt{2\pi n} \left( \frac{n}{e} \right)^n \exp(\lambda_n)}{\sqrt{2\pi n a \alpha/n} \alpha \alpha^n \exp(\lambda_{\alpha n}) \sqrt{2\pi (1-\alpha)n} \left( \frac{1-\alpha}{\alpha} \right)^{(1-\alpha)n} \exp(\lambda_{(1-\alpha)n})}
\]
\[
= \frac{\exp(\lambda_n - \lambda_{\alpha n} - \lambda_{(1-\alpha)n})}{\sqrt{2\pi \alpha (1-\alpha) n^{\alpha/n} (1-\alpha)(1-\alpha)n}} = \frac{1}{\sqrt{2\pi \alpha (1-\alpha) n}} \exp(H(\alpha)n + \lambda_n - \lambda_{\alpha n} - \lambda_{(1-\alpha)n}).
\]

\section{7.2 Proofs}

We will first show that Theorem 7.2 implies Theorem 7.1. This should be expected as the former is essentially the two family version of the latter. For convenience of the reader, we restate Theorem 7.1.

\textbf{Theorem 7.1 (restated).} For all $0 < \delta < 1$ and $0 < \alpha \leq 1$, there exists $\epsilon = \epsilon(\delta, \alpha) > 0$ such that for any subset $U \subseteq \{0,1\}^n$ of density at least $\alpha$, we have

\[
\Pr_{x,y}[x \in U, y \in U] \geq \epsilon - o(1),
\]

where $x$ is chosen uniformly at random from $\{0,1\}^n$ and $y$ is chosen at random so that $d_H(x,y) = 2\lfloor \delta n/2 \rfloor$, i.e. $\delta n$ rounded down to the closest even number.

Furthermore, one can take $\epsilon = 2(\alpha/2)^{1-\lfloor \delta n/2 \rfloor}$. If all elements of $U$ have the same parity, then it is possible to take $\epsilon = 2\alpha^{1-\lfloor \delta n/2 \rfloor}$.

\textit{Proof.} Write $U$ as the disjoint union $U = U_o \cup U_e$, where $U_o$ (respectively, $U_e$) is the subset of $U$ consisting of odd (respectively, even) Hamming weight points. Notice that any two members of $U_o$ (respectively, $U_e$) have even Hamming distance. Also note that from $U_o$ and $U_e$ at least one has density at least $\alpha/2$. We apply Theorem 7.2 with $U' = W' = U_e$ or $U' = W' = U_o$ whichever is larger. Without loss of generality assume that $|U_o| \geq |U|/2$,

\[
\Pr_{x,y}[x \in U, y \in U] \geq \Pr_{x,y}[x \in U_o, y \in U_o] = 2 \Pr_{x',y'}[x' \in U', y' \in W'] \geq 2\epsilon(\delta, \alpha/2, \alpha/2) - o(1),
\]

where $x, y$ are distributed according to the probability distribution of Theorem 7.1 and $x', y'$ are distributed according to the probability distribution of Theorem 7.2. So the theorem holds with $\epsilon(\delta, \alpha) = 2\epsilon(\delta, \alpha/2, \alpha/2)$. \hfill \&

In the language of the $S_d$ operator it is simple to see that Theorem 7.2 can be rewritten as,

\textbf{Theorem 7.2 (restated).} For all $0 < \delta < 1$ and $0 < \alpha_1, \alpha_2 \leq 1$ there exists $\epsilon' = \epsilon'(\delta, \alpha_1, \alpha_2) > 0$ such that for any two functions $f, g : \{0,1\}^n \rightarrow \{0,1\}$ such that $\mathbb{E}[f] \geq \alpha_1$ and $\mathbb{E}[g] \geq \alpha_2$, we have,

\[
f : \frac{S_d + S_{d+1}}{2} \geq \epsilon' - o(1),
\]

where $d = \lfloor \delta n \rfloor$. \hfill \&
Figure 7.1: Comparison of $\chi_U \cdot S_d \chi_U$ and $\chi_U \cdot T_\rho \chi_U$ for $n = 100$, $\rho = 1 - 2d/n$, $d = 6, 7$ for sets, $U$, of size 0 to $n$. The horizontal axis is $|U|$ and the vertical is $\chi_U \cdot S_d \chi_U$ and $\chi_U \cdot T_\rho \chi_U$.

Notice that Theorem 7.2 would follow from Theorem 7.9 if we could show that $f \cdot \frac{S_d + S_{d+1}}{2} g$ is close to $f \cdot T_\rho g$ in operator norm for the suitable $\rho$ and any two functions $f$ and $g$ (of, say, bounded $\ell_2$ norm.) It follows from (7.1) and Lemma 7.11 that $T_\rho$ and $\frac{S_d + S_{d+1}}{2}$ have the same eigenfunctions (i.e. $\chi_U$ for $U \subseteq [n]$) so it is enough to show that each eigenvalue of $T_\rho$ is close to the corresponding eigenvalue of $\frac{S_d + S_{d+1}}{2}$. Before we start the proof, Figure 7.1 compares the eigenvalues of $T_\rho$ and $S_d$ for even and odd $d$. Notice how the spectrum of $S_d$ and $T_\rho$ are close for $|U| \leq n/2$ and depending on parity of $d$ the spectrum of $S_d$ is either symmetric or antisymmetric around $|U| = n/2$. Adding $S_d$ and $S_{d+1}$ would in effect cancel out the right side of the diagram where the eigenvalues of $T_\rho$ and $S_d$ disagree. The rest of this section is in effect proving that the Figure 7.1 is accurate for general $n$ and $d$.

The rest of the proof is structured as follows. We first prove Lemma 7.15 which shows that the eigenvalues of $S_d$ and $T_\rho$ corresponding to $\chi_U$ for small sets $U$ are close. Notice that when $|U|$ is large the eigenvalue of $T_\rho$ corresponding to $\chi_U$ is $\rho^{|U|}$ which is exponentially small. We will continue the proof by showing that the corresponding eigenvalues of $S_d$ are also small (in absolute value) in Lemma 7.16 whenever $|U|$ is large but no larger than $n/2$. Unfortunately, we can only show that these eigenvalues are no bigger than $2\sqrt{\frac{n}{d|U|}}$. This means that to get a good bound for very small $d$ one has to do a careful analysis to tie the range of Lemma 7.15 and Lemma 7.16, this is done in Theorem 7.17 to conclude that for $|U| \leq n/2$ the eigenvalues of $T_\rho$ and $S_d$ corresponding to $\chi_U$ are close. Finally, while the eigenvalues of $S_d$ corresponding
to $\chi_U$ for $|U| \gg n/2$ are large in absolute value and not close to the corresponding eigenvalues of $T_{\rho}$ it is not very hard to conclude from (7.5) in Lemma 7.11 that the eigenvalues of $S_d$ and $S_{d+1}$ corresponding to $|U| > n/2$ are very close in absolute value and have opposite sign. The calculations to show these precisely are carried out in the proof of Lemma 7.18 to show that $\frac{S_d + S_{d+1}}{2}$ and $T_{\rho}$ are close in operator norm. Finally, once we have established that these operators are close in operator norm we can proceed to prove Theorem 7.2.

We start by proving the following Lemma which establishes that eigenvalues of $T_{\rho}$ and $S_d$ corresponding to $\chi_U$ for small $U$ are close.

**Lemma 7.15.** Let $d \leq n/2$, $k = |U| \leq n/2$, and define $\rho = 1 - 2d/n$. We have,

$$\chi_U \cdot S_d \chi_U = \chi_U \cdot T_{\rho} \chi_U \pm 2 \left( \exp\left(\frac{k^2}{4(n-k/2)}\right) - 1 \right)$$

(7.12)

**Proof.** Remember that from (7.1) and Lemma 7.11 we have,

$$\chi_U \cdot T_{\rho} \chi_U = \sum_i (-1)^i \binom{k}{i} \left( 1 - \rho \right)^i \left( 1 + \rho \right)^{k-i} = \sum_{i=0}^k (-1)^i \binom{k}{i} \left( \frac{d}{n} \right)^i \left( 1 - \frac{d}{n} \right)^{k-i},$$

$$\chi_U \cdot S_d \chi_U = \sum_{i=0}^k (-1)^i \binom{k}{i} \left( \frac{n-k}{d} \right)^i \left( \frac{n}{d} \right).$$

To show (7.12) we will approximate each term of the second sum by the corresponding term in the first sum.

$$\frac{(n-k)}{d-i} \leq \frac{(n-k)!}{(n-d)!} \leq \frac{(d)^i(n-d)_{k-i}}{(n)_k} \leq \frac{(d)^i(n-d)_{k-i}}{(n-i)_{k-i}} \leq \frac{(d)^i(n-d)_{k-i}}{(n-i)_{k-i}} \leq \frac{(d)^i(n-d)_{k-i} \exp\left(\frac{k^2}{4(n-k/2)}\right)}{(n-i)_{k-i}}.$$

where we have used $\frac{a}{b} \leq \frac{a+c}{b+c}$ whenever $b, c \geq 0$ and $a \leq b$ and $i(k-i) \leq k^2/4$. Similarly,

$$\frac{(n-k)}{d-i} \leq \frac{(d)^i(n-d)_{k-i}}{(n-k+i))_{k-i}} \leq \frac{(d)^i(n-d)_{k-i}}{(n-k+i))_{k-i}} \leq \frac{(d)^i(n-d)_{k-i} \exp\left(\frac{k^2}{4(n-k/2)}\right)}{(n-k+i))_{k-i}}.$$

$$\chi_U \cdot S_d \chi_U = \sum_i (-1)^i \binom{k}{i} \left( \frac{n-k}{d-i} \right) = 1 - 2 \sum_{i: \text{odd}} \binom{k}{i} \left( \frac{n-k}{d-i} \right) \leq 1 - 2 \exp\left(\frac{k^2}{4(n-k/2)}\right) \sum_{i: \text{odd}} \frac{(k)^i(n-d)^{k-i}d^i}{n^k}.$$
\[\geq 1 - 2 \sum_{i: \text{odd}} \binom{k}{i} \frac{(n-d)^{k-i}d^i}{n^k} - 2\exp\left(\frac{k^2}{4(n-k/2)}\right) - 1 = \chi_U \cdot T\rho \chi_U - 2\exp\left(\frac{k^2}{4(n-k/2)}\right) - 1,\]

where we have used \(\sum_i \binom{k}{i} (n-k)^{d-i} = n\) and \(\sum_{i: \text{odd}} \binom{k}{i} (n-d)^{k-i}d^i \leq \sum_i \binom{k}{i} (n-d)^{k-i}d^i = 1\). Similarly,

\[
\chi_U \cdot Sd\chi_U = -1 + 2 \sum_{i: \text{even}} \binom{k}{i} \frac{(n-k)^{d-i}}{n^k} \leq -1 + 2\exp\left(\frac{k^2}{4(n-k/2)}\right) \sum_{i: \text{even}} \binom{k}{i} (n-d)^{k-i}d^i
\]

\[
\leq -1 + 2 \sum_{i: \text{even}} \frac{(k)(n-d)^{k-i}d^i}{n^k} + 2\exp\left(\frac{k^2}{4(n-k/2)}\right) - 1
\]

\[
= \chi_U \cdot T\rho \chi_U + 2\exp\left(\frac{k^2}{4(n-k/2)}\right) - 1. \tag{7.13}
\]

Lemma 7.15 is sufficient for small sets \(|U| \ll \sqrt{n}\) but becomes meaningless for \(|U| > \sqrt{n}\). We can show another bound for mid values of \(|U|\) using several properties of the Krawtchouk Polynomials. In particular, note that when \(|U|d/n\) is large, say superlogarithmic, the value of \(\chi_U \cdot T\rho \chi_U\) is very small, i.e. subpolynomial, while the error term in (7.12) can be very large. The following Lemma shows that the value \(\chi_U \cdot Sd\chi_U\) is small for large \(|U|\).

**Lemma 7.16.** For any \(k = |U| \leq n/2\) and \(d \leq n/2\) bigger than a constant we have,

\[
|\chi_U \cdot Sd\chi_U| \leq 2\sqrt{\frac{n}{dk}} \tag{7.13}
\]

**Proof.** According to (7.4),

\[
\chi_U \cdot Sd\chi_U = \frac{1}{\binom{n}{d}} \sum_i (-1)^i \binom{k}{i} \binom{n-k}{d-i}.
\]

We will argue that the absolute value of the terms in sum (as a function of \(i\)) is increasing for \(i = 0, \ldots, i_0\) and decreasing for \(i = i_0, \ldots, k\). To see this notice that the ratio of any two consecutive terms is

\[
\frac{\binom{k}{i} \binom{n-k}{d-i}}{\binom{k}{i+1} \binom{n-k}{d-i-1}} = \frac{k!}{i!(k-i)!} \frac{(n-k)!}{(d-i)!(n-k-d+i)!} \frac{i!(k-i)!}{(i+1)!(k-i-1)!(d-i-1)!(n-k-d+i+1)!} = \frac{(i+1)(n-k-d+i+1)}{(k-i)(d-i)} \tag{7.14}
\]

which is clearly an increasing function of \(i\). This shows that the difference between the odd terms and the even terms is at most the biggest term in the sum, i.e.,

\[
|\chi_U \cdot Sd\chi_U| \leq \frac{1}{\binom{n}{d}} \max_i \binom{k}{i} \binom{n-k}{d-i}.
\]

Notice that the above maximum is attained when the ratio in (7.14) is equal to one, i.e. for
\[ i = i_0 \text{ where} \]
\[
\frac{(i_0 + 1)(n - k - d + i_0 + 1)}{(k - i_0)(d - i_0)} = 1 \implies (n + 2)i_0 + n - k - d - kd + 1 = 0 \implies i_0 = \frac{(k + 1)(d + 1)}{n + 2} - 1.
\]

We then have,
\[
|\chi_U \cdot S_d \chi_U| \leq \max\left(\frac{k}{\binom{n}{i_0}}\left(\frac{n - k}{d - i_0}\right)\left(\frac{n - k}{d - i_0}\right), \frac{k}{\binom{n}{i_0}}\left(\frac{n - k}{d - i_0}\right)\left(\frac{n - k}{d - i_0}\right)\right).
\]

We can now apply Lemma 7.14 to get,
\[
\frac{k}{\binom{n}{d - i}} = \sqrt{\frac{2\pi(1 - \frac{d}{n})d}{\pi i(d - i)}} \exp\left(-H\left(\frac{d}{n}\right)n - \lambda_n + \lambda_d + \lambda_{n-d}\right) \times \frac{1}{\sqrt{2\pi(1 - \frac{i}{k})i}} \exp\left(H\left(\frac{i}{k}\right)k + \lambda_k - \lambda_i - \lambda_{k-i}\right)
\]
\[
\times \frac{1}{\sqrt{2\pi(1 - \frac{d - i}{n - k})d - i)} \exp\left(H\left(\frac{d - i}{n - k}\right)(n - k) + \lambda_{n-k} - \lambda_{d-i} - \lambda_{n-k-d+i}\right).
\]

It is not hard to check the following inequalities for \( i \in \{\lfloor i_0 \rfloor, \lceil i_0 \rceil\}, \)
\[
\frac{i}{k} \leq \frac{1}{2}, \quad \frac{d - i}{n - k} \leq \frac{1}{2} + \frac{3n + 4}{(n + 2)(n - k)} < \frac{1}{2} + \frac{7}{n}, \quad \frac{d}{d - i} \leq 2 + \frac{2}{d - 1}.
\]

The first two together with \( 1 - \frac{d}{n} \leq 1 \) imply
\[
\frac{k}{\binom{n}{d - i}} \leq \sqrt{\frac{2d}{(1 - \frac{14}{n})\pi i(d - i)}} \exp\left(H\left(\frac{i}{k}\right)k + H\left(\frac{d - i}{n - k}\right)(n - k) - H\left(\frac{d}{n}\right)n + \lambda_k + \lambda_{n-k} + \lambda_d + \lambda_{n-d}\right).
\]

using (7.11) and that \( k, n - k, d, n - d \geq 1, \frac{d}{d - i} \leq 2 + 2/(d - 1) \) and \( d \) is bigger than a constant,
\[
\leq \sqrt{\frac{4e}{\pi i}} \exp\left(H\left(\frac{i}{k}\right)k + H\left(\frac{d - i}{n - k}\right)(n - k) - H\left(\frac{d}{n}\right)n\right).\]

Now we can use the concavity of the entropy function (7.9) to conclude that
\[
\leq \sqrt{\frac{4e}{\pi i}} \exp\left(H\left(\frac{i}{n}\right) + \frac{d - i}{n} - H\left(\frac{d}{n}\right)n\right) = \sqrt{\frac{4e}{\pi i}} \times 2\sqrt{\frac{n}{dk}},
\]
for \( n, \) and \( dk \) bigger than a constant. This completes the proof.

We can now combine Lemmas 7.15 and 7.16 together with Lemma 7.10 to prove the following theorem.
Theorem 7.17. For $e^2\sqrt{n} \leq d \leq n/2$ and $U \subseteq [n]$ with $|U| \leq n/2$ we have,

$$\chi_U \cdot S_d\chi_U = \chi_U \cdot T_{\rho}\chi_U \pm \max\left(\frac{5}{\sqrt{n}}, 2\frac{n}{d^2} \ln^2 \frac{d^2}{n}\right), \quad (7.15)$$

In particular, for $d = \Theta(n)$,

$$\chi_U \cdot S_d\chi_U = \chi_U \cdot T_{\rho}\chi_U \pm O\left(\frac{1}{\sqrt{n}}\right), \quad (7.16)$$

and for $d = \omega(\sqrt{n})$

$$\chi_U \cdot S_d\chi_U = \chi_U \cdot T_{\rho}\chi_U \pm o(1). \quad (7.17)$$

Proof. It is easy to see that (7.16) and (7.17) follow from the more general (7.15). Notice that it is implied from $e^2\sqrt{n} \leq n/2$ that $n > 218$; we will use this in our bounds.

We study several cases depending on $d$ and $|U|$. First assume $d \geq n/4$. We have two cases based on $|U|$; for $|U| \leq n^{2/5}$ it from Lemma 7.15 we have,

$$\left|\chi_U \cdot S_d\chi_U - \chi_U \cdot T_{\rho}\chi_U\right| \leq 2 \left(\exp\left(\frac{|U|^2}{4(n - |U|/2)}\right) - 1\right) \leq 2 \left(\exp\left(\frac{n^{4/5}}{3n}\right) - 1\right) < \frac{1}{\sqrt{n}},$$

where we have used $1/3\sqrt{n} \leq 1/3$ and $e^x < 1 + 3x/2$ for $x \leq 1/3$. For $k = |U| \geq n^{2/5}$ we will use Lemma 7.16.

$$\left|\chi_U \cdot S_d\chi_U - \chi_U \cdot T_{\rho}\chi_U\right| \leq \left|\chi_U \cdot S_d\chi_U\right| + \left|\chi_U \cdot T_{\rho}\chi_U\right| \leq 2\sqrt{\frac{n}{dk}} + |\rho|^k \leq 4\frac{1}{\sqrt{k}} + 2^{-n^{0.4}} < \frac{5}{\sqrt{n}},$$

where we have used $2^{-n^{0.4}} < \frac{1}{\sqrt{n}}$. This completes the proof for $d \geq n/4$.

For $d \leq n/4$ we will have three cases. First assume $|U| \leq \frac{n}{4} \ln \frac{d^2}{n}$. In this case we use Lemma 7.15,

$$\left|\chi_U \cdot S_d\chi_U - \chi_U \cdot T_{\rho}\chi_U\right| \leq 2 \left(\exp\left(\frac{|U|^2}{4(n - |U|/2)}\right) - 1\right) \leq 2 \exp\left(\frac{n^2 \ln^2 \frac{d^2}{n}}{3n}\right) - 1)

= 2\exp\left(\frac{n}{3d^2} \ln^2 \frac{d^2}{n} - 1\right) < \frac{n}{d^2} \ln^2 \frac{d^2}{n},$$

where in the first line we have used $|U|/2 \leq \frac{n}{21} \ln \frac{d^2}{n} \leq 2\sqrt{n} \leq n/4$ for $n \geq 64$ and in the second line we have used $\frac{n}{3d^2} \ln^2 \frac{d^2}{n} \leq 16/3e^4 < 0.1$ and $e^x < 1 + 3x/2$ for $x \leq 1/2$.

Now assume that $\frac{n}{4} \ln \frac{d^2}{n} \leq |U| \leq n/20$. Notice that Lemma 7.10 and Lemma 7.11 imply that $\chi_U \cdot S_d\chi_U$ is decreasing from $|U| = \frac{n}{4} \ln \frac{d^2}{n}$ to the first root of $\kappa_d(x)$ and that this first root
is,
\[ x_0^{(d)} \geq \left( \frac{1}{2} - \frac{d}{n(1-d)} \right) n + o(n) \geq \left( \frac{1}{2} - \frac{\sqrt{3}}{4} \right) n + o(n) \gg n/20. \]

Fixing \( U' \) as an arbitrary subset of \([n] \) of size \( n/2d \ln d^2/n \) this implies that in this range,
\[
0 \leq \chi_U \cdot S_d \chi_U \leq \chi_{U'} \cdot S_d \chi_U \leq \chi_{U'} \cdot T_\rho \chi_U \leq \frac{n}{2d} \ln^2 \frac{d^2}{n} \leq (1 - 2d/n) \frac{n}{2d} \ln \frac{d^2}{n} + \frac{n}{d^2} \ln \frac{d^2}{n} \\
\leq \exp(-2d/n) \times \frac{n}{2d} \ln^2 \frac{d^2}{n} + \frac{n}{d^2} \ln^2 \frac{d^2}{n} = \frac{n}{d^2} + \frac{n}{d^2} \ln^2 \frac{d^2}{n} \leq (1 + 1/16) \frac{n}{d^2} \ln^2 \frac{d^2}{n},
\]
\[
\chi_U \cdot T_\rho \chi_U \leq \chi_{U'} \cdot T_\rho \chi_U \leq \frac{n}{d^2},
\]

which completes the proof for this range.

Finally, when \( k = |U| \geq n/20 \) we use Lemma 7.16.
\[
\left| \chi_U \cdot S_d \chi_U - \chi_U \cdot T_\rho \chi_U \right| \leq \left| \chi_U \cdot S_d \chi_U \right| + \left| \chi_U \cdot T_\rho \chi_U \right| \leq 2 \sqrt{\frac{n}{d^2} + \rho |U|} \leq 2 \sqrt{20/d} + (1 - 2e^2 \sqrt{n}/n)^{n/20} \\
< 4/\sqrt{n} + \exp(-e^2 \sqrt{n}/10) < 5/\sqrt{n} < 5/\sqrt{n},
\]

which completes the last case of the proof.

The following Lemma follows from Theorem 7.17.

**Lemma 7.18.** For large enough \( n \) and \( d = \Theta(n) \) and \( \rho = 1 - 2d/n \),
\[
\left\| T_\rho - \frac{S_d + S_{d+1}}{2} \right\|_{2 \rightarrow 2} = O(\frac{1}{\sqrt{n}}).
\]

**Proof.** Consider an arbitrary \( f : \{0,1\}^n \rightarrow \mathbb{R} \). By (7.7),
\[
\left\| T_\rho - \frac{S_d + S_{d+1}}{2} \right\|_{2 \rightarrow 2} = \sup_{f \in \{0,1\}^n \rightarrow \mathbb{R}} \left\| \left( T_\rho - \frac{S_d + S_{d+1}}{2} \right) f \right\|_2 / \|f\|_2 \\
= \sup_f \left\| \sum_{U \subseteq [n]} \hat{f}(\{U\}) \left( T_\rho - \frac{S_d + S_{d+1}}{2} \right) \chi_U \right\|_2 / \|f\|_2 \\
= \sup_f \left\| \sum_{U \subseteq [n]} \hat{f}(\{U\})^2 \chi_U \cdot T_\rho \left( T_\rho - \frac{S_d + S_{d+1}}{2} \right) \chi_U \right\|_2 / \|f\|_2 \\
\leq \sup_f \left\| \sum_{U \subseteq [n]} \hat{f}(\{U\})^2 \max_U \chi_U \cdot \left( T_\rho - \frac{S_d + S_{d+1}}{2} \right) \chi_U \right\| / \|f\|_2 \\
= \max_U \left| \chi_U \cdot \left( T_\rho - \frac{S_d + S_{d+1}}{2} \right) \chi_U \right| \quad \text{by (2.25)}
\]
To show the conclusion of the lemma we have three cases. For $|U| = 0$ we have,

$$
\left| \chi_U \cdot \left( T_\rho - \frac{S_d + S_{d+1}}{2} \right) \chi_U \right| = \left| \rho^0 - \frac{(1)^0(\delta)}{(d+1)^0} + \frac{(1)^0(\delta)}{(d+1)^0(\delta)} \right| = 0.
$$

For $0 < |U| \leq n/2$ we have,

$$
\left| \chi_U \cdot \left( T_\rho - \frac{S_d + S_{d+1}}{2} \right) \chi_U \right|
\leq |\chi_U \cdot (T_\rho - S_d) \chi_U|/2 + |\chi_U \cdot (T_{\rho-2/n} - S_{d+1}) \chi_U|/2 + |\chi_U \cdot (T_\rho - T_{\rho-2/n}) \chi_U|/2
= O(1/\sqrt{n}) + O(1/\sqrt{n}) + |\rho| - |\rho - 2/n|^U|/2 + \max_{0 < \delta < 2/n} |U|2(|\rho - \tau|)|U|^{-1}/n
= O(1/\sqrt{n}) + O(\log n/n) = O(1/\sqrt{n}),
$$

where we have used triangle inequality to derive the second line, (7.16) and Taylor’s theorem to get the third line and a case analysis (based on whether $|U| \leq \log n$) to get the last line.

Similarly, when $|U| > n/2$ we have,

$$
\left| \chi_U \cdot \left( T_\rho - \frac{S_d + S_{d+1}}{2} \right) \chi_U \right|
\leq |\chi_U \cdot T_\rho \chi_U| + |\chi_U \cdot S_d \chi_U + \chi_U \cdot S_{d+1} \chi_U|/2
= |\chi_U \cdot S_d \chi_U + \chi_U \cdot S_{d+1} \chi_U|/2
= \exp(-\Theta(n)) + \left| \chi_U \cdot S_d \chi_U - \chi_U \cdot S_{d+1} \chi_U \right|/2
= \exp(-\Theta(n)) + O(1/\sqrt{n}) + O(1/\sqrt{n}),
$$

where in the last line we have used $|U| < n/2$ to reduce to the previous case. Putting the three cases together finishes the proof.

We are now ready to go back to the proof of Theorem 7.2 which we repeat here for convenience.

**Theorem 7.2** (restated). For all $0 < \delta < 1$ and $0 < \alpha_1, \alpha_2 \leq 1$ there exists $\epsilon' = \epsilon'(\delta, \alpha_1, \alpha_2) > 0$ such that for any two functions $f, g : \{0, 1\}^n \rightarrow \{0, 1\}$ such that $\mathbb{E}[f] = \alpha_1$ and $\mathbb{E}[g] = \alpha_2$, we have,

$$
\left| f \cdot \frac{S_d + S_{d+1}}{2} g \geq \epsilon' - \sqrt{\alpha_1 \alpha_2} O(1/\sqrt{n}),
$$

where $d = \lfloor \delta n \rfloor$ and the $O(\cdot)$ notation hides a constant independent of $\delta, \alpha_1, \alpha_2$ and $n$. One can take

$$
\epsilon' = \exp \left( \frac{\ln \alpha_1 + 2|\rho|\sqrt{\ln \alpha_1 \ln \alpha_2}}{(1 - \rho^2)} + \ln \alpha_2 \right)
$$

for $\rho = 1 - 2\delta$. 

Proof of Theorem 7.2.

\[ f \cdot \frac{S_d + S_{d+1}}{2} g = f \cdot T_\rho g + f \cdot \left( \frac{S_d + S_{d+1}}{2} - T_\rho \right) g \geq \epsilon' + f \cdot \left( \frac{S_d + S_{d+1}}{2} - T_\rho \right) g \]

\[ \geq \epsilon' - \|f\|_2 \left\| \left( \frac{S_d + S_{d+1}}{2} - T_\rho \right) g \right\|_2 \geq \epsilon' - \|f\|_2 \left\| g \right\|_2 \left\| \frac{S_d + S_{d+1}}{2} - T_\rho \right\|_{2 \rightarrow 2} \]

where we have used Theorem 7.9, Cauchy-Schwarz and Lemma 7.18 in the first, second and third lines respectively.

7.3 Application

We have the following two theorems. In both theorems the \( \Omega \) and \( O \) terms represent the asymptotics as \( d \) grows large.

**Theorem 7.4** (restated). For any fixed \( l \) the integrality gap of the level-\( l \) Sherali-Adams LP relaxation of Independent Set on graphs of bounded degree \( d \) is at least \( \Omega \left( \frac{d \log d}{\log d} \right) \).

**Theorem 7.3** (restated). For any \( l \) the integrality gap of the level-\( l \) Lovász-Schrijver SDP (\( LS^+ \)) relaxation of Vertex Cover in graphs of bounded degree \( d \) is at least \( 2 \left( 1 - \frac{1}{2} \frac{1}{12 \ln \ln d + o \left( \frac{\ln \ln d}{\ln d} \right)} \right) \).

For both these theorems we are going to start with previously known Integrality Gap results for the general case of the problem (with no degree bounds) that use the Frankl-Rödl graph as the instance. We are then going to sample the edges of this graph so that the expected degree of each node is \( d \). It is not hard to show that in the resulting graph the maximum degree is almost surely at most \( 2d \). To make the argument precise we will define the following graph.

**Definition 7.4.** For any \( n \geq 1 \) and \( \xi > 0 \) such that \( \xi n \) is an integer define the “Frankl-Rödl” graph \( G_{\xi}^n \) as follows. The vertex set of the graph \( V(G_{\xi}^n) = \{0, 1\}^n \) is the \( n \)-dimensional Hamming cube and two vertices \( x, y \in V(G) \) are connected if and only if their Hamming distance \( d_H(x, y) \) is exactly \( (1 - \xi) n \).

The above definition is consistent with that of Georgiou et. al.[GMPT07]. The following Lemma is an strengthening (due to Georgiou et. al.[GMPT07]) of one of the main theorems of Frankl-Rödl [FR87].

**Lemma 7.19** (Lemma 1 in [GMPT07]). For any \( n \geq 1 \) and \( \xi > 0 \) if \( (1 - \xi)n \) is even then the maximum independent set of \( G_{\xi}^n \) has size at most \( (n + 1)2^n (1 - \xi^2/64)^n \). In particular if \( \xi \geq 12 \sqrt{\log n}/n \) the largest independent set is of size at most \( 2^n/n \).

Lemma 7.19 serves as the “soundness” part of the integrality gap result of Georgiou et. al. For our case we need a form of the Lemma that shows that big subsets of the vertices of \( G_{\xi}^n \) are not only non-independent sets but in fact have many edges inside them. This way when we
sample a number of edges from $G_n^\xi$ at random such sets have a good chance of not becoming independent sets. Such a statement follows as a simple corollary of Theorem 7.1.

**Corollary 7.20** (of Theorem 7.1). For any $d$ and $\xi > 0$, for large enough $n$, $G_n^\xi$ has a subgraph $G''$ with $2^{n-1}$ vertices and maximum degree $d$ such that, $IS(G'') < (9 \ln(d/8)/d)^{\xi/(1-\xi)} + 2^{-d/8}$.

**Proof.** Let $n$ be even and define $G'$ to be the set of all vertices of $G_n^\xi$ of even Hamming weight. Consider $G''$ defined as a random induced subgraph of $G'$ where $d2^{n-3}$ edges of $G'$ are selected to be present in $G''$ at random with replacement. For a vertex $x \in V(G')$ the probability that the degree of $x$ in $G''$ is bigger than $d$ is at most,

$$\Pr[d(x) > d] = \Pr\left[d(x) > 2 \mathbb{E}[d(x)] + (e/2^2)^{\mathbb{E}[d(x)]} = (e/4)^{d/2}\right]$$

where we have used a multiplicative form of Chernoff bound observing that $d(x)$ as a random variables is the sum of $d2^{n-3}$ independent random variables the ith of which indicates if the ith edge of $G''$ is a neighbour of $x$. Let $B \subseteq V(G'')$ be the set of all vertices of degree more than $d$, applying Markov’s inequality we have,

$$\Pr\left[|B| > |V(G'')|2^{-d/8}\right] \leq \frac{\mathbb{E}[|B|]}{|V(G'')|2^{-d/8}} < \frac{|V(G'')|(e/4)^{d/2}}{|V(G'')|2^{-d/8}} = (e\sqrt{2}/4)^{d/2} < 2^{-d/8}. \quad (7.18)$$

Define $G'''$ as the subgraph of $G''$ where we delete all the edges of any vertex of degree more than $d$. It is clear that the maximum degree of $G'''$ is at most $d$ and that $IS(G''') \leq IS(G'') + |B|/|V(G'')|$. Thus,

$$\Pr[IS(G''') > IS(G'') + 2^{-d/8}] < 2^{-d/8}. $$

On the other hand consider a set $S \subseteq |V(G'')|$ of size $\mu 2^{n-1}$. Applying Theorem 7.1 with $\delta = (1 - \xi)$ the probability that a random edge of $G'$ lands inside $S$ is at least,

$$\Pr_{G}[e_i \in S \times S] \geq \mu^{2(1+|\rho|)/(1-\rho^2)} - o(1) = 2(\mu)^{2/(1-|\rho|)} - o(1) = \mu^{1/\xi} - o(1), $$

$$\Pr_{G}[S \text{ is independent}] < (1 - \mu^{1/\xi} + o(1))d2^{n-3}, $$

$$\Pr_{G}[IS(G'') > \mu] < (1 - \mu^{1/\xi} + o(1))d2^{n/8}(\frac{2^{n-1}}{\mu2^{n-1}}) \leq \exp(-\mu^{1/\xi}d2^{n/8} + o(2^n)) \exp(2^{n-1}H(\mu))/\sqrt{\mu2^{n-1}} = \exp\left(2^{n-1}\left(H(\mu) - \mu^{1/\xi}d/4 + o(1)\right)\right) \leq \exp\left(2^{n-1}\left(2\mu \ln(1/\mu) - \mu^{1/\xi}d/4 + o(1)\right)\right), \quad (7.19)$$

where we have used the fact that $\xi$ and $\mu$ are constants (dependent on $d$ but independent of $n$) in the first line, Lemma 7.14 in the third line and assuming $\mu < 1/3$ in the fourth line. Choosing $\mu = (9 \ln(d/8)/d)^{\xi/(1-\xi)}$ will make $(7.19) \in o(1)$.
Combining (7.18) and (7.19) we get that $G''$ has degree at most $d$ and,

$$\Pr \left[ IS(G'') > (9 \ln(d/8)/d)^{\xi/(1-\xi)} + 2^{-d/8} \right] < 2^{-d/8} + o(1). \quad \square$$

### Integrality Gap for Vertex Cover

We will use the following theorem of Georgiou et al.

**Theorem 7.21** (Reformulation of Theorem 1 of [GMPT07]). For all $\xi > 0$, $\epsilon > 5\xi$ a sufficiently small constant and $n$ a sufficiently big integer, if $1/\xi$ is even then the level-$\lceil \epsilon/6\xi \rceil$ Lovász-Schrijver SDP relaxation for VERTEX COVER on $G_n^\xi$ has a solution of objective value $2^n(1/2+\epsilon)$. Furthermore, in this solution each vertex contributes $1/2 + \epsilon$ to the objective function.

**Proof.** Let $\epsilon > 0$ be a small constant to be fixed later and let $\xi = \frac{\epsilon}{6l}$ be such that $1/\xi$ is an integer (and hence $1/\xi$ even.) The instance of the integrality gap will be $G''$ from Corollary 7.20.

We first show that Theorem 7.21 implies that the value of the level-$l$ LS$^+$ relaxation is at most $2^n - 1(1/2 + \epsilon)$. This is because $G'''$ has half the vertices of $G_n^\xi$ and some of its edges. Given that removing vertices and edges from a graph can only introduce new solutions to the LS$^+$ relaxations, the solution in Theorem 7.21 is a valid solution for level-$l$ LS$^+$ applied to our graph. Furthermore, given that the contribution of each vertex to this solution (for $G_n^\xi$) is exactly $1/2 + \epsilon$ its total objective value is exactly $2^n - 1(1/2 + \epsilon)$ for $G''$.

On the other hand it follows from Corollary 7.20 that $G'''$ has maximum degree at most $d$ and no vertex cover of size less than $2^n - 1(1 - (9 \ln(d/8)/d)^{\xi/(1-\xi)} - 2^{-d/8})$. Combining these the integrality gap is at least,

$$IG \geq \frac{2^n - 1(1 - (9 \ln(d/8)/d)^{\xi/(1-\xi)} - 2^{-d/8})}{2^n - 1(1/2 + \epsilon)} > 2(1 - (9 \ln(d/8)/d)^{\xi/(1-\xi)} - 2^{-d/8})(1 - 2\epsilon)$$

$$> 2(1 - (9 \ln(d/8)/d)^{\xi/(1-\xi)} - 2^{-d/8} - 2\epsilon) \geq 2 \left( 1 - 12l \frac{\ln d}{\ln d} + o_d \left( \frac{\ln \ln d}{\ln d} \right) \right).$$

where the last line follows for $\epsilon = 6l \frac{\ln (d/9 \ln(d/8)) - \ln \ln 12l}{\ln(d/9 \ln(d/8))}$. \quad \square

### Integrality Gap for Independent Set

As the starting point for our proof of Theorem 7.4 we will use the following corollary of Lemma 6.3.

**Corollary 7.22** (of Lemma 6.3). For any constant $l$ and large enough $n$, the objective value of the level-$l$ Sherali-Adams LP relaxation of INDEPENDENT SET for $G_n^{1/2}$ is $\Omega(2^n)$, where $\Omega()$ hides a constant (less than 1) that depends on $l$ but not $n.$
Proof. By running Algorithm 5 with \( t = l + 1, \delta \) one gets a family of consistent local distributions for Vertex Cover. Flipping the status of all the variables results is a family of local distributions for Independent Set. Notice that we are running Algorithm 5 with \( \delta = 1/2 \) instead of the typical value \( \delta \ll 1 \), however the only effect is that the size of the Independent Set will be much smaller, in fact, exponentially small in \( t \) however, for any fixed \( l \) the value will be still \( \Omega(2^n) \).

Now Theorem 7.4 follows easily from Corollaries 7.20 and 7.22.

**Theorem 7.4** (restated). For any fixed \( l \) the integrality gap of the level-\( l \) Sherali-Adams LP relaxation of Independent Set on graphs of bounded degree \( d \) is at least \( \Omega(\frac{d}{\log d}) \).

Proof. Let \( \xi = 1/2 \) and let the instance of the integrality gap be \( G'''' \) from Corollary 7.20.

On one hand according to Corollary 7.22 the value of the relaxation on \( G'''' \) is at least \( \Omega(2^n) \).

On the other hand we know from Corollary 7.20 that the size of the biggest independent set is at most,

\[
2^n((9 \ln(d/8)/d)^{\xi/(1-\xi)} + 2^{-d/8}) = 2^n(9 \ln(d/8)/d + 2^{-d/8}) < 2^n 10 \ln(d/8)/d.
\]

The theorem follows by dividing these two quantities.
Chapter 8

Conclusion

We studied the integrality gap of the Sherali-Adams SDP relaxations for various problems. The most interesting remaining problem in this area is to prove tight lower bounds for the integrality gap of the stronger Lasserre relaxations. In fact, there are very few examples of such tight lower bounds. Schoenebeck [Sch08], Tulsiani [Tul09] and Grigoriev [Gri01]\(^1\) are three notable exceptions. Tight integrality gaps for the Unique Label Cover problem (the optimization problem which is the subject of UGC) in particular would be very interesting. We note that the recent work of Barak et al. [BBH\(^+\)12] indicates that the Lasserre relaxations might have no integrality gap on many of the instances previously thought to be good candidates for integrality gap constructions.

8.1 MAX \(k\)-CSP\(_q\)(\(P\))

We showed tight integrality gaps for level-\(\Theta(n)\) Sherali-Adams SDP relaxations of MAX \(k\)-CSP\(_q\)(\(P\)) when \(P\) is promising. Our construction heavily uses the expansion properties of the integrality gap instance and is very generic.

We also gave approximation algorithms for MAX \(k\)-CSP(\(P\)) for two classes of predicates \(P\) not previously known to be approximable. The first case, signs of symmetric quadratic forms, follows from the condition that the low-degree part of the Fourier expansion behaves “roughly” like the predicate in the sense of Theorem 4.6. The second case, Monarchy, is interesting since it does not satisfy the condition of Theorem 4.6. As far as we are aware, this is the first example of a predicate which does not satisfy this property but is still approximable. Monarchy is of course only a very special case of Conjecture 4.3 which we leave open.

A further interesting special case of the conjecture is a generalization of Monarchy called “republic”, defined as sign\(\left(\frac{k}{2}x_1 + \sum_{i=2}^{k} x_i\right)\). In this case the \(x_1\) variable needs to get a \(1/4\) fraction of the other variables on its side. We do not even know how to handle this seemingly

\(^1\)Grigoriev [Gri01]’s lowerbound is for the degree parameter of the, so-called, “Positivstellenstatz Calculus Proofs” for MAX \(k\)-XOR. It is now known that such lowerbounds imply integrality gap lowerbounds for the Lasserre hierarchy.
innocuous example.

It is interesting that the condition on $P$ for our $(\epsilon, \eta)$-rounding to succeed turned out to be precisely the same as the condition previously found by Hast [Has05a] using a completely different algorithm. It would be interesting to know whether this is a coincidence or there is a larger picture that we cannot yet see.

In general there are not many results that give integrality gap lower bounds or approximation algorithms for MAX $k$-CSP$_q(P)$ for a large classes of predicates, $P$. Any new such result would be very interesting.

### 8.2 Quadratic Programming and Max Cut Gain

We saw that the level-poly($n$) Sherali-Adams SDP relaxation of QUADRATIC PROGRAMMING has the same asymptotic integrality gap as the canonical SDP, namely $\Omega(\log n)$. It is interesting to see other problems for which this kind of construction can prove meaningful integrality gap results. It is easy to see that as long as a problem does not have “hard” constraints, and a super constant integrality gap for its canonical SDP relaxation is known, one can get super constant integrality gaps for super constant levels of the Sherali-Adams SDP hierarchy just by plugging in the uniform distribution for $\nu$ in Lemma 5.3.

It is possible to show that the same techniques apply when the objective function is a polynomial of degree greater than 2 (but still constant.) This is particularly relevant to MAX CSP($P$) problems. When formulated as a maximization problem of a polynomial $q$ over $\pm 1$ valued variables, $q$ will have degree $r$, the arity of $P$. In order to adapt Lemma 5.3 to this setting, the Fourier expansion of $f_S$’s should be adjusted appropriately. Specifically, their Fourier expansion would match that of the starting SDP solution up to level $r$ and that of $\nu$ beyond level $r$. It is also possible to define “gain” versions of MAX CSP($P$) problems in this setting and extend superconstant integrality gaps to the Sherali-Adams SDP of superconstant level.

### 8.3 Vertex Cover

We presented tight integrality gaps for the level-5 Sherali-Adams SDP relaxation of VERTEX COVER. We also showed that if a certain analytical conjecture is proved these can be extended to any constant number of rounds. Along the way we also gave an intuitive and geometric proof of tight Sherali-Adams LP integrality gaps for the same problem. While these LP integrality gaps apply to less rounds than [CMM09] they remain highly nontrivial, yet significantly simplified.

For large $t$, proving Conjecture 6.18 seems challenging. We leave it as an open problem. Another important (but hard) open problem is to extend the ideas to construct tight Lasserre gaps for Vertex Cover, thus giving the strongest evidence that VERTEX COVER cannot be solved using SDP hierarchies.
8.4 Isoperimetric Inequality and Integrality Gaps in Degree-bounded Graph

We proved a density variant of the Frankl-Rödl theorem [FR87] using tools from discrete Fourier analysis. An immediate observation about Theorems 7.1 and 7.2 is that some form of an error term (the $-o(1)$ factor on the right hand side) is required. This is because for sufficiently small (sub-constant) values of $\mu$ the quantity on the left hand side is simply 0. The error term we get however is polynomially small in $n$, where as (by [FR87]) for left hand side to vanish $\mu$ has to be exponentially small. In other words it might be possible to improve the error term to an exponentially small quantity. In fact such an improvement would make Theorems 7.1 and 7.2 strong enough to imply the original theorem of [FR87]. Given that the proof of [FR87] is combinatorial and iterative in nature where as ours is analytical this would be a very interesting alternative proof of the Frankl-Rödl theorem. Such a proof might have applications in the context of the new results of Barak et al. [BBH+12].

We note that by contrast Lemma 7.18 seems to require a polynomially small error term as even the low eigenvalues of $S_d$ and $S_{d+1}$ are at least polynomially different. So if one hopes to prove Theorems 7.1 and 7.2 with a smaller error term one needs to avoid Lemma 7.18.

We also showed how our theorems can be used to prove integrality gap lower bounds for graph problems in the presence of bounds on the maximum degree of the graphs. Our applications are preliminary at best. In fact, (with enough effort) one should be able to extend Theorem 7.4 to the Sherali-Adams SDP relaxation by making the machinery of Section 6.3 work in this setting.
Bibliography


[ABL02] Sanjeev Arora, B{\textsc{\i}}la Bollob{\textsc{\i}}}s, and L{\textsc{\aa}}szl{\textsc{o}} Lov{\textsc{\aa}}sz, *Proving integrality gaps without knowing the linear program*, FOCS’02, 2002, pp. 313–322.


[GMPT07] Konstantinos Georgiou, Avner Magen, Toniann Pitassi, and Iannis Tourlakis, *Integality Gaps of $2 − o(1)$ for Vertex Cover SDPs in the Lovász-Schrijver Hierarchy*,


Personal communication, 2009.


Subhash Khot and Oded Regev, *Vertex Cover Might be Hard to Approximate to Within 2 − ε*, Journal of Computer and System Sciences 74 (2008), no. 3, 335–349.


[Sch08] Grant Schoenebeck, Linear Level Lasserre Lower Bounds for Certain k-CSPs, FOCS’08, 2008, pp. 593–602.


[ST00] Alex Samorodnitsky and Luca Trevisan, A PCP characterization of NP with optimal amortized query complexity, STOC’00, 2000, pp. 191–199.


[Tou05] Iannis Tourlakis, Towards optimal integrality gaps for hypergraph vertex cover in the Lovász-Schrijver hierarchy, APPROX’05, 2005, pp. 233–244.


# Appendix A

## Glossary of Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Page or Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{E}$</td>
<td>Expectation.</td>
<td>Chapter 2, p. 6</td>
</tr>
<tr>
<td>$[n]$</td>
<td>${1, \ldots, n}$.</td>
<td>Chapter 2, p. 6</td>
</tr>
<tr>
<td>$\varphi(S)$</td>
<td>Power set of $S$.</td>
<td>Chapter 2, p. 6</td>
</tr>
<tr>
<td>$\binom{S}{k}$</td>
<td>$k$th falling power of $n$.</td>
<td>Chapter 2, p. 6</td>
</tr>
<tr>
<td>$\binom{S}{\leq k}$</td>
<td></td>
<td>Chapter 2, p. 6</td>
</tr>
<tr>
<td>$(n)_k$</td>
<td>$n$ vector concatenation.</td>
<td>Chapter 2, p. 6</td>
</tr>
<tr>
<td>$x \circ y$</td>
<td>Projection.</td>
<td>Chapter 2, p. 6</td>
</tr>
<tr>
<td>$\alpha(S)$</td>
<td>Marginal.</td>
<td>(2.1), p. 6</td>
</tr>
<tr>
<td>Mar$_S \mu$</td>
<td>$S^n$-sphere of radius 1.</td>
<td>Chapter 2, p. 7</td>
</tr>
<tr>
<td>$\varphi, \Phi$</td>
<td>PDF/CDF of a Standard Normal Random Variable.</td>
<td>(2.2), p. 6</td>
</tr>
<tr>
<td>$S^n$</td>
<td>Spherical cap</td>
<td>Chapter 2, p. 7</td>
</tr>
<tr>
<td>Gram matrix</td>
<td></td>
<td>Definition 2.1, p. 7</td>
</tr>
<tr>
<td>$A \succeq 0$</td>
<td>Positive Semidefinite.</td>
<td>Chapter 2, p. 7</td>
</tr>
<tr>
<td>$\text{sign}(x)$</td>
<td>$x$-approximation</td>
<td>(2.3), p. 7</td>
</tr>
<tr>
<td>$IS(G)$</td>
<td>Maximum independent set of $G$.</td>
<td>Chapter 2, p. 7</td>
</tr>
<tr>
<td>$VC(G)$</td>
<td>Minimum vertex cover of $G$.</td>
<td>Chapter 2, p. 7</td>
</tr>
<tr>
<td>Approximation ratio</td>
<td></td>
<td>Section 2.1, p. 7</td>
</tr>
<tr>
<td>$\alpha$-approximation</td>
<td></td>
<td>Section 2.1, p. 8</td>
</tr>
<tr>
<td>Approximability Threshold</td>
<td></td>
<td>Section 2.1, p. 8</td>
</tr>
<tr>
<td>Vertex Cover</td>
<td></td>
<td>Subsection 2.1.1, p. 8</td>
</tr>
<tr>
<td>Independent Set</td>
<td></td>
<td>Subsection 2.1.1, p. 8</td>
</tr>
<tr>
<td>Max Cut</td>
<td></td>
<td>Subsection 2.1.2, p. 8</td>
</tr>
<tr>
<td>Max Cut Gain</td>
<td></td>
<td>Subsection 2.1.2, p. 9</td>
</tr>
<tr>
<td>Quadratic Programming</td>
<td></td>
<td>Definition 2.2, p. 9</td>
</tr>
<tr>
<td>MAX $k$-CSP$_q(P)$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix A. Glossary of Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition or Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clause</td>
<td></td>
</tr>
<tr>
<td>MAX $k$-CSP$_q$</td>
<td>Definition 2.2, p. 10</td>
</tr>
<tr>
<td>Approximation Resistant</td>
<td></td>
</tr>
<tr>
<td>Approximable</td>
<td></td>
</tr>
<tr>
<td>Promising Predicate</td>
<td></td>
</tr>
<tr>
<td>IP</td>
<td>Integer Program</td>
</tr>
<tr>
<td>QP</td>
<td>Quadratic Program</td>
</tr>
<tr>
<td>Relaxation</td>
<td></td>
</tr>
<tr>
<td>LP</td>
<td>Linear Program</td>
</tr>
<tr>
<td>Vector Program</td>
<td></td>
</tr>
<tr>
<td>SDP</td>
<td>Semidefinite Program</td>
</tr>
<tr>
<td>IG</td>
<td>Integrality Gap</td>
</tr>
<tr>
<td>SA LP relaxation</td>
<td>level-$k$ Sherali-Adams LP relaxation.</td>
</tr>
<tr>
<td>SA SDP relaxation</td>
<td>level-$k$ Sherali-Adams SDP relaxation.</td>
</tr>
<tr>
<td>Lasserre relaxation</td>
<td>level-$k$ Lasserre relaxation.</td>
</tr>
<tr>
<td>local distributions</td>
<td></td>
</tr>
<tr>
<td>compatibility</td>
<td></td>
</tr>
<tr>
<td>${ -1, 1 }^n$</td>
<td>Hamming cube</td>
</tr>
<tr>
<td>$w_H(\cdot)$</td>
<td>Hamming weight</td>
</tr>
<tr>
<td>$d_H(\cdot, \cdot)$</td>
<td>Hamming distance</td>
</tr>
<tr>
<td>$\chi_S$</td>
<td>Characters, Fourier basis</td>
</tr>
<tr>
<td>$\hat{f}(S)$</td>
<td>Fourier coefficient</td>
</tr>
<tr>
<td>$f = d$</td>
<td></td>
</tr>
<tr>
<td>$|f|_2$</td>
<td>Euclidian norm</td>
</tr>
<tr>
<td>$G_{\gamma}^m$</td>
<td>The Frankl-Rödl Graph</td>
</tr>
<tr>
<td>$B_{\gamma}^n$</td>
<td>The Borsuk Graph</td>
</tr>
<tr>
<td>$G_{\Phi}$</td>
<td>Constraint graph</td>
</tr>
<tr>
<td>$\Gamma(C_i), \Gamma(x_j), \Gamma(\mathcal{C})$</td>
<td>Neighbor Set</td>
</tr>
<tr>
<td>$\mathcal{C}(S)$</td>
<td>$S$-dominated constraints</td>
</tr>
<tr>
<td>$G_{\downarrow S}$</td>
<td></td>
</tr>
<tr>
<td>$\partial X$</td>
<td>Boundary of $X$</td>
</tr>
<tr>
<td>Boundary expansion</td>
<td></td>
</tr>
<tr>
<td>Monarchy</td>
<td></td>
</tr>
<tr>
<td>$v \otimes u$</td>
<td>Tensor product</td>
</tr>
<tr>
<td>$v^{\otimes d}$</td>
<td>Tensor power</td>
</tr>
<tr>
<td>$T_P$</td>
<td>Tensorsing using a polynomial</td>
</tr>
<tr>
<td>$T_\rho$</td>
<td>Bonami-Beckner operator</td>
</tr>
<tr>
<td>$S_d$</td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
</tr>
<tr>
<td>$r_d^{(n)}(x)$</td>
<td>Krawtchouk Polynomial</td>
</tr>
<tr>
<td>$|T|_{2\to 2}$</td>
<td>$\ell_2 \to \ell_2$ operator norm</td>
</tr>
<tr>
<td>$H(x)$</td>
<td>Entropy Function</td>
</tr>
</tbody>
</table>