Local Constraints in Combinatorial Optimization

by

Madhur Tulsiani

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Computer Science

in the

GRADUATE DIVISION
of the
UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor Luca Trevisan, Chair
Professor Alistair Sinclair
Professor Satish Rao
Professor David Aldous

Fall 2009

Local Constraints in Combinatorial Optimization

Abstract

Local Constraints in Combinatorial Optimization

by

Madhur Tulsiani
Doctor of Philosophy in Computer Science

University of California at Berkeley

Professor Luca Trevisan, Chair

Hard combinatorial optimization problems are often approximated using linear or semidefinite programming relaxations. In fact, most of the algorithms developed using such convex programs have special structure: the constraints imposed by these algorithms are *local* i.e. each constraint involves only few variables in the problem. In this thesis, we study the power of such local constraints in providing an approximation for various optimization problems.

We study various models of computation defined in terms of such programs with local constraints. The resource in question for these models is are the sizes of the constraints involved in the program. Known algorithmic results relate this notion of resources to the time taken for computation in a natural way.

Such models are provided by the "hierarchies" of linear and semidefinite programs, like the ones defined by Lovász and Schrijver[LS91]; Sherali and Adams[SA90]; and Lasserre [Las01]. We study the complexity of approximating various optimization problems using each of these hierarchies.

This thesis contains various lower bounds in this computational model. We develop techniques for reasoning about each of these hierarchies and exhibiting various combinatorial objects whose local properties are very different from their global properties. Such lower bounds *unconditionally* rule out a large class of algorithms (which captures most known ones) for approximating the problems such as MAX 3-SAT, Minimum Vertex Cover, Chromatic Number and others studied in this thesis.

We also provide a positive result where a simple semidefinite relaxation is useful for approximating a constraint satisfaction problem defined on graphs, if the underlying graph is expanding. We show how expansion connects the local properties of the graph to the global properties of interest, thus providing a good algorithm.

Acknowledgements

The past four years that I spent in Berkeley have been an extremely interesting and educative experience, much of which I owe to my advisor Luca Trevisan. I arrived at Berkeley, quite unconfident and unsure of myself. I am indebted to Luca for being a very patient and inspiring mentor, for introducing me to a lot of beautiful theory while always giving me my own space (and time), and for innumerable espressos at Brewed Awakenings; all of which were instrumental in giving me the confidence to pursue my own research ideas.

Besides being an encouraging mentor, Luca has also been a great friend and a wonderful source of advice about academia, the theory community and Asian food. I benefitted greatly from his unique perspective on theory, and also research in general. I hope I have acquired at least some bit of his ability for "not understanding" seemingly simple things (and asking very interesting questions in the process).

I thank the theory faculty at Berkeley for the many discussions, seminars, courses and theory lunches that I was fortunate to attend. I was greatly inspired by the gentle personality of Dick Karp and the extremely enthusiatic one of Christos Papadimitriou, and the beautiful questions asked by both of them. I thank Satish Rao for many conversations about metric embeddings and politics, Alistair Sinclair and David Aldous for removing my initial fears of probability theory and Umesh Vazirani for his research koans and various discussions on semidefinite programming and complexity theory. Also, many thanks to David, Satish and Alistair for serving on my quals and thesis committees and for kindly accommodating all my requests.

I worked with many excellent collaborators over the past few years. I thank: Sanjeev Arora, Anindya De, Costis Georgiou, Subhash Khot, Alexandra Kolla, Avner Magen, Omer Reingold, Grant Schoenebeck, David Steurer, Luca, Salil Vadhan and Nisheeth Vishnoi; for the things I learnt from each one of them.

I would not have even half of my fond memories of Berkeley without my fellow students. Thanks to Grant Schoenebeck, for always being a great companion, be it explorations of the Lovász-Schrijver hierarchy or of restaurants in Beijing. Thanks to Costis Daskalakis, Omid Etesami, Henry Lin, Alex Fabrikant, Alexandra Kolla, Mani Narayanan and Lorenzo Orecchia for everything they taught me in my initial years at Berkeley. Thanks to Per Austrin for many discussions about CSPs during his stay at Berkeley, and to Anindya De for many enjoyable conversations. Thanks to my other fellow students: Anand Bhaskar, Siu-Man Chan, Siu-On Chan, James Cook, Rafael Frongillo, Omar Khan, Anupam Prakash, Jonah Sherman, Yaron Singer, Isabelle Stanton, Alexandre Stauffer, Greg Valiant, Thomas Vidick and Thomas Watson; for the many group retreats, reading groups, TGIFs and whiteboard battles in general.

My journey through graduate school was greatly aided by the support of many other friends. I thank my house-mates Debopriyo Chowdhury and Arkaprabha Sengupta for always being very helpful and tolerant of my idiosyncrasies. Thanks to Pulkit Grover, Subhranshu Maji, Vinayak Nagpal, Koushik Pal, Nikhil Shetty and Ambuj Tewari for all the fun moments I shared with them.

As a graduate student, I spent a semester at Princeton University, one month at Tsinghua University and two summers at Microsoft Research India. I thank my hosts at each one of these places for their kind hospitality. Thanks to Sushobhan Avasthi, Abhishek Bhattacharjee, Eugene Brevdo, Eden

Chlamtac, Shougata Ghosh, Parikshit Gopalan, Moritz Hardt, Aman Jain, Rajsekar Manokaran, Prasad Raghavendra and David Steurer; whose company I enjoyed while at Princeton.

It would not have been possible for me to embark on this journey without the unwavering support of family, who instilled in me a love for independent and original thought (and did not hang up the phone whenever I tried explaining what is it that I *actually* do). To them, with love, I dedicate this thesis.

To my family

Contents

List of Figures

Chapter 1

Introduction

Combinatorial optimization problems play an important role in computer science. However, most such problems like MAX 3-SAT, MAX-CUT, Minimum Vertex Cover etc. are NP-complete and a common way to deal with it is to settle for an approximately optimal solution.

The approximation algorithms developed for these problems have a special structure. Most approximation algorithms are developed by formulating the problem at hand as an integer program. One then relaxes the integer program to a convex program which can be solved in polynomial time: such as a program with linear constraints over real variables (called a linear program or an LP) or a program with linear matrix inequalities over positive semidefinite matrices (called a semidefinite program or an SDP). One is then left with the task of designing an algorithm to convert the solution of such a convex relaxation, to an integer solution for the combinatorial problem, often referred to as "rounding".

If we are dealing with (say) a maximization problem for which the true combinatorial optimum is OPT, then the convex relaxation will achieve a value FRAC which is at least as large as OPT (as the integer solution is also a feasible solution to the convex program). The rounding algorithm then uses the solution of the convex relaxation with objective value FRAC to produce an integer solution with (possibly suboptimal) value ROUND. The analysis of the algorithm then boils down to a comparison of these three quantities which satisfy $\mathsf{ROUND} \leq \mathsf{OPT} \leq \mathsf{FRAC}$. The inequalities are reversed for a minimization problem.

If one just thinks of the combinatorial problem as a question of finding the optimum *value* of the objective (e.g. the *size* of the minimum vertex cover in a graph), then the rounding algorithm is not needed and the quantities of interest are the values OPT and FRAC. If instead, the question is to *search* for an optimum integer solution (e.g. a minimum vertex cover), then one is interested in comparing the quantities ROUND and OPT. However, in the analysis of the problem, it is often not possible to compare the solution against the value OPT, simply because OPT is not known for an arbitrary instance of the problem! Hence, the approximation guarantees for algorithms are given by giving an upper bound on the ratio $\mathsf{FRAC}/\mathsf{ROUND}$, which in turn is an upper bound on the ratio $\mathsf{OPT}/\mathsf{ROUND}$. To establish a lower bound for both these notions, we need to establish a lower bound on the ratio $\mathsf{FRAC}/\mathsf{OPT}$ (for an explicit instance where we know OPT), which is called the *integrality gap* of the program. Figure 1.1 shows the relationship between these quantities.

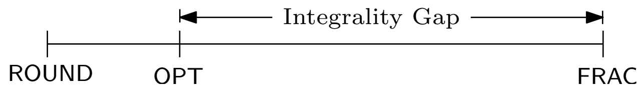

Figure 1.1: The integrality gap

Definition 1.1 *Let $\mathcal{I}_\Pi$ denote an arbitrary instance of a maximization problem Π formulated as an integer program. Then for a given convex relaxation of the program, let* $\mathsf{OPT}(\mathcal{I}_\Pi)$ *denote the optimum of the integer program and let* $\mathsf{FRAC}(\mathcal{I}_\Pi)$ *denote the optimum of the convex relaxation. Then the* integrality gap *of the relaxation is defined as*

$$\textit{Integrality Gap} \quad = \quad \sup_{\mathcal{I}_\Pi} \frac{\mathsf{FRAC}(\mathcal{I}_\Pi)}{\mathsf{OPT}(\mathcal{I}_\Pi)}$$

For a minimization problem, the integrality gap is defined to be the supremum of the inverse ratio.

Note that according to the above definition, the integrality gap is always at least 1 and a large gap indicates a poor approximation ratio. In cases when the integrality gap is infinite, we express it as a function of the size of the instance $\mathcal{I}_\Pi$.

1.1 Programs with Local Constraints

The convex relaxations for most problems have, in fact, another feature in common. Most relaxations only impose *local constraints* on the variables i.e. each constraint only affects a few variables. Consider for example the following linear and semidefinite[1] relaxations for the problem of finding the Maximum Independent Set (the largest subset of vertices not containing any edges) in a graph $G = (V, E)$. Constraints in both programs correspond to at most two vertices of the graph.

LP relaxation		SDP relaxation	
maximize	$\sum_{i \in V} x_i$	maximize	$\sum_{i \in V} \lvert\mathbf{u}_i\rvert^2$
subject to	$x_i + x_j \leq 1 \quad \forall (i,j) \in E$	subject to	$\left\langle \mathbf{u}_i, \mathbf{u}_j \right\rangle = 0 \quad \forall (i,j) \in E$
	$x_i \in [0,1]$		$\langle \mathbf{u}_i, \mathbf{u}_0 \rangle = \lvert\mathbf{u}_i\rvert^2 \quad \forall i \in V$
			$\lvert\mathbf{u}_0\rvert = 1$

Figure 1.2: LP and SDP relaxations for Maximum Independent Set

The strengthenings of these relaxations that have been considered before, involve adding constraints on a larger number of variables (but the number is still small compared to the size of the problem instance). This process of generating stronger relaxations by adding larger (but still, local) constraints is captured by various hierarchies of convex relaxations such as the ones defined by

[1]The program here is written in terms of vectors, but the constraints can be verified to be linear matrix inequalities on the positive semidefinite matrix X with $X_{ij} = \left\langle \mathbf{u}_i, \mathbf{u}_j \right\rangle$.

Lovász and Schrijver [LS91], Sherali and Adams [SA90] and Lasserre [Las01]. These hierarchies define various *levels* of convex relaxations for a problem, with the relaxations at a higher level being more powerful than the ones at lower levels.

These hierarchies are known to capture the best available algorithms for many problems, such as the SDP for Sparsest Cut by Arora, Rao and Vazirani [ARV04] and the ϑ-function of Lovász for Maximum Independent Set [Lov79], within a constant number of levels. It is also known that for an integer program with n variables taking values in $\{0, 1\}$, the program obtained by n levels of any of the above hierarchies has integrality gap 1 i.e. it gives the exact solution. However, solving the program obtained by t levels of these hierarchies takes time $n^{O(t)}$ which is exponential in n for $t = \Omega(n/\log n)$.

A lower bound for a convex program obtained after many levels (say $\Omega(n)$) of such a hierarchy, is then a strong lower bound against a class of algorithms capturing most known ones. Such a lower bound is proved by showing that the integrality gap of the program obtained after many levels of the hierarchy remains large. Note that such a lower bound is *unconditional* (it does not assume $\mathsf{P} \neq \mathsf{NP}$!) and rules out *exponential time algorithms in a powerful algorithmic model*.

Another motivation for studying lower bounds for these hierarchies is to study the *effect of local constraints in globally constraining the optimum* of a combinatorial problem. Levels in these hierarchies provide a natural notion of increasingly powerful local constraints. A lower bound, then involves study of problem instances whose local properties are very different from their global properties, which are interesting objects in their own right.

We describe below each of these hierarchies. We shall use the example of Maximum Independent Set throughout this chapter to illustrate the differences in the programs obtained by the various hierarchies. An excellent comparison of all the three hierarchies mentioned above is also available in [Lau03].

1.2 The Lovász - Schrijver Hierarchies

Lovász and Schrijver [LS91] describe two versions of a “lift-and-project” method. This can be thought of as an operator which when applied to a convex programming relaxation K of a 0/1 integer linear program, produces a tighter relaxation. A weaker version of the method, denoted LS, adds auxiliary variables and linear inequalities, and the projection of the new relaxation on the original variables is denoted by $N(K)$; a stronger version, denoted LS+, adds semidefinite programming constraints as well, and the projection on the original variables is denoted by $N_+(K)$.

Starting from a basic relaxation and iteratively applying the operator N (N_+) one gets higher and higher levels (which are called *rounds* for the Lovász-Schrijver hierarchies due to their iterative nature) of the LS (LS+) hierarchy. Thus, the relaxation obtained by r rounds of the hierarchy is given by $N(\cdots N(K)\cdots)$ where the operator is applied t times. We denote it as $N^t(K)$.

Lovász and Schrijver also prove that if we start from a linear programming relaxation of a 0/1 integer program with n variables, then n applications of the LS procedures are sufficient to obtain a tight relaxation where the only feasible solutions are convex combinations of integral solutions. If one starts from a linear program with poly(n) inequalities, then it is possible to optimize over the set of solutions defined by t rounds of LS or LS+ in $O(n^{O(t)})$ time.[2]

1.2.1 Definitions of the hierarchies

To describe these hierarchies it will be more convenient to work with *convex cones* rather than arbitrary convex subsets of $[0,1]^n$. Recall that a *cone* is a subset K of $\mathbb{R}^d$ such that if $\mathbf{x}, \mathbf{y} \in K$ and $\alpha, \beta \geq 0$ then $\alpha\mathbf{x} + \beta\mathbf{y} \in K$, that is, a cone is a set of vectors that is closed under non-negative linear combinations. (Note that, in particular, a cone is always convex.)

If we are interested in a convex set $R \subseteq [0,1]^n$ (which might be the feasible region of our starting convex relaxation), we first convert it into the cone $K \subseteq \mathbb{R}^{n+1}$ defined as the set of all vectors $(\lambda, \lambda y_1, \ldots, \lambda y_N)$ such that $\lambda \geq 0$ and $(y_1, \ldots, y_n) \in R$. For example, in the "cone" linear programming relaxation of the Maximum Independent Set problem $(y_0, y_1, \ldots, y_n)$ is in the feasible region (denoted by $IS(G)$) if and only if

$$
\begin{aligned}
y_i + y_i &\leq y_0 \quad \forall (i,j) \in E \\
0 \leq y_i &\leq y_0 \quad \forall i \in V \\
y_0 &\geq 0 \qquad\qquad (IS(G))
\end{aligned}
$$

We would now like to "tighten" the relaxation by adding inequalities (on the solution obtained after scaling to get $y_0 = 1$) that are valid for 0/1 solutions but that are violated by other solutions. Ideally, we would like to say that a solution $(1, y_1, \ldots, y_n)$ must satisfy the conditions $y_i^2 = y_i$, because such a condition is satisfied only by 0/1 solutions. Equivalently, we could introduce n^2 new variables $Y_{i,j}$ and add the conditions (i) $Y_{i,j} = y_i \cdot y_j$ and (ii) $Y_{i,i} = y_i$. Unfortunately, condition (i) is neither linear nor convex, and so we will instead "approximate" condition (i) by enforcing a set of linear conditions that are implied by (but not equivalent to) (i). This is formalized in the definition below.

Definition 1.2 *For a cone $K \subseteq \mathbb{R}^d$ we define the set $N(K)$ (also be a cone in $\mathbb{R}^d$) as follows: a vector $\mathbf{y} = (y_0, \ldots, y_{d-1}) \in \mathbb{R}^d$ is in $N(K)$ if and only if there is a matrix $Y \in \mathbb{R}^{d \times d}$ such that*

1. *Y is symmetric;*
2. *For every $i \in \{0, 1, \ldots, d-1\}$, $Y_{0,i} = Y_{i,i} = y_i$*
3. *Each row Y_i is an element of K*
4. *Each vector $Y_0 - Y_i$ is an element of K*

In such a case, Y is called the protection matrix *of* $\mathbf{y}$. *If, in addition, Y is* positive semidefinite, *then $\mathbf{y} \in N_+(K)$. We define $N^0(K)$ and $N_+^0(K)$ as K, and $N^t(K)$ (respectively, $N_+^t(K)$) as $N(N^{t-1}(K))$ (respectively, $N_+(N_+^{t-1}(K))$).*

[2]It is also possible to optimize over feasible solutions for $N^t(K)$ and $N_+^t(K)$ in time $n^{O(t)}$ provided that a separation oracle for K is computable in time poly(n). (That is, it is not necessary for K to be a linear or semidefinite programming relaxation with a polynomial number of inequalities.)

If $\mathbf{y} = (1, y_1, \ldots, y_{d-1}) \in \{0,1\}^d$, then we can set $Y_{i,j} = y_i \cdot y_j$. Such a matrix Y is clearly positive semidefinite, and it satisfies $Y_{i.i} = y_i^2 = y_i$ if the y_i are in $\{0,1\}$. Consider now a row Y_i of Y, that is, the vector $\mathbf{r}$ such that $r_j := Y_{i,j} = y_i \cdot y_j$. Then, either $y_i = 0$, in which case $\mathbf{r} = (0, \ldots, 0)$ is in every cone, or $y_i = 1$, and $\mathbf{r} = \mathbf{y}$. Similarly, if we consider $r_j := Y_{0,j} - Y_{i,j} = (1 - y_i) \cdot y_j$ we find that it either equals the all-zero vector or it equals $\mathbf{y}$. This shows that if $\mathbf{y} = (1, y_1, \ldots, y_{d-1}) \in \{0,1\}^d$ and $\mathbf{y} \in K$, then also $\mathbf{y} \in N_+^t(K)$ for every t. Hence, if $K \cap \{y_0 = 1\}$ defines a relaxation of the integral problem, so does $N_+^t(K) \cap \{y_0 = 1\}$, and hence also $N^t(K) \cap \{y_0 = 1\}$.

For a graph G, the relaxation of the Maximum Independent Set problem resulting from t rounds of LS+ is the result of

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^{n} y_i \\ \text{subject to} & (y_0, \ldots, y_n) \in N_+^t(IS(G)) \\ & y_0 = 1 \end{array}$$

1.2.2 The Prover-Adversary Game

It is sometimes convenient to think of the LS and LS+ hierarchies in terms of a prover-adversary game. This formulation was first used by Buresh-Oppenheim et. al. [BOGH+03] who used it to prove lower bounds on the LS+ procedure as a proof system. For example, the following is a formulation of which is convenient to use for proving that a certain vector belongs to $N_+^t(IS(G))$. The treatment for relaxations of other problems is identical. Also, the formulation remains the same for talking about $N(IS(G))$, except that we omit the positive semidefiniteness constraint in the matrix below.

A *prover* $\mathcal{P}$ is an algorithm that, on an input vector $(y_0, \ldots, y_n)$, either fails or outputs a matrix $Y \in \mathbb{R}^{(n+1)\times(n+1)}$ and a set of vectors $O \subseteq \mathbb{R}^{n+1}$ such that

1. Y is symmetric and positive semidefinite.
2. $Y_{i,i} = Y_{0,i} = y_i$.
3. Each vector Y_i and $Y_0 - Y_i$ is a non-negative linear combination of vectors of O.
4. Each element of O is in $IS(G)$.

Consider now the following game played by a prover against another party called the *adversary*. We start from a vector $\mathbf{y} = (y_0, \ldots, y_n)$, and the prover, on input $\mathbf{y}$, outputs Y and O as before. Then the adversary chooses a vector $\mathbf{z} \in O$, and the prover, on input $\mathbf{z}$, outputs a matrix Y' and a set O', and so on. The adversary *wins* when the prover fails.

Lemma 1.3 *Suppose that there is a prover such that, starting from a vector $\mathbf{y} \in IS(G)$, every adversary strategy requires at least $t + 1$ moves to win. Then $\mathbf{y} \in N^t(IS(G))$.*

Proof: We proceed by induction on t, with $t = 0$ being the simple base case. Suppose that, for every adversary, it takes at least $t + 1$ moves to win, and let Y and O be the output of the prover on input $\mathbf{y}$. Then, for every element $\mathbf{z} \in O$, and every prover strategy, it takes at least t moves to win starting from $\mathbf{z}$. By inductive, hypothesis, each element of O is in $N_+^{t-1}(IS(G))$, and since $N_+^{t-1}(IS(G))$ is closed under non-negative linear combinations, the vectors Y_i and $Y_0 - Y_i$ are all in $N_+^{t-1}(IS(G))$, and so Y is a protection matrix that shows that $\mathbf{y}$ is in $N_+^t(IS(G))$, ■

1.3 The Sherali-Adams Hierarchy

The Sherali-Adams hierarchy [SA90] defines a hierarchy of linear programs which give increasingly tighter relaxations. To see the intuition behind the hierarchy, we can see it as an extension of the LS procedure. Recall that the solution to a 0/1 integer program can be specified by a vector $\mathbf{y} \in \{0,1\}^n$. In the Lovász-Schrijver hierarchy we defined auxiliary variables Y_{ij} and wanted to express the constraint that $Y_{ij} = y_i \cdot y_j$. We then expressed it by some implied linear conditions on the variables Y_{ij}.

Consider a solution $(1, y_1, \ldots, y_n)$ which is feasible at the second level of the LS hierarchy. Then the row Y_i of the protection matrix must also define a feasible solution to the "cone" version of the relaxation, say $IS(G)$. Since, $Y_{i0} = 1$ the solution $\mathbf{y}' = (1, Y_{i1}/y_i, \ldots, Y_{in}/y_i)$ must also be feasible (assuming $y_i \neq 0$), for the first level, there exists a protection matrix Y' for it. Now, we would also like to think of $Y'_{jk} = (Y_{ij}/y_i)(Y_{jk}/y_i) = y_i y_j y_k$. However, notice that the choice of Y' was dependent on the fact that we chose the row Y_i. In particular, if we looked at the protection matrix Y'' for the solution $\mathbf{y}'' = (1, Y_{j0}/y_j, \ldots, Y_{jn}/y_j)$, it need not be true that $Y'_{jk} = Y''_{ik}$.

The Sherali-Adams hierarchy solves this problem by introducing all the auxiliary variables at once instead of an inductive process. In particular, we define a variable Y_S for each $S \subseteq [n]$ with $|S| \leq t+1$. The intuition again is that we want to impose $Y_S = \prod_{i \in S} y_i$. However, we instead impose some linear conditions implied by this. For every constraint $a^\mathsf{T}\mathbf{y} - b \leq 0$ of the starting LP relaxation, we consider sets S, T such that $|S| + |T| \leq t$ and impose a linear implication of $(a^\mathsf{T}\mathbf{y} - b) \cdot \prod_{i \in S} y_i \cdot \prod_{j \in T}(1 - y_j) \leq 0$, by requiring that

$$\sum_{T' \subseteq T} (-1)^{|T'|} \cdot \left(\sum_{i=1}^{n} a_i \cdot Y_{S \cup T' \cup \{i\}} - b \cdot Y_{S \cup T'} \right) \leq 0$$

Note again that the number of variables and constraints in the LP at level t is $n^{O(t)}$ and hence it can be solved in time $n^{O(t)}$. Also, each such program is a relaxation, since for any $\mathbf{y} \in \{0,1\}^n$ satisfying the initial constraints, $Y_S = \prod_{i \in S} y_i$ defines a valid level-t solution. The program below gives the relaxation of Maximum Independent Set obtained at the t^{th} level of the Sherali-Adams hierarchy.

Since the above program is a convex relaxation, any convex combination of 0/1 solutions is also a solution to the program. It is convenient to think of the convex combination as defining a *distribution* over 0/1 solutions. With this interpretation, we can think of Y_S as the *probability* that all variables in set S are equal to 1. The following lemma says that a solution which "locally" looks like a valid distribution, is a feasible solution for the above relaxation.

Lemma 1.4 *Consider a family of distributions $\{\mathcal{D}(S)\}_{S \subseteq [n]:|S| \leq t+2}$, where each $\mathcal{D}(S)$ is defined over $\{0,1\}^S$. If the distributions satisfy*

$$
\begin{array}{lll}
\text{maximize} & \displaystyle\sum_{i=1}^{n} Y_{\{i\}} & \\
\text{subject to} & \displaystyle\sum_{T'\subseteq T} (-1)^{|T'|}\cdot\left[Y_{S\cup T'\cup\{j\}} + Y_{S\cup T'\cup\{i\}} - Y_{S\cup T'}\right] \ \le\ 0 & |S|+|T|\le t,\ (i,j)\in E \\
& 0\ \le\ \displaystyle\sum_{T'\subseteq T} (-1)^{|T'|}\cdot Y_{S\cup T'\cup\{i\}}\ \le\ \sum_{T'\subseteq T}(-1)^{|T'|}\cdot Y_{S\cup T'} & |S|+|T|\le t \\
& Y_{\emptyset}\ =\ 1 &
\end{array}
$$

Figure 1.3: Sherali-Adams relaxation for Maximum Independent Set

1. *For all* $(i, j) \in E$ *and* $S \supseteq \{i, j\}$, $\mathbb{P}_{\mathcal{D}(S)}[(y_i = 1) \wedge (y_j = 1)] = 0$, *and*
2. $S' \subseteq S \subseteq [n]$ *with* $|S| \le t + 1$, *the distributions* $\mathcal{D}(S'), \mathcal{D}(S)$ *agree on* S'

then $Y_S = \mathbb{P}_{\mathcal{D}(S)}\left[\bigwedge_{i\in S}(y_i = 1)\right]$ *is a feasible solution for the above level-t Sherali-Adams relaxation.*

Proof: We have $Y_\emptyset = 1$ by definition. We first verify the second constraint. For given S, T, i, let W denote $S \cup T \cup \{i\}$. Using the fact that distributions are consistent over subsets, we can write

$$
\begin{aligned}
\sum_{T'\subseteq T}(-1)^{|T'|}\cdot Y_{S\cup T'\cup\{i\}} &= \sum_{T'\subseteq T}(-1)^{|T'|}\cdot \mathop{\mathbb{P}}_{\mathcal{D}(S\cup T'\cup\{i\})}\left[\bigwedge_{j\in S\cup T'\cup\{i\}}(y_j=1)\right] \\
&= \sum_{T'\subseteq T}(-1)^{|T'|}\cdot \mathop{\mathbb{P}}_{\mathcal{D}(W)}\left[\bigwedge_{j\in S\cup T'\cup\{i\}}(y_j=1)\right] \\
&= \mathop{\mathbb{P}}_{\mathcal{D}(W)}\left[\bigwedge_{j\in S\cup\{i\}}(y_j=1)\bigwedge_{k\in T}(y_k=0)\right]
\end{aligned}
$$

where the last equality follows from inclusion-exclusion. With the manipulation on the RHS, the second constraint becomes

$$
0\ \le\ \mathop{\mathbb{P}}_{\mathcal{D}(W)}\left[\bigwedge_{j\in S\cup\{i\}}(y_j=1)\bigwedge_{k\in T}(y_k=0)\right]\ \le\ \mathop{\mathbb{P}}_{\mathcal{D}(W)}\left[\bigwedge_{j\in S}(y_j=1)\bigwedge_{k\in T}(y_k=0)\right]
$$

which is obviously satisfied since $\mathcal{D}(W)$ is a distribution. Similarly, for $W_1 = S \cup T \cup \{i, j\}$, we can transform the first constraint to (probabilities below are taken with respect to $\mathcal{D}(W_1)$)

$$
\mathbb{P}\left[\bigwedge_{k\in S\cup\{i\}}(y_k=1)\bigwedge_{l\in T}(y_l=0)\right]+\mathbb{P}\left[\bigwedge_{k\in S\cup\{j\}}(y_k=1)\bigwedge_{l\in T}(y_l=0)\right]\le\mathbb{P}\left[\bigwedge_{k\in S}(y_k=1)\bigwedge_{l\in T}(y_l=0)\right]
$$

which is satisfied since the assumptions of the lemma imply that the events $\left\{\bigwedge_{k\in S\cup\{i\}}(y_k = 1)\bigwedge_{l\in T}(y_l = 0)\right\}$ and $\left\{\bigwedge_{k\in S\cup\{j\}}(y_k = 1)\bigwedge_{l\in T}(y_l = 0)\right\}$ are disjoint. ■

1.4 The Lasserre Hierarchy

The Lasserre hierarchy gives a sequence of increasingly tight semidefinite programming relaxations for a quadratic integer program for variables taking values 0 and 1. As in the case of the Sherali-Adams hierarchy, the semidefinite program after t rounds of the Lasserre hierarchy also introduces a new (vector valued) variable for the product of every t variables in the original program.

For concreteness, we consider the program for Maximum Independent Set. The same procedure can be used to derive the level-t SDP for any problem formulated as a quadratic integer program, with variables taking values in $\{0, 1\}$. Given a graph $G = (V, E)$, the integer program would have a variable X_i for each $i \in V$ with $y_i = 1$ if i is in the independent set and 0 otherwise. To ensure that the solution is an independent set, we would enforce that $y_i \cdot y_j = 0$ for all $(i, j) \in E$.

To obtain the Lasserre relaxation, we first think of a an integer program which has a variable Y_S for each $S \subseteq V, |S| \leq t$ where the intended solution, as before, is $Y_S = 1$ iff all vertices in S are in the independent set. We can then add the constraint that the product $Y_{S_1} \cdot Y_{S_2}$ must only depend on $S_1 \cup S_2$. For homogenization, we introduce an extra variable $Y_\emptyset$ which is always supposed to be 1. Replacing the integer variables Y_S by vectors $\mathbf{U}_S$ gives the semidefinite relaxation as below.

$$
\begin{array}{lll}
\text{maximize} & \sum_{i \in V} \left|\mathbf{U}_{\{i\}}\right|^2 & \\
\text{subject to} & \left\langle \mathbf{U}_{\{i\}}, \mathbf{U}_{\{j\}} \right\rangle = 0 & \forall\, (i, j) \in E \\
& \langle \mathbf{U}_{S_1}, \mathbf{U}_{S_2} \rangle = \langle \mathbf{U}_{S_3}, \mathbf{U}_{S_4} \rangle & \forall\, S_1 \cup S_2 = S_3 \cup S_4 \\
& \langle \mathbf{U}_{S_1}, \mathbf{U}_{S_2} \rangle \in [0, 1] & \forall S_1, S_2 \\
& |\mathbf{U}_\emptyset| = 1 &
\end{array}
$$

Figure 1.4: Lasserre SDP for Maximum Independent Set

Note that the program for level t only has vectors for sets of size at most t. It can be shown that for any set S with $|S| \leq t$, the vectors $\mathbf{U}_{S'}, S' \subseteq S$ induce a probability distribution over valid independent sets of the subgraph induced by S. However, unlike the Sherali-Adams hierarchy, the existence of such distributions is not a sufficient condition for the existence of a feasible solution for the semidefinite program.

1.5 A comparison

Let $SA^{(t)}(P)$ denote the feasible set of the program obtained by starting from a basic 0/1 relaxation P and augmenting variables for t levels of the Sherali-Adams hierarchy. Similarly, let $LS^{(t)}(P)$, $LS_+^{(t)}(P)$, $Las^{(t)}(P)$ represent feasible sets corresponding respectively to t levels of the LS, LS+ and Lasserre hierarchies. We summarize in the facts below, a comparison of these relaxations. The reader is referred to the excellent survey by Laurent [Lau03] for a more detailed comparison.

1. $LS^{(n)}(P) = LS_+^{(n)}(P) = SA^{(n)}(P) = Las^{(n)}(P) = \mathcal{I}$, where $\mathcal{I}$ denotes the convex hull of the 0/1 solutions to the starting integer program with n variables.

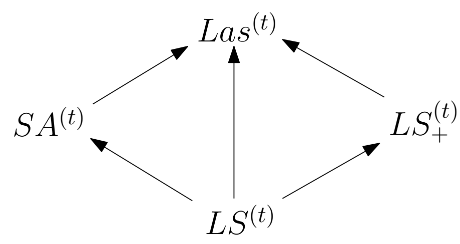

Figure 1.5: A comparison of the hierarchies

2. For all $t \leq n$, $LS^{(t)}(P) \subseteq LS_+^{(t)}(P) \subseteq Las^{(t)}(P)$, and also $LS^{(t)}(P) \subseteq SA^{(t)}(P) \subseteq Las^{(t)}(P)$. Hence, the relaxations provided by the Lasserre hierarchy at each level are the strongest (most constrained) among the relaxations at the corresponding level of all the hierarchies discussed above.

3. If the starting relaxation P has $n^{O(1)}$ constraints, then one can optimize over the sets $LS^{(t)}(P), LS_+^{(t)}(P), SA^{(t)}(P)$ and $Las^{(t)}(P)$ in time $n^{O(t)}$. This is known to be true for $LS^{(t)}(P), LS_+^{(t)}(P)$ and $SA^{(t)}(P)$ even if we only assume that P has a weak separation oracle running in time $n^{O(1)}$. It is not known if one can optimize efficiently over $Las^{(t)}(P)$ using an efficient separation oracle for P.

1.6 Results in this thesis

This thesis primarily contains lower bounds for various hierarchies. Chapters 2 and 3 investigate the problem of approximating Minimum Vertex Cover using the LS and LS+ hierarchies. Chapter 2 is based on [STT07a] and shows that $\Omega(n)$ levels of the LS+ hierarchy do not provide an approximation factor better than $7/6$. Chapter 3 contains an optimal lower bound (factor 2) for $\Omega(n)$ levels of the (weaker) LS hierarchy and is based on results in [STT07b].

Chapters 4 and 5 consider various constraint satisfaction problems. Chapter 4, based on work in [Tul09], provides optimal lower bounds for approximating the number of satisfiable constraints, for a general class of constraint satisfaction problems. Chapter 5, drawing on work in [GMT09] further extends this class and proves optimal lower bounds in the Sherali-Adams hierarchy.

In Chapter 6, we explore the topic of "reductions" and whether lower bounds for one problem can be translated to lower bounds for other problems as in the case of NP-hardness. Using the results from Chapter 4, we obtain lower bounds for Maximum Independent Set, Minimum Vertex Cover and graph coloring problems, in the Lasserre hierarchy. This is based on work in [Tul09].

Chapter 7 uses SDP relaxations to get algorithms for special cases of Unique Games, which is an important constraint satisfaction problem in complexity theory. These results appear in [AKK+08].

Chapter 2

Integrality Gaps for Lovász-Schrijver SDP relaxations

In this chapter we will prove an integrality gap for the Lovász-Schrijver SDP relaxations of Minimum Vertex Cover. Recall that the problem requires one to find a minimum subset S of vertices in a graph $G = (V, E)$, such that S contains at least one vertex from every edge of the graph. We will prove that for graphs on n vertices, the integrality gap may be as large as $7/6 - \varepsilon$ for the relaxation obtained by $\Omega_\varepsilon(n)$ rounds[1] of the LS+ hierarchy.

This result is actually subsumed by a later result of Schoenebeck [Sch08], who showed a gap of $7/6 - \varepsilon$, even after $\Omega_\varepsilon(n)$ levels of the *Lasserre* hierarchy. We choose to present it here, primarily because it provides a nice illustration of the proof techniques of inductive nature, used for reasoning about the Lovász-Schrijver hierarchy. The arguments in this chapter are technically much simpler than the ones in Chapter 3, where we present strong results for Minimum Vertex Cover in the LS hierarchy (the analogues of which are not known in the other hierarchies).

Previous work

The study of Lovász-Schrijver relaxations of Vertex Cover was initiated by Arora, Bollobás, Lovász, and Tourlakis [ABL02, ABLT06, Tou06] who lower bounds for the LS hierarchy (of linear programs). They showed that even after $\Omega_\varepsilon(\log n)$ rounds the integrality gap is at least $2 - \varepsilon$, and that even after $\Omega_\varepsilon((\log n)^2)$ rounds the integrality gap is at least $1.5 - \varepsilon$.

Buresh-Oppenheim, Galesy, Hoory, Magen and Pitassi [BOGH$^+$03], and Alekhnovich, Arora, Tourlakis [AAT05] studied the LS+ hierarchy and proved $\Omega(n)$ LS+ round lower bounds for proving the unsatisfiability of random instances of 3SAT (and, in general, kSAT with $k \geq 3$) and $\Omega_\varepsilon(n)$ round lower bounds for achieving approximation factors better than $7/8 - \varepsilon$ for MAX 3-SAT, better than $(1 - \varepsilon) \ln n$ for Minimum Set Cover, and better than $k - 1 - \varepsilon$ for Hypergraph Vertex Cover in k-uniform hypergraphs. They left open the question of proving LS+ round lower bounds for approximating the Vertex Cover problem.

[1] Due to the iterative nature of its definition, the levels of Lovász-Schrijver hierarchy are often referred to as "rounds".

The standard reduction from Max 3SAT to Vertex Cover shows that if one is able to approximate Vertex Cover within a factor better than $17/16$ then one can approximate MAX 3-SAT within a factor better than $7/8$. This fact, and the $7/8-\varepsilon$ integrality gap for MAX 3-SAT of [AAT05], however do not suffice to derive an integrality gap result for Vertex Cover. The reason is that reducing an instance of MAX 3-SAT to a graph, and then applying a Vertex Cover relaxation to the graph, defines a semidefinite program that is possibly tighter than the one obtained by a direct relaxation of the MAX 3-SAT problem.

Feige and Ofek [FO06] were able to analyze the value of the Lovász Theta function of the graph obtained by taking a random 3SAT instance and then reducing it to an instance of Independent Set (or, equivalently, of Vertex Cover). Their result immediately implies a $17/16 - \varepsilon$ integrality gap for one round of LS+, and the way in which they prove their result implies also the stronger $7/6 - \varepsilon$ bound. For one round of LS+ (or, equivalently, for the function defined as number of vertices minus the Theta function) Goemans and Kleinberg [KG98] had earlier proved a $2 - o(1)$ integrality gap result by using a different family of graphs. Charikar [Cha02] proves a $2-o(1)$ integrality gap result for a semidefinite programming relaxation of Vertex Cover that includes additional inequalities. Charikar's relaxation is no tighter than 3 rounds of LS+, and is incomparable with the relaxation obtained after two rounds.

It was compatible with previous results that after a constant number of rounds of LS+ or after poly $\log n$ rounds of LS the integrality gap for Vertex Cover could become $1 + o(1)$.

Our Result

We prove that after $\Omega_\varepsilon(n)$ rounds of LS+ the integrality gap remains at least $7/6 - \varepsilon$. (For a stronger reason, the lower bound applies to LS as well.)

We combine ideas from the work of Alekhnovich, Arora and Tourlakis [AAT05] and Feige and Ofek [FO06]. As in [FO06], we study the instance obtained by starting from a random instance of 3XOR and then reducing it to the independent set problem; we also define our semidefinite programming solutions in a way that is similar to [FO06] (with the difference that we need to define such solutions inside an inductive argument, while only one solution is needed in [FO06]). As in [AAT05], our argument proceeds by considering an "expansion" property of the underlying instance of 3XOR and maintaining it as an invariant throughout the proof. Our way of modifying the instance at every step to ensure expansion (called "expansion correction" in previous works) is new.

Subsequent to this work, such a result was also shown for the more powerful Lasserre hierarchy, by Schoenebeck[Sch08]. A larger gap (1.36) was also exhibited for $\Omega(n^\delta)$ levels of the Lasserre hierarchy in a result presented in Chapter 6.

2.1 Preliminaries

Recall that the LS+ hierarchy is defined by considering the cone of feasible solutions to the (homogenized version of) starting LP and repeatedly applying the N_+ operator which applies certain constraints to the solutions in the cone. The cone of solutions for the vertex cover problem on a graph $G = (V, E)$ where $V = \{1, \ldots .N\}$ is

$$\begin{aligned} y_i + y_j &\geq y_0 \quad \forall (i, j) \in E \\ 0 \leq y_i &\leq y_0 \quad \forall i \in V \\ y_0 &\geq 0 \qquad\qquad\qquad (VC(G)) \end{aligned}$$

We shall be interested in optimizing over a smaller cone which is obtained by t applications of the N_+ operator to the above. In particular, we are interested in the following SDP

$$\begin{aligned} \text{Minimize} \quad & \sum_{i=1}^{N} y_i \\ \text{subject to:} \quad & (y_0, \ldots, y_N) \in N_+^t(VC(G)) \\ & y_0 = 1 \end{aligned}$$

Even though we are interested in the LS+ relaxations of vertex cover, it shall be more convenient to argue about the maximum independent set problem i.e. for the cone $IS(G)$ where the constraint $y_i + y_j \geq y_0$ is replaced by $y_i + y_j \leq y_0$. The following lemma shows that the two settings are equivalent.

Lemma 2.1 *Let $G = (V, E)$ be a graph and $V = 1, \ldots, N$. Then, for every $k \geq 0$, $(y_0, y_1, \ldots, y_N) \in N_+^k(VC(G))$ if and only if $(y_0, y_0 - y_1, \ldots, y_0 - y_N) \in N^k(IS(G))$.*

Proof: We prove it by induction, with the $k = 0$ base case being clear. If $(y_0, y_1, \ldots, y_N) \in N_+^{k+1}(VC(G))$ then there is a protection matrix Y that is symmetric, positive semidefinite, and such that $Y_{i,i} = Y_{0,i} = y_i$, and such that the vectors Y_i and $Y_0 - Y_i$ are in $N^k(VC(G))$. Since Y is positive semidefinite, there must be vectors $\mathbf{u}_0, \ldots, \mathbf{u}_N$ such that $Y_{i,j} = \mathbf{u}_i \cdot \mathbf{u}_j$.

Consider now the vectors $\mathbf{c}_0, \ldots, \mathbf{c}_N$ defined as follows: $\mathbf{c}_0 := \mathbf{u}_0$ and $\mathbf{c}_i := \mathbf{u}_0 - \mathbf{u}_i$ for $i > 0$. Define the matrix Z as $Z_{i,j} := \mathbf{c}_i \cdot \mathbf{c}_j$. Thus the matrix Z is symmetric and positive semidefinite. We will argue that Z is a protection matrix by showing that the vector $\mathbf{z} := (y_0, y_0 - y_1, \ldots, y_0 - y_N) \in N_+^{k+1}(IS(G))$.

First, we see that $Z_{0,0} = Y_{0,0} = y_0$ and that, for $i > 0$,

$$Z_{i,i} = (\mathbf{u}_0 - \mathbf{u}_i) \cdot (\mathbf{u}_0 - \mathbf{u}_i) = Y_{0,0} - 2Y_{0,i} + Y_{i,i} = y_0 - y_i$$

Consider now the row vector Z_i, which is equal to $(r_0, \ldots, r_N)$ where

$$r_0 = \mathbf{u}_0 \cdot (\mathbf{u}_0 - \mathbf{u}_i) = y_0 - y_i$$

and, for $j > 0$,

$$r_j = (\mathbf{u}_0 - \mathbf{u}_j) \cdot (\mathbf{u}_0 - \mathbf{u}_i) = y_0 - y_j - y_i + Y_{i,j}$$

We need to show $(r_0, \ldots, r_N) \in N_+^k(IS(G))$ which, by the inductive hypothesis, is equivalent to $(r_0, r_0 - r_1, \ldots, r_0 - r_N) \in N_+^k(VC(G))$. But $(r_0, r_0 - r_1, \ldots, r_0 - r_N) = Y_0 - Y_i$ which belongs to $N_+^k(VC(G))$ by our assumption that Y is a protection matrix for $\mathbf{y}$. The other conditions are similarly verified. ■

We recall the formulation membership problem in $N_+(IS(G))$ as a prover-adversary game which shall be useful in our argument. Recall that a *prover* $\mathcal{P}$ is an algorithm that, on an input vector $(y_0, \dots, y_N)$, either fails or outputs a matrix $Y \in \mathbb{R}^{(N+1)\times(N+1)}$ and a set of vectors $O \subseteq \mathbb{R}^{N+1}$ such that

1. Y is positive semidefinite
2. $Y_{i,i} = Y_{0,i} = y_i$
3. Each vector Y_i and $Y_0 - Y_i$ is a non-negative linear combination of vectors of O
4. Each element of O is in $IS(G)$

To prove that a vector $\mathbf{y} \in N_+(IS(G))$, the prover plays the following game against an *adversary*. On input $\mathbf{y}$, the prover outputs Y and O as before. Then the adversary chooses a vector $\mathbf{z} \in O$, and the prover, on input $\mathbf{z}$, outputs a matrix Y' and a set O', and so on. The adversary *wins* when the prover fails. The following lemma connects the game to the LS+ hierarchy.

Lemma 2.2 *Suppose that there is a prover such that, starting from a vector* $\mathbf{y} \in IS(G)$*, every adversary strategy requires at least* $t + 1$ *moves to win. Then* $\mathbf{y} \in N^t(IS(G))$.

2.2 Overview of Our Result

Let φ be an instance of 3XOR, that is, a collection of linear equations mod 2 over variables $x_1, \dots, x_n$ such that each equation is over exactly 3 variables. We denote by $\mathsf{OPT}(\varphi)$ the largest number of simultaneously satisfiable equations in φ.

Given a 3XOR instance φ with m equation, we define the FGLSS graph G_φ of φ as follows: G_φ has $4m$ vertices, one for each equation of φ and for each assignment to the three variables that satisfies the equation. We think of each vertex as being labeled by a partial assignment to three variables. Two vertices u and v are connected if and only if the partial assignments that label u and v are inconsistent. For example, for each equation, the four vertices corresponding to that equation form a clique. It is easy to see that $\mathsf{OPT}(\varphi)$ is precisely the size of the largest independent set of G_φ. Note that, in particular, the independent set size of G_φ is at most $N/4$, where $N = 4m$ is the number of vertices.

We say that φ is (k, c)-expanding if every set S of at most k equations in φ involves at least $c|S|$ distinct variables. Our main result is that if φ is highly expanding, then even after a large number of rounds of Lovász-Schrijver, the optimum of the relaxation is $N/4$, the largest possible value.

Lemma 2.3 (Main) *Let* φ *be a* $(k, 1.95)$*-expanding instance of 3XOR such that any two clauses share at most one variable, and let* G_φ *be its FGLSS graph.*

Then $(1, \frac{1}{4}, \dots, \frac{1}{4})$ *is in* $N_+^{(k-4)/44}(IS(G_\varphi))$.

Our integrality gap result follows from the well known fact that there are highly expanding instances of 3XOR where it is impossible to satisfy significantly more than half of the equations.

Lemma 2.4 *For every $c < 2$ and $\varepsilon > 0$ there are $\eta, \beta > 0$ such that for every n there is an instance φ of 3XOR with n variables and $m = \beta n$ equations such that*

- *No more than $(1/2 + \varepsilon)m$ equations are simultaneously satisfiable;*
- *Any two clauses share at most one variable*
- *φ is $(\eta n, c)$-expanding.*

The rest of the chapter is devoted to the proof of Lemma 2.3. We prove it in Section 2.3 by describing a prover strategy that survives for at least $(k-4)/44$ rounds. Proofs of variants of Lemma 2.4 have appeared before in the literature, for example in [BSW01, BOT02]; we give a proof of a more general version in the in the Appendix.

The two lemmas combine to give our lower bound.

Theorem 2.5 *For every $\varepsilon > 0$ there is a $c_\varepsilon > 0$ such that for infinitely many t there is a graph G with t vertices such that the ratio between the minimum vertex cover size of G and the optimum of $N^{c_\varepsilon t}(VC(G))$ is at least $7/6 - \varepsilon$.*

Proof: Using Lemma 2.4, construct an instance φ of 3XOR with n clauses and $O_\varepsilon(m)$ equations such that (i) no more than an $1/2 + \varepsilon$ fraction of equations can be simultaneously satisfied; (ii) any two clauses share at most one variable; and (iii) φ is $(\Omega_\varepsilon(n), 1.95)$-expanding.

The minimum vertex size in the graph G_φ is at least $4m - (1/2 + \varepsilon)m$, but, by Lemma 2.3, the solution $(1, 3/4, \cdots, 3/4)$ is feasible for $N^{\Omega_\varepsilon(n)}(VC(G_\varphi))$, and so the optimum of $N^{\Omega_\varepsilon(n)}(VC(G_\varphi))$ is at most $3m$. ■

2.3 The Prover Algorithm

For the sake of this section, we refer to a fixed formula φ with n variables $X = \{x_1, \ldots, x_n\}$ and m clauses which is $(k, 1.95)$-expanding and such that two clauses share at most one variable. The graph G_φ has $N = 4m$ vertices, which is also the number of variables in our starting linear program. Recall that each vertex i of G_φ corresponds to an equation C of φ and to an assignment of values to the three variables of C that satisfies C. (In the following, if i is one of the vertices corresponding to an equation C, we call C the *equation of i*.)

2.3.1 Some Intuition

Suppose that φ were satisfiable, and let D be a distribution over satisfying assignments for φ. Then define the vector $\mathbf{y} = \mathbf{y}(D)$ as follows: $y_0 = 1$ and

$$y_i := \mathop{\mathbb{P}}_{a \in D}[a \text{ agrees with } i]$$

We claim that this solution is in $N_+^k(IS(G))$ for all k. This follows from the fact that it is a convex combination of 0/1 solutions, but it is instructive to construct the protection matrix for $\mathbf{y}$. Define the matrix Y such that $Y_{0,i} = Y_{i,0} = y_i$ and

$$Y_{i,j} := \mathbb{P}_{a \in D}[a \text{ agrees with } i \text{ and with } j]$$

It is easy to see that this matrix is positive semidefinite.

Consider now the i-th row of Y. If $y_i = 0$, then the row is the all-zero vector. Otherwise, it is y_0 times the vector $\mathbf{z} := (1, Y_{1,j}/y_0, \ldots, Y_{N,i}/y_0)$. Observe that, for $j \neq 0$,

$$z_j = \frac{\mathbb{P}_{a\in D}[a \text{ agrees with } j \text{ and with } i]}{\mathbb{P}_{a\in D}[a \text{ agrees with } i]} = \mathbb{P}_{a \in D}[a \text{ agrees with } j | a \text{ agrees with } i]$$

This is the vector $\mathbf{y}(D|i)$ where $(D|i)$ is the distribution D conditioned on assignments that agree with i.

Consider now the vector $Y_0 - Y_i$. If $y_i = 1$, then this is the all-zero vector. Otherwise, it is $(y_0 - y_i)$ times the vector $\mathbf{z} := (1, (y_1 - Y_{1,i})/(1 - y_i), \ldots, (y_n - Y_{n,i})/(1 - y_i))$. We have

$$z_i = \frac{\mathbb{P}_{a\in D}[a \text{ agrees with } j \text{ but not with } i]}{\mathbb{P}_{a\in D}[a \text{ does not agree with } j]} = \mathbb{P}_{a \in D}[a \text{ agrees with } j | a \text{ does not agree with } i]$$

And this is the same as $\mathbf{y}(D|\neg i)$, where $(D|\neg i)$ is the distribution D conditioned on assignments that do not agree with i. Note, also, that $\mathbf{y}(D|\neg i)$ can be realized as a convex combination of vectors $\mathbf{y}(D|j)$, where j ranges over the other vertices that correspond to satisfying assignments for the equation of i.

These observations suggest the following prover algorithm: on input a vector of the form $\mathbf{y}(D)$, output a matrix Y as above, and then set

$$O := \{\mathbf{y}(D|i) : i \in V \text{ and } \mathbb{P}_{a \in D}[a \text{ consistent with } i] > 0\}$$

To prove Lemma 2.3, we need to find a prover strategy that succeeds for a large number of rounds starting from the vector $(1, \frac{1}{4}, \ldots, \frac{1}{4})$. The above prover strategy would work if there is a distribution over satisfying assignments for φ such that, for each equation C, each of the four satisfying assignments for C occurs with probability $\frac{1}{4}$ in the distribution. Since we want to prove an integrality gap, however, we will need to work with highly unsatisfiable instances, and so no such distribution exists.

In our proof, we essentially proceed by pretending that such a distribution exists. Every time we "look" at certain equations and have certain conditions, we refer to the distribution that is uniform over all assignments that satisfy the equations and meet the conditions; this will mean that, for example, when defining the matrix Y we will refer to different distributions when filling up different entries. If the instance is highly expanding, however, it will take several rounds for the adversary to make the prover fail. This is because if there is an adversary strategy that makes the prover fail after k rounds, we can find a non-expanding subset of the formula of size $O(k)$.

2.3.2 Fractional Solutions and Protection Matrices Based on Partial Assignments

All the fractional solutions and protection matrices produced by the prover algorithm have a special structure and are based on *partial assignments* to the variables of φ. Before describing the prover algorithm, we will describe such solutions and matrices, and prove various facts about them.

A *partial assignment* $\alpha \subseteq X \times \{0, 1\}$ is a set of assignments of values to some of the variables of φ such that each variable is given at most one value. For example, $\{(x_3, 0), (x_5, 1)\}$ is a partial assignment. A partial assignment α *contradicts* an equation of φ if it assigns values to all the three variables of the equations, and such values do not satisfy the equation; a partial assignment is *consistent* with φ if it contradicts none of the equations of φ.

If α is a consistent partial assignment, then the *restriction of* φ *to* α, denoted $\varphi_{|\alpha}$, is the set of equations that we obtain by applying the assignments of values to variables prescribed by α. (We remove the equations in which all variables are assigned, and that are reduced to $0 = 0$.)

$\varphi_{|\alpha}$ contains some equations with three variables, some equations with two variables and some equations with one variable (as we said, we remove the equations with zero variables). If an equation has two variables, we say those variables are *α-equivalent*. Note that α-equivalence is an equivalence relation, and so the variables in X not fixed by α are split into a collection of equivalence classes.

We make the following observations.

Claim 2.6 *If $\varphi_{|\alpha}$ is $(2, 1.51)$-expanding, then*

1. *Each equivalence class contains at most two variables;*
2. *If an equation contains three variables, then those variables belong to three distinct classes.*

The first part of the claim follows from the fact that, under the expansion assumption, all equations of size two are over disjoint sets of variables (otherwise, two equations of size two with a variable in common would form a set of two equations with only three occurring variables). The second part of the claim follows from the first part and from the assumption that in φ (and, for a stronger reason, in $\varphi_{|\alpha}$) two clauses can share at most one variable.

Definition 2.7 (Good partial assignment) *A partial assignment α is* good *for a formula φ if: (i) it is consistent with φ; (ii) $\varphi_{|\alpha}$ has no equation with only one variable; (iii) the set of all equations of $\varphi_{|\alpha}$ with two variables is satisfiable.*

The third condition seems very strong, but it is implied by expansion.

Claim 2.8 *Suppose that α is consistent with φ and that $\varphi_{|\alpha}$ is $(2, 1.51)$-expanding. Then α is a good partial assignment.*

Proof: $\varphi_{|\alpha}$ cannot contain an equation with a single variable, otherwise it would not even be $(1, 1.1)$-expanding. Furthermore, any pair of equations with two variables cannot have any variable in common (otherwise we would have two equations involving only 3 variables), and so it is trivial to simultaneously satisfy all the size-2 equations. ■

Definition 2.9 (α-Consistent Assignment) *If α is a good partial assignment for φ, then we say that an assignment $r \in \{0, 1\}^n$ is α-consistent if it agrees with α and if it satisfies all the equations with two variables in $\varphi_{|\alpha}$.*

Definition 2.10 (Fractional solution associated to a good partial assignment) *Let α be a good partial assignment for φ. We describe the following fractional solution $\mathbf{y} = \mathbf{y}(\alpha)$ of the independent set problem in G_φ: $y(\alpha)_0 := 1$, and for every vertex i*

$$y(\alpha)_i := \mathop{\mathbb{P}}_{r \in \{0,1\}^n} [\ r \text{ agrees with } i \mid r \text{ agrees with } \alpha \text{ and satisfies } C\ ]$$

where C is the equation of i.

Another way of thinking of $\mathbf{y}(\alpha)$ is to remove from G_φ all the vertices that are inconsistent with α, and then, for each equation, split equally among the surviving vertices for that equation a total weight of 1. Note that, in $\mathbf{y}(\alpha)$ each entry is either 1, or 1/2, or 1/4 or 0.

Claim 2.11 *Let α be a good partial assignment for φ. Then $\mathbf{y}(\alpha)$ is a feasible solution in the cone $IS(G)$.*

Proof: If two vertices i and j are connected in G, then there is a variable x such that i and j assign different values to x. If $\mathbf{y}(\alpha)$ assigns non-zero weight to both i and j, then it means that x is not assigned a value by α, and so both $y(\alpha)_i$ and $y(\alpha)_j$ are at most 1/2. ■

We also define the following "semidefinite solution."

Definition 2.12 (Protection Matrix Associated to a Partial Assignment) *Let α be a good partial assignment. To every vertex v we associate a $(d+1)$-dimensional vector $\mathbf{u}_i = \mathbf{u}_i(\alpha)$, where d is the number of equivalence classes in the set of variables of $\varphi_{|\alpha}$. When two variables are α-equivalent, we choose a representative. (Recall the earlier discussion about variables being α-equivalent.)*

- *If v is inconsistent with α, then we simply have $\mathbf{u}_i := (0, \dots, 0)$.*
- *If α assigns values to all the variables of v (consistently with v), then $\mathbf{u}_i = (1, 0, \dots, 0)$.*
- *If the equation of v has only two free variables in $\varphi_{|\alpha}$, they are in the same class, say the t-th class, and one of them is the representative. Then $\mathbf{u}_i = (1, 0, \dots, 0, \pm\frac{1}{2}, 0, \dots, 0)$, where the only non-zero entries are the 0th and the tth. The tth entry is $1/2$ if i requires the representative of the tth class to be 1; the tth entry is $-1/2$ otherwise.*
- *If the equation of v has three free variables in $\varphi_{|\alpha}$, then they are in three distinct classes, say the t_1th, the t_2th and the t_3th. Then $\mathbf{u}_i = (\frac{1}{4}, 0, \dots, \pm\frac{1}{4}, \dots \pm \frac{1}{4}, \dots, \pm\frac{1}{4}, \dots, 0)$, where the only nonzero entries are the 0th, the t_1th, the t_2th and the t_3th. The t_1th entry is $1/4$ if i requires the representative of the t_1th class to be 1, and $-1/4$ otherwise, and similarly for the other classes.*
- *Finally, let $\mathbf{u}_0(\alpha) = (1, 0 \dots, 0)$.*

Define the matrix $Y(\alpha)$ as

$$Y_{i,j}(\alpha) := \mathbf{u}_i(\alpha) \cdot \mathbf{u}_j(\alpha) \tag{2.1}$$

Note that, by definition, $Y(\alpha)$ is positive semidefinite. The matrix has the following equivalent characterization

Claim 2.13 *Let α be a good partial assignment such that $\varphi_{|\alpha}$ is $(4, 1.51)$-expanding. Then, for two vertices i, j, let C_1, C_2 be their equations; we have:*

$$Y_{i,j}(\alpha) = \mathop{\mathbb{P}}_{r \in \{0,1\}^n} [r \text{ agrees with } i \text{ and } j \,|r \text{ satisfies } C_1, C_2,\, r \text{ is } \alpha\text{-consistent }]$$

Furthermore, $Y_{0,i}(\alpha) = y_i(\alpha)$.

Proof: To simplify notation we will omit the dependency on α.

If i and j correspond to two distinct assignments for the same equation, then it is easy to see that $Y_{i,j} = 0$.

If the equation of i and the equation of j have variables in disjoint classes, then $Y_{i,j} = \mathbf{u}_i \cdot \mathbf{u}_i = \mathbf{u}_{i,0}\mathbf{u}_{j,0}$, where

$$\mathbf{u}_{i,0} = \mathop{\mathbb{P}}_{r \in \{0,1\}^n} [r \text{ agrees with } i \mid r \text{ is } \alpha\text{-consistent }]$$

and

$$\mathbf{u}_{j,0} = \mathop{\mathbb{P}}_{r \in \{0,1\}^n} [r \text{ agrees with } j \mid r \text{ is } \alpha\text{-consistent }]$$

and, using independence

$$\mathbf{u}_{i,0}\mathbf{u}_{j,0} = \mathop{\mathbb{P}}_{r \in \{0,1\}^n} [r \text{ agrees with } i \text{ and } j \mid r \text{ is } \alpha\text{-consistent }]$$

If the equation of i and the equation of j each share precisely one variable from the same class t, then either both equations must involve three variables, or one equation involves two variables and the second involves two variables from the same class. In either case we have $Y_{i,j} = \mathbf{u}_{i,0}\mathbf{u}_{j,0} + \mathbf{u}_{j,t}\mathbf{u}_{j,t}$. In the first case, if the label of i and the label of j assign consistent values to the variable(s) in class t, then $Y_{i,j} = 1/8$, otherwise $Y_{i,j} = 0$, in accordance with the claim. In the second case, if the label of i and the label of j assign consistent values to the variable(s) in class t, then $Y_{i,j} = 1/4$, otherwise $Y_{i,j} = 0$, again, in accordance with the claim.

Finally, it is impossible for two distinct equations to have each two variables in common classes. Otherwise, we would have four equations involving at most six variables and contradict expansion. ■

The matrix has also the following useful property.

Claim 2.14 *For a vertex i, let S denote the set of vertices corresponding to the equation of i which are consistent with α.*

Then

$$Y_0 - Y_i = \sum_{j \in S - \{i\}} Y_j$$

Proof: The claim follows from the fact that

$$\sum_{j \in S} \mathbf{u}_j(\alpha) = \mathbf{u}_0(\alpha) = (1, 0, \ldots, 0)$$

a fact that can be established by a simple cases analysis:

- If S contains only one element, then that element must be i, and it must be the case that $\mathbf{u}_i(\alpha) = (1, 0, \ldots, 0)$.
- If $S = \{i, j\}$ contains two elements, then i and j have $1/2$ in the first coordinate and then one has $1/2$ and another has $-1/2$ in the coordinate corresponding to the equivalence class of the two unassigned variables in the equation.
- If $S = \{i_1, i_2, i_3, i_4\}$ has four elements, then each one has $1/4$ in the first coordinates and then they have $\pm 1/4$ entries in the three coordinates corresponding to the three classes of the variables occurring in the equation. Each variable is given value zero in 2 vertices and value one in 2 vertices, so the entries in these three coordinates all cancel out.

■

2.3.3 Expansion and Satisfiability

Let α be a good partial assignments for φ, let C be an equation whose three variables are not assigned in α, and i be one of the vertices in G_φ corresponding to C. For the sake of this subsection, we think of i as being itself a partial assignment.

We define the "closure" of $\alpha \cup i$ as the output of the following algorithm

- $\beta := \alpha \cup i$;
- while φ_β has at least an equation with only one variable, of the form $x_i = b$
  - $\beta := \beta \cup \{(x_t, b)\}$
- return β

If φ_α is highly expanding, then the above algorithm terminates almost immediately, and it outputs an assignment β such that every small enough subset of the equations of φ_β are mutually satisfiable.

Lemma 2.15 *Suppose that φ_α is a $(k, 1.9)$-expanding instance of 3XOR let i be a vertex of G_φ that is not inconsistent with α. Let β be the closure of $\alpha \cup i$. Then β is a consistent partial assignment, and it fixes at most one variable not fixed in $\alpha \cup v$.*

Proof: First, we note that $\varphi_{|\alpha \cup i}$ has at most one equation with only one variable. (Otherwise we would have three equations with a total of only 5 variables in $\varphi_{|\alpha}$.)

Let α' be $\alpha \cup i$ possibly extended to assign a value to the only equation of size one in $\varphi_{|\alpha \cup i}$ so that the equation is satisfied.

Then α' is a consistent partial assignment for φ such that $\varphi_{|\alpha'}$ has no equation of size one. (Otherwise, if $\varphi_{|\alpha'}$ had an equation of size one, then there would be three equations with five variables in $\varphi_{|\alpha}$.) We conclude that $\beta = \alpha'$ and the lemma follows. ■

Lemma 2.16 (Satisfiability of Subsets of Expanding Instances) *Suppose that φ_α is a $(k, 1.9)$-expanding instance of 3XOR, let i be a vertex of G_φ corresponding to an equation involving variables not assigned by α. Let β be the closure of $\alpha \cup i$.*

Let S be any subset of at most $k-2$ equations of $\varphi_{|\beta}$. Then there is assignment that satisfies all the equations of S. Furthermore, for every equation C in S and every assignment to the variables of C that satisfies C, it is possible to extend such an assignment to an assignment that satisfies all the equations in S.

Proof: Recall that the difference between $\varphi_{|\beta}$ and $\varphi_{|\alpha}$ is that $\varphi_{|\beta}$ has either one fewer equation and at most three fewer variables than $\varphi_{|\alpha}$, or two fewer equations and at most four fewer variables than φ_α. (Depending on whether the closure algorithm performs zero steps or one step.)

Let C be an equation in $\varphi_{|\beta}$, let a be an assignment to the free variables in C that satisfies C, and let S be a smallest set of equations in φ_β such that S cannot be satisfied by an extension of a.

Suppose towards a contradiction that S contains at most $k-3$ equations.

Observe that, in $\varphi_{|\beta\cup a}$, every variable that occurs in S must occur in at least two equations of S, otherwise we would be violating minimality.

We will need to consider a few cases.

1. S cannot contain just a single equation C_1, because C_1 must have at least two variables in $\varphi_{|\beta}$, and it can share at most one variable with C.

2. Also, S cannot contain just two equations C_1 and C_2, because, for this to happen, C_1 and C_2 can have, between them, at most one variable not occurring in C, so that C, C_1 and C_2 are three clauses involving at most 4 variables in $\varphi_{|\beta}$; this leads to having either 4 equations involving at most 7 variables in $\varphi_{|\alpha}$, or to 5 equations involving at most 8 variables. In either case, we contradict the expansion assumption.

3. Consider now the case $|S| = 3$. We note that no equation in S can have three free variables in $\varphi_{|\beta\cup a}$, because then one of those three variables would not appear in the other two equations. Thus, each equation has at most two variables, each variable must occur in at least two equations, and so we have at most three variables occurring in S in $\varphi_{|\beta\cup a}$. In $\varphi_{|\alpha}$, this corresponds to either 5 clauses involving at most 9 variables, or 6 clauses involving at most 10 variables, and we again violate expansion.

4. If $|S| = 4$, then we consider two cases. If each equation in S has three free variables in $\varphi_{|\beta\cup a}$, then there can be at most 6 variables occurring in S, and we have a set of 4 equations in $\varphi_{|\alpha}$ involving only 6 variables.

   If some of the equations in S have less than three free variables, then at most a total of 5 variables can occur S in $\varphi_{|\beta\cup a}$. This means that we can find either 6 equations in φ_α involving at most 11 variables, or 7 equations involving at most 12 variables.

5. If $|S| \geq 5$, then at most $1.5 \cdot |S|$ variables can occur in S in $\varphi_{|\beta \cup a}$, and so we find either $|S|+2$ equations in $\varphi_{|\alpha}$ involving at most $\lfloor 1.5 \cdot |S| \rfloor + 6$ variables, or $|S|+3$ equations involving at most $\lfloor 1.5 \cdot |S| \rfloor + 7$ variables. Either situation violates expansion if $|S| \geq 5$.

■

2.3.4 Expansion-Correction

We will make use of the following simple fact.

Lemma 2.17 *Let ψ be an instance of 3XOR, and k be an integer. Then there is a subset $|S|$ of at most k equations such that:*

- *The instance $\psi - S$ is a $(k - |S|, 1.9)$-expanding instance of 3XOR;*
- *There is at most a total of $1.9|S|$ variables occurring in the equations in S.*

Proof: Take a largest set S of equations in ψ such that $|S| \leq k$ and at most $1.9|S|$ variables occur in S. (Note that, possibly, S is the empty set.)

Suppose towards a contradiction that $\psi - S$ is not $(k - |S|, 1.9)$-expanding. Then there is a set T of equations in $\psi - S$ such that $|T| \leq k - |S|$ and at most $1.9|T|$ variables occur in T. Then the union of S and T and observe that it contradicts te maximality assumption about S. ■

2.3.5 The Output of the Prover Algorithm

The prover algorithm takes in input a vector $\mathbf{y} = \mathbf{y}(\alpha)$ such that α is a consistent partial assignment and $\varphi_{|\alpha}$ is $(k, 1.9)$-expanding, $k \geq 4$. The output is a positive semidefinite matrix Y that is a protection matrix for $\mathbf{y}$ and a set of vectors $O \subseteq \mathbb{R}^{1+4m}$ such that each column Y_i of Y and each difference $Y_0 - Y_i$ are positive linear combinations of elements of O.

As we will see, each element of O is itself a vector of the form $\mathbf{y}(\beta)$, where β is an extension of α.

The Positive Semidefinite Matrix

The matrix Y is the matrix $Y(\alpha)$ defined in (2.1). By definition, Y is positive semidefinite. It also follows from the definition that $Y_{0,i} = Y_{i,i} = \mathbf{y}_i$.

The Set of Vectors

Because of Claim 2.14, each vector $Y_0 - Y_j$ is a non-negative linear combination of vectors Y_i, and so it is enough to prove that Y_i can be obtained as a non-negative combination of O.

In order to define the set O, we will define a set O_i for each vertex of the graph, and show that Y_i is a positive linear combination of elements of O_i. We will then define O to be the union of the sets O_i.

Let us fix a vertex i and let β be the closure of $\alpha \cup i$. Let us now find a set $|S|$ of equations of $\varphi_{|\beta}$ as in Lemma 2.17 with parameter $k-3$. The equations in S (if any), are simultaneously satisfiable by Lemma 2.16. Let A be the set of assignments that satisfy all equations in S. Define

$$O_i = \{\mathbf{y}(\beta \cup a) | a \in A\}$$

Lemma 2.18 *The vector Y_i is a non-negative combination of elements of O_i.*

Proof: We will argue that

$$Y_i = Y_{0,i} \cdot \frac{1}{|A|} \sum_{a \in A} \mathbf{y}(\beta \cup a)$$

As a warm-up, we note that $Y_{i,i} = Y_{0,i}$ (as observed before) and that $\mathbf{y}_i(\beta \cup a) = 1$ for every a (because β already sets all the variables of i consistently with the label of i).

Let us now consider $j \neq i$, and let C be the equation of i. Recall that $Y_{i,j}$ has the following probabilistic interpretation:

$$Y_{i,j} = \mathop{\mathbb{P}}_{r \in \{0,1\}^n} [\ r \text{ agrees with } i \text{ and } j \mid r \text{ preserves } \alpha\text{-consistency}, r \text{ satisfies } C, C'\ ]$$

where C is the equation of i and C' is the equation of j. We can also derive a probabilistic interpretation of the right-hand side of the equation we wish to prove

$$\frac{1}{|A|} \sum_{a \in A} \mathbf{y}_i(\beta \cup a) = \mathop{\mathbb{P}}_{a \in A, r \in \{0,1\}^n} [\ r \text{ agrees with } j \mid r \text{ satisfies } C \text{ and agrees with } \beta \cup a\ ] \tag{2.2}$$

Now we claim that the probability (2.2) is precisely the same as

$$\mathop{\mathbb{P}}_{r \in \{0,1\}^n} [\ r \text{ agrees with } j \mid r \text{ satisfies } C \text{ and agrees with } \beta\ ] \tag{2.3}$$

This is clear if the clauses in S and C share no variable outside β, because the conditioning on a has no influence on the event we are considering.

If C shares some, but not all, of its variables outside β with the clauses in S, then a random element a of A assigns uniform and independent values to such variables. This is because A is an affine space, and so if the above were not true, then A would force an affine dependency among a strict subset of the variables of C outside β; this would mean that there is a satisfying assignment for C that is inconsistent with each assignment in A, that is, there is a satisfying assignment for C that is inconsistent with S (in $\varphi_{|\beta}$), thus violating Lemma 2.16.) If C shares all its variables with the clauses of S, then a random a in A must assign to the variables of C a random satisfying assignment. (Otherwise, we would again conclude that there is a satisfying assignment for C that is inconsistent with S in $\varphi_{|\beta}$.)

The next step is to observe that, thus,

$$\frac{1}{|A|} \sum_{a \in A} \mathbf{y}(\beta \cup a)$$

is the same as the probability that a random extension of α is consistent with j conditioned on: (i) being consistent with i; (ii) satisfying the equation of j; (iii) preserving α-consistency. If we multiply by $Y_{0,i}$, what we get is the probability that a random extension of α is consistent with the labels of i and j conditioned on: (i) satisfying the equation of i; (ii) satisfying the equation of j; (iii) preserving α-consistency. And this is just the definition of $Y_{i,j}$. ■

2.3.6 Putting Everything Together

Let φ be a $(k, 1.95)$-expanding instance of 3XOR, and suppose that there is an adversary strategy that makes the game terminate after t steps.

The game begins with the solution $(1, 1/4, \ldots, 1/4)$, which is $\mathbf{y}(\emptyset)$, and, at each round, the prover picks a solution of the form $\mathbf{y}(\alpha)$ for a partial assignment α. The game ends when $\varphi_{|\alpha}$ is not a $(4, 1.9)$-expanding instance of 3XOR.

Let us denote by $\mathbf{y}(\emptyset), \mathbf{y}(\alpha_1), \ldots, \mathbf{y}(\alpha_t)$ the solutions chosen by the adversary. Note that α_r is an extension of α_{r-1} in which the variables occurring in a set S_r of clauses have been fixed, in addition to the the variables occurring in one or two clauses (call this set T_r). We also have that S_r contains at most $1.9|S_r|$ variables that do not occur in $T_{r'}$ $r' \le r$ or in $S_{r'}$, $r' \le i - 1$. By the properties of the expansion correction, $\mathbf{y}(\alpha_r)$ is $(k - \sum_{r' \le r} |S_{r'}| + |T_{r'}|, 1.9)$-expanding.

When the game terminates, we have

$$k \ge \sum_r |S_r| + |T_r| \ge k - 4$$

Let v be total number of variables occurring in the S_r and T_r. We have

$$t_v \ge 1.95(\sum_r |S_r| + |T_r|)$$

because of the expansion in φ. But we also have

$$t_v \le 3|T_r| + 1.9 \sum_r |S_r|$$

so

$$\sum_r |S_r| \le 21 \sum_r |T_r|$$

and

$$k \le 4 + 22 \sum_r |T_r| \le 4 + 44t$$

which gives $t \ge k/44 - 1/11$.

Chapter 3

LP relaxations of Vertex Cover and Max-Cut

In this chapter we study the effectiveness of the linear programs in the Lovász-Schrijver hierarchy in approximating Minimum Vertex Cover and Maximum Cut. The study of integrality gaps for these programs was initiated by Arora, Bollobás, Lovász, and Tourlakis [ABL02, ABLT06, Tou06] who proved gaps for LS relaxations of Vertex Cover. They show that even after $\Omega_\varepsilon(\log n)$ rounds the integrality gap is at least $2 - \varepsilon$ [ABLT06], and that even after $\Omega_\varepsilon((\log n)^2)$ rounds the integrality gap is at least $1.5 - \varepsilon$ [Tou06]. We show that the integrality gap is at least $2 - \varepsilon$ even after $\Omega_\varepsilon(n)$ rounds. We also show that the integrality gap for Maximum Cut is at least $2 - \varepsilon$ after $\Omega_\varepsilon(n)$ rounds.

Let $G = (V, E)$ be a graph, and assume $V = \{1, \ldots, n\}$. Recall that the cone of the linear programing relaxation of the vertex cover problem is the set of vectors $\mathbf{y} \in \mathbb{R}^{n+1}$ such that

$$\begin{aligned} y_i + y_j &\geq y_0 \quad \forall (i, j) \in E \\ 0 \leq y_i &\leq y_0 \quad \forall i \in V \\ y_0 &\geq 0 \qquad\qquad\qquad (VC(G)). \end{aligned}$$

The linear programming relaxation for MAX-CUT is a set of constraints on *n vertex variables* and *m edge variables*. For a vector $\tilde{\mathbf{y}} \in \mathbb{R}^{n+m+1}$, let $\tilde{y}_0$ be the extra coordinate for homogenization, $(\tilde{y}_1, \ldots, \tilde{y}_n)$ denote the vertex variables and $(\tilde{y}_{e_1}, \ldots, \tilde{y}_{e_m})$ denote the the edge-variables. Then the cone is the solution set of the constraints

$$\begin{aligned} \tilde{y}_e &\leq \tilde{y}_i + \tilde{y}_j & \forall e = (i, j) \in E \\ \tilde{y}_e &\leq 2\tilde{y}_0 - (\tilde{y}_i + \tilde{y}_j) & \forall e = (i, j) \in E \\ 0 \leq \tilde{y}_i &\leq \tilde{y}_0 & \forall i \in V \\ 0 \leq \tilde{y}_e &\leq \tilde{y}_0 & \forall e \in E \\ \tilde{y}_0 &\geq 0 & (MC(G)) \end{aligned}$$

The integrality gap for vertex cover is taken to be the ratio of the size of the minimum vertex cover in the graph to the optimum of the linear program. In case of MAX-CUT, we take the gap to be the inverse of this ratio. Note that in both cases the integrality gap is at least 1.

The instances for which we prove the integrality gap results are (slight modifications of) sparse random graphs. In such graphs, the size of the minimum vertex cover is $\approx n$, where n is the number of vertices, while we show the existence of a fractional solution of cost $n \cdot (\frac{1}{2} + \varepsilon)$ that remains feasible even after $\Omega_\varepsilon(n)$ rounds. The size of a maximum cut is $\approx \frac{m}{2}$, where m is the number of edges, while we show the existence of a fractional solution of cost $m \cdot (1 - \varepsilon)$ that also remains feasible after $\Omega_\varepsilon(n)$ rounds.

We use two properties of (modified) sparse random graphs. The first property is large girth; it suffices for our application that the girth be a large constant depending on ε. The second property is that for every set of $k = o(n)$ vertices, such vertices induce a subgraph containing at most $(1 + o(1))k$ edges. The same properties are also used in [ABLT06, Tou06].

In order to prove that a certain fractional solution y is feasible for a relaxation $N^k(K)$, it is sufficient to construct a matrix Y such that certain vectors obtained from the rows and columns of Y are all feasible solutions for $N^{k-1}(K)$. (By convention, $N^0(K) := K$.) As before, we use an inductive approach, where we have a theorem that says that all solutions satisfying certain conditions are feasible from $N^k(K)$; to prove the theorem we take a solution y that satisfies the conditions for a certain value of k, and then we construct a matrix Y such that all the derived vectors satisfy the conditions of the theorem for $k-1$, and hence, by inductive hypothesis, are feasible from $N^{(k-1)}(K)$, thus showing that y is feasible for $N^k(K)$. We will also use the fact that the set $N^{k-1}(K)$ is convex. This gives that, once we define the matrix Y, and we have to prove that the associated vectors are in $N^{k-1}(K)$, it suffices to express each such vector as a *convex combination* of vectors that satisfy the conditions of the theorem for $k - 1$.

Roughly speaking, in the work of Arora et al. [ABLT06] on Vertex Cover, the appropriate theorem refers to solutions where all vertices are assigned the value $1/2 + \varepsilon$, except for a set of exceptional vertices that belong to a set of constant-diameter disks. Oversimplifying, to prove a lower bound of k rounds, one needs to consider solutions that have up to k disks, and for the argument to go through one needs the union of the disks to induce a forest, hence the lower bound is of the same order as the girth of the graph. Tourlakis [Tou06] does better by showing that, due to extra conditions in the theorem, the subgraph induced by k "disks" has diameter $O(\sqrt{k})$, and so it contains no cycle provided that the girth of the graph is sufficiently larger than $\sqrt{k}$. This yields an integrality gap result that holds for a number of rounds up to a constant times the square of the girth of the graph.[1]

The solutions in our approach have a similar form, but we also require the disks to be far away from each other. When we start from one such solution y, we construct a matrix Y, and consider the associated vectors, we find solutions where disks are closer to each other than allowed by the theorem, and we have to express such solutions as convex combinations of allowed solutions. Roughly speaking, we show that such a step is possible provided that the union of the "problematic" disks (those that are too close to each other) induces a very sparse graph. Due to our choice of random graph, this is true provided that there are at most $c_\varepsilon \cdot n$ disks, where c_ε is a constant that depends only on ε. We also show that, in order to prove an integrality gap for k rounds, it is sufficient to consider solutions with $O(k)$ disks, and so our integrality gap applies even after $\Omega_\varepsilon(n)$ rounds.

[1] Arora et al. [ABLT06, Tou06] present their proofs in the language of a "prover-verifier" game, but they can be equivalently formulated as inductive arguments.

Hence (again, roughly speaking) our improvement over previous work comes from the fact that it suffices that the union of the disks induce a sparse graph (something which is true for a sublinear number of disks) rather than induce a forest (a requirement that fails once we have a logarithmic or polylogarithmic number of disks). This oversimplified sketch ignores some important technical points: We will give a more precise overview in Section 3.2.

Linear versus Semidefinite Relaxations

It is interesting to compare our results for Maximum Cut in the LS hierarchy with the known results for the LS+ hierarchy. After applying one round of LS+ to the basic linear programming relaxation of Max Cut one obtains the Goemans-Williamson relaxation, which yields a .878 approximation. In contrast, we show that even after $\Omega_\varepsilon(n)$ rounds of LS the integrality gap remains $\frac{1}{2} + \varepsilon$. This gives a very strong separation between the approximability of LS versus LS+ for a natural problem.

Also, for vertex cover, proving strong integrality gaps for the LS+ hierarchy of semidefinite programs is likely to require significantly different techniques than the ones used here. For proving LS integrality gaps discussed in this chapter, we rely on random graphs. However, it is known that the Theta function of a random sparse graph is very small, and so already one round of LS+ provides a good approximation to Vertex Cover on random graphs.

For one round of LS+ (or, equivalently, for the function defined as number of vertices minus the Theta function) Goemans and Kleinberg [KG98] had earlier proved a $2 - o(1)$ integrality gap result by using a different family of graphs studied initially by Frankl and Rödl [FR87]. Charikar [Cha02] proves a $2-o(1)$ integrality gap result for a semidefinite programming relaxation of Vertex Cover that includes additional inequalities using the same family of graphs. Charikar's relaxation is no tighter than 3 rounds of LS_+, and is incomparable with the relaxation obtained after two rounds. These results were later extended to a gap of $2 - \varepsilon$ for $\Omega_\varepsilon(\sqrt{\log n / \log\log n})$ rounds of LS+ by Georgiou et. al. [GMPT07]. Proving optimal gaps in the LS+ hierarchy (for $\Omega(n)$) rounds remains a challenging open problem.

Other Related Work

Independently of this work, Fernandez de la Vega and Kenyon [dlVKM07] proved, for every $\varepsilon > 0$, that the integrality gap of Max Cut remains at least $2 - \varepsilon$ even after a constant number of rounds of LS. Their result also applies to the more powerful Sherali-Adams method [SA90]. (The relaxation obtained after r levels of Sherali-Adams is at least as tight as the relaxation obtained after r round of Lovász-Schrijver.) The techniques used by them were very similar the ones discussed in this chapter.

These techniques were also used later by Charikar, Makarychev and Makarychev [CMM07a] in construction of metrics on n points, such that the metric defined by any k of these points embeds isometrically into ℓ_1 (in fact, into ℓ_2) but embedding the entire metric on n points into ℓ_1 requires distortion at least $\Omega(\log n/(\log k + \log\log n))$. They also used these metrics to show that integrality gap for Minimum Vertex Cover and Maximum Cut remains at least $2 - \varepsilon$ for the linear program obtained by $\Omega(n^{\delta})$ levels of the Sherali-Adams hierarchy, for $\delta = \delta(\varepsilon)$ [CMM09], and to establish integrality gaps for Unique Games.

3.1 Our Results

Define an $(\alpha, \delta, \gamma, \eta)$ graph G on n vertices as a graph with girth $\delta \log(n)$, and such that no vertex cover of size $(1-\alpha)n$ exists and each induced subgraph of G with $k \leq \gamma n$ vertices, has at most $(1+\eta)k$ edges.

Lemma 3.1 *For every $0 < \alpha < 1/125$, $\eta > 0$, there exists a $d = d(\alpha) \in \mathbb{N}$, $\delta, \gamma > 0$, and $N \in \mathbb{N}$ such that for $n \geq N$ there exists an $(\alpha, \delta, \gamma, \eta)$ graph with max cut less than $\frac{1}{2}|E|(1+\alpha)$ and maximum degree at most d on n vertices. Here $d(\alpha)$ is an explicit function that depends only on α.*

Lemma 3.2 *For every η, δ, $\gamma > 0$, $0 < \varepsilon < 1/20$, $d \in \mathbb{N}$ if G is an $(\alpha, \delta, \gamma, \eta)$ graph with maximum degree at most d on n vertices then $(1, 1/2+\varepsilon, \ldots, 1/2+\varepsilon) \in N^{\Omega_{\varepsilon,\eta,\delta,\gamma,d}(n)}(VC(G))$ if $\eta \leq \eta(\varepsilon, d)$ where $\eta(\varepsilon, d)$ is an explicit function that depends only on ε and d.*

Lemma 3.3 *For every η, δ, $\gamma > 0$, $0 < \varepsilon < 1/20$, $d \in \mathbb{N}$ if G is an $(\alpha, \delta, \gamma, \eta)$ graph with maximum degree at most d on n vertices then the solution $\mathbf{y}$ defined as $y_0 := 1$, $y_i := 1/2 + \varepsilon$ and $y_e := 1 - 2\varepsilon$ is in $N^{\Omega_{\varepsilon,\eta,\delta,\gamma,d}(n)}(MC(G))$ if $\eta \leq \eta(\varepsilon, d)$ where $\eta(\varepsilon, d)$ is an explicit function that depends only on ε and d.*

Theorem 3.4 *For all $0 < \zeta < 1/50$, there is a constant $c_\zeta > 0$ such that, for all sufficiently large n, the integrality gap for vertex cover after $c_\zeta n$ rounds is at least $2 - \zeta$.*

Proof: Let $\alpha = \zeta/6$ and $\varepsilon = \zeta/6$. Let $d = d(\alpha)$ where $d(\alpha)$ is as in Lemma 3.1. Let $\eta = \eta(\varepsilon, d)$ where $\eta(\varepsilon, d)$ is as in Lemma 3.2. Then by Lemma 3.1, there exists a $\delta, \gamma > 0$, $N \in N$ such that such that for $n \geq N$ there exists an $(\alpha, \delta, \gamma, \eta)$ graph with maximum degree at most d on n vertices. By Lemma 3.2, the vector $(1, 1/2+\varepsilon, \ldots, 1/2+\varepsilon) \in N^{\Omega_{\varepsilon,\eta,\delta,\gamma,d}(n)}(VC(G))$ because $\eta = \eta(\varepsilon, d)$. This exhibits an integrality gap of $\frac{1-\alpha}{1/2+\varepsilon} = \frac{1-\zeta/6}{1/2+\zeta/6} \geq 2 - \zeta$. ■

Similarly, we have

Theorem 3.5 *For all $0 < \zeta < 1/50$, there is a constant $c_\zeta > 0$ such that, for all sufficiently large n, the integrality gap for max cut after $c_\zeta n$ rounds is at least $2 - \zeta$.*

Lemma 3.1 is very similar to results already known in the literature (for example [ABLT06]) and so we only prove the additional properties that we require in the appendix. Most of this chapter is dedicated to a proof of Lemma 3.2. Lemma 3.3 will follow via a relative simple "reduction" to Lemma 3.2.

3.2 Overview of the Proof

If D is a random variable ranging over vertex covers, then the solution $\mathbf{y}_D$ where $y_0 = 1$ and $y_i = \mathbb{P}[i \in D]$ is a convex combination of integral solutions, and so it survives an arbitrary number of rounds of LS. The protection matrix for $\mathbf{y}_D$ is the matrix $Y = Y_D$ such that $Y_{i,j} = \mathbb{P}[i \in D \wedge j \in D]$.

In trying to show that a given vector $\mathbf{y}$ survives several rounds of LS, it is a good intuition to think of $\mathbf{y}$ as being derived from a probability distribution over vertex covers (even if $\mathbf{y}$ is not a convex combination of integral solutions, and cannot be derived in this way) and, in constructing the protection matrix Y, to think of Y as being derived from the said distribution as above.

Note that for the above matrix, the vectors $\mathbf{z} = \mathbf{Y}_i/y_i$ and $\mathbf{w} = (\mathbf{Y}_0 - \mathbf{Y}_i)/(1 - y_i)$ correspond to conditional distributions with $z_j = \mathbb{P}[j \in D | i \in D]$ and $w_j = \mathbb{P}[j \in D | i \notin D]$. To show that $\mathbf{y} \in N^k(VC(G))$, we must show that $\mathbf{z}, \mathbf{w} \in N^{k-1}(VC(G))$ for the vectors $\mathbf{z}$ and $\mathbf{w}$ corresponding to every i. The kth row in the protection matrices may now be interpreted as the distribution obtained by further conditioning on k. Intuitively, more rounds of LS correspond to further conditioning on other vertices which do not already have probability 0 or 1 in these conditional distributions. We often refer to vertices having probability 0/1 as being *fixed* in the distribution.

Since only r vertices can be conditioned upon in r rounds, we only need to create solutions that look "locally" like distributions over vertex covers for small sized subgraphs. Also, because the given graph has large girth, subgraphs of size $O(\log n)$ are trees. We thus start by expressing the vector $\mathbf{y} = (1, 1/2 + \varepsilon, \ldots, 1/2 + \varepsilon)$ as a probability distribution over vertex covers for a tree. This distribution we define has the property that conditioning on a vertex i only affects the vertices upto a constant distance ℓ from i. In fact, the effect of conditioning decreases exponentially with the distance from i and we explicitly truncate it at distance $\ell = O(\frac{1}{\varepsilon}\log(\frac{1}{\varepsilon}))$. The conditional distribution is referred to as a *splash* around i as it creates "ripples" (change in probabilities) which decrease with distance from i. Fernandez de la Vega and Kenyon [dlVKM07, Section 5] describe essentially the same distribution of vertex covers over trees in their paper, suggesting its usefulness for proving integrality gaps for the vertex cover problem.

We start with the vector $(1, 1/2 + \varepsilon, \ldots, 1/2 + \varepsilon)$ for the given graph G. After one round of LS, each row i of the protection matrix is defined by changing only weights of vertices within distance a distance ℓ of vertex i according to a splash. Since it affects only a small subgraph, which is a tree rooted at i, the solution "looks" locally like a valid conditional distribution.

Now consider trying to extend this strategy to a second round. Say we want to show that the ith row of the protection matrix above survives another round. We thus need to create another protection matrix for this row. Each row of this new matrix corresponds to conditioning on some other vertex j. If i and j are at distance greater than 2ℓ, the weights (probabilities) of vertices within a distance ℓ from j are still $1/2 + \varepsilon$. The conditional distribution can then be created by replacing these values according to a splash around j and leaving the weights of the other vertices as unchanged. If the distance between i and j is less than 2ℓ and k is a vertex within distance ℓ of either i or j, we modify the weight of k according to the probability that *both* i and j are in the vertex cover.

It would become, unfortunately, very complex to proceed for a large number of rounds with this kind of analysis, and it would appear that the girth of the graph would be a natural limit for the number of rounds for which we can extend this line of argument. (See indeed [ABLT06, Tou06].)

We note however that certain cases are simpler to handle. Suppose that we are given a vector **y** that is $1/2 + \varepsilon$ everywhere except in a number of balls, all at distance at least 5ℓ from each other, in which the values of **y** are set according to splashes. Then the above ideas can be used to define a valid protection matrix. Unfortunately, this does not seem to help us in setting up an inductive argument, because the structure of the vector that we start from is not preserved in the rows of the protection matrix: we may end up with splash areas that are too close to each other, or with the more special structures that we get by conditioning on a vertex less than distance 2ℓ from the root of a splash.

Our idea, then, is to take such more complex vectors and express them as convex combinations of vectors that are $1/2 + \varepsilon$ everywhere except in splash areas that are at distance at least 5ℓ from each other. We will refer to such solutions as *canonical* solutions. Since we are trying to show that the complex vector belongs to some convex cone, it suffices to show that each one of these simpler vectors is in the cone. Now we are back to the same type of vectors that we started from, and we can set up an inductive argument.

Our inductive argument proceeds as follows: we start from a solution **y** in a "canonical" form, that is, such that all vertices have value $1/2+\varepsilon$ except for the vertices belonging to at most k splashes; furthermore, the roots of any two splashes are at distance at least 5ℓ from each other. We need to construct a protection matrix Y for this vector. To define the jth row $\mathbf{Y}_j$ of the protection matrix we reason as follows: if j is far (distance $> 2\ell$) from the roots of all the splashes in **y**, then $\mathbf{Y}_j$ looks like **y**, plus a new splash around j. If j is at distance $\leq 2\ell$ from a splash (and, necessarily, far from all the others) rooted at a vertex r, then we replace the splash rooted at r with a new splash which corresponds to our original distribution over trees conditioned on both r and j.

If $\mathbf{Y}_j$ happens to be a vector in canonical form, we are done, otherwise we need to express it as a convex combination of vectors in canonical form. There are two ways in which $\mathbf{Y}_j$ can fail to be canonical: j may be at distance more than 2ℓ but less than 5ℓ from the closest splash; in this case the new splash we create around j is too close to an already existing one. The other possibility is that j is at distance less than 2ℓ from an existing splash, in which case $\mathbf{Y}_j$ contains a "doubly-conditioned" splash which is not an allowed structure in a canonical solution.

Our idea is then to define a set S of "problematic vertices," namely, the vertices in the two close splashes, in the first case, or the vertices in the doubly-conditioned splash, in the second case. Then we prove that[2] that the restriction of **Y** to small (sub-linear) subset S of vertices can be expressed as a distribution of valid integral vertex covers over S. We would then like to use this fact to express **y** itself as a convex combination of solutions that are integral over S and agreeing with **y** outside S; if we could achieve this goal, we would have expressed **y** as a convex combination of vectors where the "problematic" coordinates of **y** are fixed, and the other coordinate are as nice as they were in **y**.

Unfortunately, some complications arise. In order to express **y** as a convex combination $\sum_a \lambda_a \mathbf{y}_a$ of vectors such that each $\mathbf{y}_a$ is fixed in S, it is necessary that each $\mathbf{y}_a$ contains a splash around each of the newly fixed variables. The new splashes may themselves be at distance less than 5ℓ from each other, making the $\mathbf{y}_a$ not canonical. To remedy this problem, we define S (the set of vertices that will be fixed in the $\mathbf{y}_a$) via the following process: we initialize S to the initial set of problematic vertices, then we add all vertices that are at distance less than ℓ from S and that can be connected via

[2] Assuming some added conditions on the fractional solution **y**, called *saturation*.

a path of length $\leq 5\ell$ that does not pass through S, and so on. At the end of this process, we express $\mathbf{y}$ restricted to S as a convex combination of integral covers, and we extend each of these integral covers over S to a fractional solution over all vertices (by putting splashes around the vertices of S) and so express $\mathbf{y}$ as a convex combination of solutions that, now, are canonical.

The argument works provided that S is of sublinear size. A careful accounting guarantees that, if we want to show that our solution survives k rounds, we only need to consider instances where S is of size $O(k)$. Intuitively, this is due to the fact that each time we make S larger we discover a short path of length $t \leq 5\ell$ in the graph, and we add to the subgraph induced by S $t-1$ new vertices and t new edges. The subgraph induced by S can only include at most $|S|(1+\eta)$ edges, for some very small η, so it cannot happen that S grows too much at each step, because it is not possible to consistently add more edges than vertices to the subgraph induced by S without causing a contradiction to the sparsity condition.

Since this ensures that it takes $\Omega(n)$ rounds before the set of fixed vertices grows to size γn, we can survive $\Omega(n)$ rounds.

3.3 Distributions of Vertex Covers in Trees

As a first (and useful) idealized model, suppose that our graph is a rooted tree. Consider the following distribution over valid vertex covers:

- The root belongs to the cover with probability $1/2 + \varepsilon$
- For every other vertex i, we make (independently) the following choice: if the parent of i does not belong to the vertex cover, then i is in the cover with probability one; if the parent of i is in the cover, then with probability $2\varepsilon/(\frac{1}{2}+\varepsilon)$ we include i in the cover, and with probability $1 - 2\varepsilon/(\frac{1}{2}+\varepsilon)$ we do not include i in the cover.

(The distribution is sampled by considering vertices in the order of a BFS, so that we make a decision about a vertex only after having made a decision about the parent.)

This is an instantiation of the Ising Model, about which much is known, but we will need only very elementary facts about it. The proofs of these facts are contained in the appendix.

A first observation is that each vertex has probability $1/2 + \varepsilon$ of being in the cover and $1/2 - \varepsilon$ of not being in the cover. The second observation is that, if we condition on the event that, say, the root is in the cover, then this condition affects very heavily the vertices that are close to root, but this effect decreases exponentially with the distance. In particular, for each vertex whose distance from the root is about $4\varepsilon^{-1} \cdot (\log \varepsilon^{-1})$, the probability of the vertex being in the cover condition on the root being in the cover is between $1/2 + \varepsilon - \varepsilon^4$ and $1/2 + \varepsilon + \varepsilon^4$, and the same is true conditioning on the root not being in the cover.

This second observation will show that reasoning about this distribution is useful to deal with graphs that are only *locally* like trees, that is, graphs of large girth. Before discussing this application, we slightly change the distribution so that, after a certain distance from the root, there is no effect (rather than a small effect) if we condition on the root being or not being in the cover. Hence the effect of conditioning on the root is explicitly cut-off after a certain distance.

In particular, consider the following two distributions which sample from the vertex covers of a tree rooted at a vertex i. The conditioning on the root only affects vertices upto a distance $\ell = \frac{8}{\varepsilon} \log \frac{1}{\varepsilon}$ of i.

Definition 3.6 *For $b \in \{0,1\}$ we define a b-Splash around a vertex i as the distribution which modifies vertices upto a distance of 2ℓ as follows*

1. $i = b$

2. *For every vertex upto distance ℓ (and at distance greater than $\ell+1$), we independently decide to include it with probability 1 if its parent is not in the vertex cover and with probability $2\varepsilon/(\frac{1}{2}+\varepsilon)$ if its parent is already in the vertex cover.*

3. *For u and v at distances ℓ, $\ell+1$ respectively, we include v with probability 1 if u is not in the vertex cover and with probability*

$$\frac{\mathbb{P}[u = 1|i = b] - \left(\frac{1}{2} - \varepsilon\right)}{\mathbb{P}[u = 1|i = b]}$$

otherwise.

Where $u = 1$ denotes the event $u \in D$ for a random variable D (with distribution defined by the splash) ranging over the vertex covers of the graph.

For the above to be well defined, we need $\mathbb{P}[u = 1|i = b] > 1/2 - \varepsilon$ for a vertex u at distance ℓ from i. Claim 3.7 shows that in fact $\mathbb{P}[u = 1|i = b] \in [1/2 + \varepsilon - \varepsilon^4, 1/2 + \varepsilon + \varepsilon^4]$ for u at distance greater than $\ell/2$ and hence the probability at distance ℓ is non-negative.

Claim 3.7 *Consider a b-Splash around any vertex i such that all vertices upto distance ℓ are labeled $\frac{1}{2}+\varepsilon$. Let j be a vertex such that $d(i,j) \le \ell$. Then,*

1. $\mathbb{P}[j = 1|i = 1, d(i,j) = k] = (1/2+\varepsilon)\left[1 + (-1)^k \left(\frac{1/2-\varepsilon}{1/2+\varepsilon}\right)^{k+1}\right]$ *for* $0 \le k \le \ell$

   $\mathbb{P}[j = 1|i = 0, d(i,j) = k] = \mathbb{P}[j = 1|i' = 1, d(i',j) = k-1]$ *for* $1 \le k \le \ell$

2. $|\mathbb{P}[j = 1|i = b, d(i,j) = \ell/2] - (1/2+\varepsilon)| \le \varepsilon^4$

3. $\mathbb{P}[j = 1|i = 1, d(i,j) = k] + \mathbb{P}[j = 1|i = 1, d(i,j) = k+1] \ge 1 + 4\varepsilon^2$ *for* $0 \le k \le \ell$

Note, in particular, that the probabilities are independent of i and j and depend only on their distance $d(i,j)$. Also, the difference of the probabilities from $1/2+\varepsilon$ decreases exponentially with distance. The following claim shows that the vertices outside a radius of ℓ from i are independent of whether or not i is in the cover.

Claim 3.8 *If we pick a 0-Splash with probability* $1/2 - \varepsilon$ *and a 1-Splash with probability* $1/2 + \varepsilon$*, then all vertices have probability* $1/2 + \varepsilon$*. Furthermore, vertices at distance* $\ell + 1$ *or more from i have weight* $1/2 + \varepsilon$ *in the 0-Splash as well as 1-Splash around i.*

The vectors that appear in our argument may involve conditioning on a vertex i that has value different from $1/2 + \varepsilon$ based on a splash distribution around a vertex r close to it. The following claims allow us to compute $\mathbb{P}[i = 1, j = 1 | r = b]$, the probability of two vertices i, j being simultaneously present in a b-Splash at r, and also $\mathbb{P}[i = 0, j = 1 | r = b]$, which is the probability that j is present and i is not. We defer the proofs to the appendix.

Claim 3.9 *Let* $i = v_0, v_1, \ldots, v_{m-1}, v_m = j$ *be the path to j,* $m \leq \ell$*, and let u be the vertex on this path which is closest to r. Then*

1. $$\mathbb{P}[i = 1, j = 1 | r = b] = \mathbb{P}[u = 1 | r = b] \cdot \mathbb{P}[i = 1 | u = 1] \cdot \mathbb{P}[j = 1 | u = 1] + \mathbb{P}[u = 0 | r = b] \cdot \mathbb{P}[i = 1 | u = 0] \cdot \mathbb{P}[j = 1 | u = 0]$$

2. *If* $\mathbb{P}[u = 1 | r = b] = 1/2 + \varepsilon$*, then* $\mathbb{P}[i = 1, j = 1 | r = b] = (1/2 + \varepsilon)\, \mathbb{P}[j = 1 | i = 1]$

The first part of the above claim states that once we condition on u, then i and j are independent. The second part states that if u is sufficiently far r, we can ignore r completely and just compute the probability of j as determined by a splash around i.

Claim 3.10 *Let i be a vertex and (j,k) be an edge in a b-Splash around r and let* $b' \in \{0, 1\}$*.*

$$\mathbb{P}[i = b', j = 1 | r = b] + \mathbb{P}[i = b', k = 1 | r = b] \geq \mathbb{P}[i = b' | r = b] \cdot (1 + 4\varepsilon^3)$$

The next claim allows us to treat vertices that are sufficiently far from each other as almost independent in the distribution conditioned on r.

Claim 3.11 *Let i and j be two vertices in a b-Splash around r, such that* $d(i, j) \geq \ell$*. Then*

$$\left|\mathbb{P}[i = b', j = 1 | r = b] - \mathbb{P}[i = b' | r = b] \cdot \mathbb{P}[j = 1 | r = b]\right| \leq 2\varepsilon^4$$

3.4 Distribution of Vertex Covers in Sparse Graphs

To reduce solutions with more complicated structure to simpler solutions, we will need to show that if we look at a sufficiently small subgraph of our original graph obtained in Lemma 3.1, then the more complicated solution can be expressed as a convex combination of 0/1 solutions.

The following result is proved in [ABLT06].

Lemma 3.12 ([ABLT06]) *Let* $\eta \leq \frac{2\varepsilon}{3+10\varepsilon}$ *and let* $G = (V, E)$ *be a graph such that*

1. *for each* $S \subseteq V$*,* $G(S) = (V_{G(S)}, E_{G(S)})$*, then* $|E_{G(S)}| \leq (1 + \eta)|V_{G(S)}|$*.*

2. *girth*$(G) \geq \frac{1+2\varepsilon}{\varepsilon}$.

Then there exists a distribution over vertex covers on G such that each vertex belongs to the vertex cover with probability $1/2 + \varepsilon$.

We will need a slight generalization. Instead of requiring the solution to have the value $1/2 + \varepsilon$ everywhere, we only require that the sum of the values on each edge should be at least $1 + 2\varepsilon$, if both of its endpoints are not already fixed.

Definition 3.13 *We call a fractional solution y for a graph G ε-*saturated *if for each edge (i, j) in graph G either:*

- *Both i and j are fixed and $y_i + y_j \geq 1$ or,*
- $y_i + y_j \geq 1 + 2\varepsilon$.

We now show that the under the conditions of the previous lemma, every ε-saturated solution can be written as a convex combination of vertex covers of the graph.

Lemma 3.14 *Let $\eta \leq \frac{2\varepsilon}{3+10\varepsilon}$ and let $G = (V, E)$ be a graph such that*

1. *for each $S \subseteq V$, $G(S) = (V_{G(S)}, E_{G(S)})$, then $|E_{G(S)}| \leq (1 + \eta)|V_{G(S)}|$.*
2. *girth*$(G) \geq \frac{1+2\varepsilon}{\varepsilon}$.

and let $\mathbf{y}$ *be an ε-saturated solution. Then there exists a distribution over vertex covers on G such that each vertex i belongs to the vertex cover with probability y_i.*

Proof: For the graph G, we will create a set of feasible fractional solutions $\mathbf{y}(k) \in \{0, 1/2 + \varepsilon, 1\}^{|V|}$ such that $\mathbf{y}$ is a convex combination of these vectors.

We partition V into V_0, $V_{1/2+\varepsilon}$, and V_1, as follows:

$$i \in \begin{cases} V_0 & y_i < 1/2 + \varepsilon \\ V_{1/2+\varepsilon} & y_i = 1/2 + \varepsilon \\ V_1 & y_i > 1/2 + \varepsilon \end{cases}$$

We define $t(i)$ as follows:

$$t(i) = \begin{cases} 1 - \frac{y_i}{1/2+\varepsilon} & i \in V_0 \\ 1 & i \in V_{1/2+\varepsilon} \\ \frac{y_i-(1/2+\varepsilon)}{1/2-\varepsilon} & i \in V_1 \end{cases}$$

We can order the $t(i)$'s: $0 \leq t(i_1) \leq t(i_2) \leq \cdots \leq t(i_{|V|}) \leq 1$. For each $k : 1 \leq k \leq |V|$ we create the vector $\mathbf{y}(k)$ where

$$\mathbf{y}(k)_i = \begin{cases} 0 & i \in V_0 \text{ and } t(i) \leq t(i_k) \\ 1 & i \in V_1 \text{ and } t(i) \leq t(i_k) \\ 1/2 + \varepsilon & \text{otherwise} \end{cases}$$

We claim the distribution where $\mathbf{y}(k)$ occurs with probability $t_{i_k} - t_{i_{k-1}}$ gives us $\mathbf{y}$.

If $i \in V_0$, then it will be 0 with probability t_i and $1/2 + \varepsilon$ with probability $1 - t_i = \frac{y_i}{1/2+\varepsilon}$. Therefore the probability that i is in the vertex cover is y_i. If $i \in V_1$, then it will be 1 with probability $t_i = \frac{y_i-(1/2+\varepsilon)}{1/2-\varepsilon}$ and $1/2 + \varepsilon$ with probability $1 - t_i = 1 - \frac{y_i-(1/2+\varepsilon)}{1/2-\varepsilon}$. Therefore the probability that i is in the vertex cover is $\frac{y_i-(1/2+\varepsilon)}{1/2-\varepsilon} + (1/2 + \varepsilon)(1 - \frac{y_i-(1/2+\varepsilon)}{1/2-\varepsilon}) = y_i$. If $i \in V_{1/2+\varepsilon}$, then it is clear that the probability that i is in the vertex cover is $1/2 + \varepsilon$.

Note that all the weights in each $\mathbf{y}(k)$ are 0, 1 or $1/2+\varepsilon$. It remains to show that in each of these $\mathbf{y}(k)$ any edge which contains one vertex fixed to 0 has the other vertex fixed to 1. First, note that all neighbors of vertices in V_0 are in V_1. It suffices to show that if i and j are adjacent, $i \in V_1$, $j \in V_0$, that $t(i) \geq t(j)$. However

$$\begin{aligned}
t(i) - t(j) &= \frac{y_i - (1/2+\varepsilon)}{1/2-\varepsilon} - \frac{(1/2+\varepsilon) - y_j}{1/2+\varepsilon} \\
&= \frac{(y_i + y_j)/2 + \varepsilon(y_i - y_j) - (1/2+\varepsilon)}{1/4-\varepsilon^2} \\
&\geq \frac{(1+2\varepsilon)/2 + \varepsilon(y_i - y_j) - (1/2+\varepsilon)}{1/4-\varepsilon^2} \\
&= \frac{\varepsilon(y_i - y_j)}{1/4-\varepsilon^2} \geq 0
\end{aligned}$$

which concludes the proof of the lemma. ■

3.5 The Main Lemma

We now define the type of solutions that will occur in our recursive argument.

Let $G = (V, E)$ be an $(\alpha, \delta, \gamma, \eta)$ graph with n vertices and degree at most d, as in the assumption of Lemma 3.2. We define the constant $C = \sum_{i=1}^{\ell+1} d^i$ as the maximum number of vertices within a distance ℓ from some vertex and $D = 5\ell C$ as the maximum number of vertices within distance ℓ of all the vertices in a path of length 5ℓ. Choose $\eta = \frac{1}{3D}$. Note that η depends on only ε and d. Also, we assume that n is large enough that the girth of the graph is larger than various fixed constants throughout. We fix G for the rest of this section.

Let $R = \frac{\gamma n}{C+2D}$

Let $G_{|S} = (S, E_{|S})$ be the subgraph of G induced by $S \subseteq V$. For some set $S \subseteq V$, define $N_S(i) = \{j$: there exists path of length ℓ from i to j using only edges in $E \backslash E_{|S}$).

Definition 3.15 *We say that a vector $y = (y_0, \ldots, y_n)$ is r-canonical if there exists a set $S \subseteq V$ such that:*

- $\forall j \in S$ $y_j \in \{0, 1\}$ *and* $\mathbf{y}_{|S}$ *is a vertex cover of* $G_{|S}$
- *For every two vertices in S the shortest path between them that uses only vertices not in S has length $> 5\ell$. (Therefore if $i, j \in S$, $i \neq j$, then $N_S(i) \cap N_S(j) = \emptyset$).*

- $$y_i = \begin{cases} \mathbb{P}[i = 1 | j = y_j] & \exists j \in S \;\; s.t.\; i \in N_S(j) \\ 1/2 + \varepsilon & o.w \end{cases}$$

- $|S| \leq rC + 2rD$

- *Let* $|S| = rC + kD$ $(k \leq 2r)$ *and* $G_{|S} = (S, E_{|S})$ *is the subgraph of* G *induced by* S*, then*

$$|E_{|S}| - |S| \geq k - r$$

We call a set S as in Definition 3.15 a witness.

Claim 3.16 *If* $\mathbf{y}$ *is an r-canonical vector then,* $\mathbf{y} \in VC(G)$*. Moreover,* $\mathbf{y}$ *is* ε^2*-saturated.*

Proof: This follows from the fact all edges are either internal to S, internal to some $N_S(i)$, internal to $V \setminus \cup_{i \in S} N(i)$ or between some $N(i)$ and $V \setminus \cup_{i \in S} N(i)$. In the first case, it follows because $\mathbf{y}_{|S}$ is a valid vertex cover having only 0/1 values. In the second because of the fact that a $N(i)$ is weighted according to a splash and Claim 3.7. In the third case, because the weights are all $1/2 + \varepsilon$. The final case just concerns the vertices at distance ℓ and $\ell + 1$ from the center of a splash and again follows from Claim 3.7. ■

Lemma 3.2 follows from the above claim, the following result and the fact that $(1, 1/2 + \varepsilon, \ldots, 1/2 + \varepsilon)$ is 0-canonical.

Lemma 3.17 *Let* $\mathbf{y}$ *be an r-canonical solution, and* $r \leq R$*. Then* $\mathbf{y}$ *is in* $N^{R-r}(VC(G))$*.*

Proof: We prove it by induction on $R - r$. By Claim 3.16, an R-canonical solution is feasible for $VC(G)$, and this gives the basis for the induction.

Let $\mathbf{y}$ be an r-canonical solution and let S be a witness to $\mathbf{y}$. We show that there is a protection matrix Y for $\mathbf{y}$ such that $(\mathbf{Y}_i)/y_i$ and $(\mathbf{Y}_0 - \mathbf{Y}_i)/(y_0 - y_i)$ are distributions over $(r+1)$-canonical vectors for $y_i \neq 0, y_0$. If $y_i = 0$, then we take $Y_i = 0$ which is in $N^k(VC(G))$ for all k and $\mathbf{Y}_0 - \mathbf{Y}_i = \mathbf{Y}_0$ which is r-canonical.

The protection matrix is defined as follows. (When we talk about distance between vertices, we mean distance via paths that do not go through any vertex in S.)

- $Y_{i,0} = Y_{0,i} = Y_{i,i} = y_i$.
- If i and j are at distance greater than ℓ from each other, then $Y_{i,j} = y_i \cdot y_j$
- If i is at distance greater than 2ℓ from the closest vertex in S, and j is at distance at most ℓ from i, then $Y_{i,j}$ is the probability that i and j both belongs to a vertex cover selected according to a splash distribution around $Y_{ij} = y_i \, \mathbb{P}[j = 1 | i = 1]$
- If i is at distance at most 2ℓ from a vertex $r \in S$, and j is at distance at most ℓ from i, then Y_{ij} is the probability that i and j both belong to a vertex cover selected according to a b-Splash distribution around r i.e. $Y_{ij} = \mathbb{P}[i = 1, j = 1 | r = b]$

Claim 3.18 *The matrix Y is symmetric.*

Proof: If $d(i, j) > \ell$, clearly $Y_{ij} = Y_{ji}$. There remain three additional cases.

- First, if both i and j are at distance greater than 2ℓ from any vertex in S, then $y_i = y_j = 1/2+\varepsilon$ and also $\mathbb{P}[j = 1|i = 1] = \mathbb{P}[j = 1|i = 1]$ as it depends only on the distance by Claim 3.7, and hence $Y_{ij} = Y_{ji}$.
- Second, both i and j are at distance at most 2ℓ from any vertex in S. Both i and j cannot be close to two different vertices in S because then $d(i, j) \le \ell$ would imply a path of length at most 5ℓ between the two vertices which is not possible. Hence, in this case, $Y_{ij} = Y_{ji} = \mathbb{P}[i = 1, j = 1|r = b]$, where r is the vertex in S close to both i and j.
- Finally, if $d(i, r) \le 2\ell$ for some $r \in S$ and $d(j, r) > 2\ell \; \forall r \in S$, then the path from i to j cannot come closer than distance $\ell + 1$ to r. If l is the vertex on this path closest to r, then we have $P^b_r(l) = 1/2 + \varepsilon$ and by Claim 3.9, $\mathbb{P}[i = 1, j = 1|r = b] = (1/2 + \varepsilon)\,\mathbb{P}[j = 1|i = 1] = y_i\,\mathbb{P}[j = 1|i = 1]$. Therefore, $Y_{ij} = \mathbb{P}[i = 1, j = 1|r = b] = y_i\,\mathbb{P}[j = 1|i = 1] = Y_{ji}$.

■

Let us fix a vertex i, and consider the vectors $\mathbf{z} := \mathbf{Y}_i/y_i$ and $\mathbf{w} := (\mathbf{Y}_i - \mathbf{Y}_0)/y_i$. We will show that they are (convex combinations of) $(r+1)$-canonical vectors. (If $y_i = 0$ we do not need to analyze $\mathbf{z}$, and if $y_i = 1$ we do not need to analyze $\mathbf{w}$.)

Note that $\mathbf{z}$ and $\mathbf{w}$ are same as $\mathbf{y}$ except for vertices that are within distance ℓ of i.

Lemma 3.19 *If $\mathbf{y}$ is an r-canonical solution and Y is the matrix as defined above, then $\forall 1 \le i \le n$, the solutions $\mathbf{z} := Y_i/y_i$ and $\mathbf{w} := (\mathbf{Y}_i - \mathbf{Y}_0)/y_i$ are ε^3-saturated*

Proof: We first give the proof for $\mathbf{z}$. Note that for $d(i, j) > \ell$ $z_j = y_j$ and hence edges as distance greater than ℓ from i are ε^2 saturated because they were in $\mathbf{y}$ by Claim 3.16. If $d(i, r) > 2\ell \; \forall r \in S$ then the distribution up to distance 2ℓ from i is same as a $1 - Splash$, which is in fact ε^2-saturated by Claim 3.7 and Claim 3.8.

Let i be within distance 2ℓ of $r \in S$ and let (j, k) be an edge such that $d(i, j) \le \ell$ or $d(i, k) \le \ell$. If both j and k are within distance ℓ of i, then by Claim 3.10

$$\begin{aligned} Y_{ij} + Y_{ik} &= \mathbb{P}[i = 1, j = 1|r = b] + \mathbb{P}[i = 1, k = 1|r = b] \\ &\ge (1 + 4\varepsilon^3)\,\mathbb{P}[i = 1|r = b] = (1 + 4\varepsilon^3)y_i \end{aligned}$$

and we are done. Finally, if $d(i, j) = \ell$ and $d(i, k) = \ell + 1$, then we know by Claim 3.11 that $|\,\mathbb{P}[i = 1, k = 1|r = b] - \mathbb{P}[i = 1|r = b]\,\mathbb{P}[k = 1|r = b]| \le 2\varepsilon^4$. This gives

$$\begin{aligned} Y_{ij} + Y_{ik} &= \mathbb{P}[i = 1, j = 1|r = b] + \mathbb{P}[i = 1|r = b]\,\mathbb{P}[k = 1|r = b] \\ &\ge \mathbb{P}[i = 1, j = 1|r = b] + \mathbb{P}[i = 1, k = 1|r = b] - 2\varepsilon^4 \\ &\ge (1 + 4\varepsilon^3)\,\mathbb{P}[i = 1|r = b] - 2\varepsilon^4 \ge (1 + 3\varepsilon^3)y_i \end{aligned}$$

using the fact that $\mathbb{P}[i = 1|r = b]$ is at least 2ε. We prove this for $\mathbf{w}$ similarly. ■

We shall now express **z** and **w** as a convex combination of $(r+1)$-canonical vectors.

Claim 3.20 *If $i \in S$, or if $\forall r \in S$, $d(i,r) > 5\ell$, then* **z** *is $r+1$ canonical.*

Proof: If $i \in S$, then $z_k = y_k$ (or $w_k = y_k$) for all $k \in V$ by construction of protection matrix. Because **y** is r-canonical **z** (or **w**) is also and this thus also $(r+1)$-canonical.

If $\forall r \in S$, $d(i,r) > 5\ell$, then it is easily seen that $S \cup \{i\}$ is a witness to **z** and **w** being $(r+1)$-canonical. ■

If neither of these cases is true, we treat only **z**, because the same argument works for **w**. We first define the subset of vertices which is fixed in these vectors.

Recall that for $i \in S$, $N_S(i) = \{j$: there exists path of length at most ℓ from i to j using only edges in $E \backslash E_{|S}$). In addition let $\partial N_S(i) = \{j : d(i,j) = \ell + 1$ in the graph $(V, E \backslash E_{|S})\}$. Also, let $N'_S(i) = N_S(i) \cup \partial N_S(i)$.

Then we make the following definition:

Definition 3.21 *For a fixed vertex i, we construct $F \subseteq V \setminus S$ as follows:*

Start with $F = N'_S(i)$. If there is a path P of length less that 5ℓ between any two vertices in $F \cup S$ that uses only edges in $V \setminus (F \cup S)$, then $F = F \cup P$. Also, if P intersects $N_S(j)$ for $j \in S$, then $F = F \cup P \cup (N'_S(j) \backslash \{j\})$.

Note that it follows from the above definition that for every $j \in S$, either $N_S(j) \cap F = \emptyset$ or $N'_S(j) \subseteq F$. Also if $\partial F = \{j \in F$: *j has neighbors in* $V \backslash (S \cup F)\}$, then $\forall j \in \partial F$, $z_j = 1/2 + \varepsilon$ (because for every intersecting $N_S(j')$, we also included $\partial N_S(j')$). We now bound the size of F.

Claim 3.22 *$|F| \leq C + (2r + 2 - k)D$, where $|S| = rC + kD$.*

Proof: Every path added in the construction of F has length at most 5ℓ. Also, each vertex in a path can be within distance ℓ of at most one $j \in S$. Thus, the number of vertices added due to a path is at most $5\ell C = D$. Thus, if p paths are added during the construction, then $|F| \leq C + pD$ since C is the size of the $N'_S(i)$, which we start with.

Since the paths are added incrementally, it suffices to show that adding $2r+2-k$ paths implies a contradiction. This would imply that $p \leq 2r + 2 - k$ and hence the claim. Let F' be F after addition of $2r + 2 - k$ paths. Then

$$\frac{|E_{|S \cup F'}|}{|S \cup F'|} = 1 + \frac{|E_{|S \cup F'}| - |S \cup F'|}{|S \cup F'|}$$

Note that $|E_{|S \cup F'}| - |S \cup F'| \geq (k - r) + (2r + 2 - k) - 1$, since $|E_{|S}| - |S| \geq (k - r)$ to begin with and addition of $N'_S(i)$, which is a tree adds one more vertex than edge (hence contributing -1), while the addition of each path adds one more edge than vertex. For any $j \in S$ including the region $N_S(j)$

intersected by the path includes a tree of which at least one vertex is already in F and can only contribute positively. This gives

$$\frac{|E_{|S\cup F'}|}{|S\cup F'|} \geq 1+\frac{k-r+2r-k+1}{|S|+|F'|} \geq 1+\frac{r+1}{rC+kD+C+(2r+2-k)D} = 1+\frac{1}{C+2D} > 1+\eta$$

since $\eta = \frac{1}{3D} < \frac{1}{C+2D}$. But this is a contradiction since $|S \cup F'| \leq \gamma n$ and hence $|E_{|S\cup F'}| \leq (1+\eta)|S \cup F'|$. ■

Now, because $r \leq R = \frac{\gamma n}{C+2D}$, $|S \cup F| \leq \gamma n$ and we employ Lemma 3.14 to $T = S \cup F$ using the fact that $\mathbf{z}$ is ε^3-saturated.

We obtain vertex covers on $S \cup F$, $T^1, \ldots, T^m$ such that $\lambda_1 T^1 + \ldots + \lambda_m T^m = \mathbf{z} \mid_T$ where $\sum_{l=1}^{m} \lambda_l = 1$. Note that the values for the vertices in S are 0/1 and are hence unchanged in all these solutions. To extend these solutions to fractional solutions over the whole graph, we look at each vertex j on the boundary of the set F and change the values of vertices upto a distance ℓ from it in $V \setminus (S \cup F)$ according to a splash around j. We first prove that all the vertices upto distance ℓ from the boundary of F have value $1/2+\varepsilon$ in $\mathbf{z}$.

Claim 3.23 *For all $j \in F$, either*

- *all neighbors of j are in $S \cup F$, or*
- *For all $k \in N_{S\cup F}(j)$, $z_k = 1/2+\varepsilon$*

Proof: Assume not, then for some $j \in F$ which has some neighbor not in $S \cup F$, there exists $k \in N_{S\cup F}(j)$ such that $z_k \neq 1/2+\varepsilon$. First, we show that it must be that $z_j = 1/2+\varepsilon$. The only elements of $\mathbf{z}$ which do not have weight $1/2+\varepsilon$ are elements of $N_S(l)$ for $l \in F$ and $N_S(i)$. However, $N'_S(i) \subseteq F \cup S$ so no element of $N_S(i)$ has a neighbor outside of F. Similarly, if $j \in N_S(l)$, then because $j \in F$, it must be that $N'_S(l) \subseteq F \cup S$ and thus j has no neighbors outside $S \cup F$.

So, say that $k \neq j$, then $k \notin S \cup F$. But there exists a path P of length $\leq \ell$ which avoids $S \cup F$ from j to k. Because $\mathbf{y}$ is r-canonical, and $\mathbf{z}$ is the same as $\mathbf{y}$ except possibly at the vertices in $N_S(i)$, it must be that $k \in N_S(i)$ or $k \in N_S(j')$ for some $j' \in S$. But, it cannot be that $k \in N_S(i)$ because $N_S(i) \subseteq F$. Also if $k \in N_S(j')$ for some $j' \in S$, then there is a path from j to j' length at most 2ℓ and so either k must be in $S \cup F$ or $j = j'$. The former cannot be true by assumption. The later cannot be true because $j \in F$ which is disjoint from S. ■

Create $\mathbf{y}^{(l)}$ as follows.

$$y_k^{(l)} = \begin{cases} \mathbb{P}\left[k=1 | j = y_j^{(l)}\right] & k \in N_{S\cup F}(j) \text{ for some } j \in F \\ \quad y_k^{(l)} = z_i & \text{o. w.} \end{cases}$$

First note that this is well defined, because if any vertex were in $N_{S\cup F}(j)$ and $N_{S\cup F}(j')$ for $j, j' \in F$, $j \neq j'$, then there would be path between two vertices in F of length 2ℓ which does not go through $S \cup F$.

We wish to show that $\lambda_1 \mathbf{y}^{(1)}, \ldots, \lambda_m \mathbf{y}^{(m)} = \mathbf{z}$. Consider first some $k \in N_{S\cup F}(j)$ for some $j \in F$. First note that $\lambda_1 \mathbf{y}_j^{(1)} + \ldots + \lambda_m y_j^{(m)} = z_j$. By Claim 3.23 if $k \neq j$, then it must be that $z_j = z_k = 1/2 + \varepsilon$. Therefore by Claim 3.8

$$\lambda_1 y_k^{(1)} + \ldots + \lambda_m y_k^{(m)} = z_j \, \mathbb{P}[k = 1 | j = 1] + (1 - z_j) \, \mathbb{P}[k = 1 | j = 0] = 1/2 + \varepsilon = z_k$$

If $k \notin \cup_{j \in F} N_{S\cup F}(j)$, then $y_k^{(l)} = z_k$ for all k, and so $\lambda_1 y_k^{(1)}, \ldots, \lambda_m y_k^{(m)} = z_k$. We now must show that for each k, $\mathbf{y}^{(k)}$ is an $(r+1)$-canonical solution. We show that $T = S \cup F$ is a witness for $\mathbf{y}^{(k)}$.

Since the solution $T^{(k)}$ given by Lemma 3.14 is a vertex cover $\mathbf{y}_{|T}^{(k)} = T^{(k)}$ is a vertex cover for T. Also, by construction of F, there is no path of length less than 5ℓ between any vertices of $S \cup F$ using only vertices outside $S \cup F$. By Claim 3.22 $|T| = |S| + |F| \leq rC + kD + C + (2r + 2 - k)D = (r+1)C + 2(r+1)D$. If the number of paths added in constructing F is p, then $|T| \leq (r+1)C + (k+p)D$. Also, as argued in Claim 3.22, $|E_{|S\cup F}| - |S \cup F| \geq (k - r) + p - 1 = (k + p) - (r + 1)$.

Finally, we need to show that $y_j^{(k)} = \mathbb{P}[j = 1 | j' = y_{j'}]$ if $j \in N_{S\cup F}(j')$ and $1/2 + \varepsilon$ otherwise. Let $y_j^{(k)} \neq 1/2 + \varepsilon$. Then either $j \in N_{S\cup F}(j')$ for some $j' \in F$ (since these vertices were set according to a splash pattern while creating $\mathbf{y}^{(k)}$) and we are done, or $z_k \neq 1/2 + \varepsilon$. However, $\mathbf{z} = \mathbf{Y}_i / y_i$ differs from $\mathbf{y}$ only in $N_S(i)$. Therefore, $z_k \neq 1/2 + \varepsilon$ in turn implies $j \in N_S(i)$ and hence $j \in F$, or $y_j \neq 1/2 + \varepsilon$. To finish off, we note that $y_j \neq 1/2 + \varepsilon$ would mean $j \in N_S(j')$ for some $j' \in S$ (by assumption on S). Since $N_S(j')$ is either contained in or disjoint with F, we must have $j \in S \cup F$ or $j \in N_{S\cup F}(j')$ respectively.

Since each $\mathbf{y}^{(k)}$ is an $(r+1)$-canonical solution, by our inductive hypothesis $\forall 1 \leq k \leq m$ $\mathbf{y}^{(k)} \in N^{R-r-1}(VC(G))$ and hence $\mathbf{z} \in N^{R-r-1}(VC(G))$. Using a similar argument for show $\mathbf{w}$, we get that $\mathbf{y} \in N^{R-r}(VC(G))$. This completes the proof of Lemma 3.17. ■

3.6 Lower bounds for MAX-CUT

Let $G = (V, E)$ be a graph with n vertices and m edges. We prove a $1/2 + \zeta$ integrality gap for $\Omega(n)$ rounds of LS on MAX-CUT.

The solutions we define for MAX-CUT are simple extensions of vertex cover solutions. For a vector $\mathbf{y} \in \mathbb{R}^{n+1}$, we define an extension $Ext(\mathbf{y})$ as the vector $\tilde{\mathbf{y}} \in \mathbb{R}^{n+m+1}$ such that, $u_i = y_i \;\; \forall 0 \leq i \leq n$ and $u_e = 2y_0 - y_i - y_j$ for $e = (i, j) \in E$. Also, we define $Res(\tilde{\mathbf{y}})$ as the inverse operation i.e. the projection of the first $n+1$ coordinates of $\tilde{\mathbf{y}}$. It is easy to verify that if $\mathbf{y} \in VC(G)$ then $Ext(\mathbf{y}) \in MC(G)$. Notice that with $R = \frac{\gamma n}{C+D}$ as defined in the previous section, it is sufficient to prove the following

Lemma 3.24 *If $\mathbf{y} \in \mathbb{R}^{n+1}$ is a $2r$-canonical solution for $VC(G)$, then $Ext(\mathbf{y}) \in N^{R/2-r}(MC(G))$.*

The integrality gap follows because $\mathbf{y} = (1, 1/2 + \varepsilon, \ldots, 1/2 + \varepsilon)$ is 0-canonical and for $\tilde{\mathbf{y}} = Ext(\mathbf{y})$, $\sum_{e \in E} u_e = (1 - 2\varepsilon)m$.

Proof: We proceed by induction on $R/2 - r$. The base case follows because if $\mathbf{y}$ is an R-canonical solution, then $\mathbf{y} \in VC(G)$ which implies $Ext(\mathbf{y}) \in MC(G) = N^0(MC(G))$. For the inductive step, let $\mathbf{y}$ be an $2r$-canonical solution and let $\tilde{\mathbf{y}} = Ext(\mathbf{y})$. We create a protection matrix U, such that $\forall 1 \le i \le n$ and $\forall e \in E$, $Res(\tilde{\mathbf{Y}}_i), Res(\tilde{\mathbf{Y}}_e), Res(\mathbf{U}_0 - \tilde{\mathbf{Y}}_i)$ and $Res(\tilde{\mathbf{Y}}_0 - \tilde{\mathbf{Y}}_e)$ can be expressed as convex combinations of $(2r+2)$-canonical solutions. This suffices because for a vector $\tilde{\mathbf{y}}$ if $Res(\tilde{\mathbf{y}}) = \lambda_1 \tilde{\mathbf{y}}^{(1)} + \ldots + \lambda_m \tilde{\mathbf{y}}^{(m)}$ then $\tilde{\mathbf{y}} = Ext(\lambda_1 \tilde{\mathbf{y}}^{(1)}) + \ldots + Ext(\lambda_m \tilde{\mathbf{y}}^{(m)})$, since the coordinates of $Ext(\mathbf{v})$ are affine functions of the coordinates of $\mathbf{v}$.

Let Y be the protection matrix of a $2r$-canonical solution as defined in the previous section. We define the matrix $\tilde{Y}$ as

$$\begin{aligned} \tilde{\mathbf{Y}}_i &= Ext(\mathbf{Y}_i) && \forall 0 \le i \le n \\ \tilde{\mathbf{Y}}_e &= Ext(2\mathbf{Y}_0 - (\mathbf{Y}_i + \mathbf{Y}_j)) && \forall e = (i,j) \in E \end{aligned}$$

We can write out the entries of $\tilde{Y}$ as follows, showing that it is symmetric.

$$\begin{aligned} \tilde{Y}_{i,j} &= Y_{ij} && 0 \le i,j \le n \\ \tilde{Y}_{i,e} &= \tilde{Y}_{e,i} = 2Y_{i0} - Y_{ij} - Y_{ik} && 0 \le i \le n,\ e = (j,k) \in E \\ \tilde{Y}_{e_1,e_2} &= 4Y_{00} - 2(Y_{i0} + Y_{j0} + Y_{k0} + Y_{l0}) + (Y_{ik} + Y_{jk} + Y_{il} + Y_{jl}) && e_1 = (i,j), e_2 = (k,l) \in E \end{aligned}$$

Note that for $i \in V$ and $e = (j,k) \in E$, $Res(\tilde{\mathbf{Y}}_i) = \mathbf{Y}_i, Res(\tilde{\mathbf{Y}}_0 - \tilde{\mathbf{Y}}_i) = \mathbf{Y}_0 - \mathbf{Y}_i$ and $Res(\tilde{\mathbf{Y}}_e) = \mathbf{Y}_0 - \mathbf{Y}_j + \mathbf{Y}_0 - \mathbf{Y}_k$, which are convex combinations of $(2r+1)$-canonical solutions as proved in the previous section. It only remains to tackle $Res(\tilde{\mathbf{Y}}_0 - \mathbf{U}_e) = \mathbf{Y}_j + \mathbf{Y}_k - \mathbf{Y}_0$. We first prove that it is ε^3-saturated.

Claim 3.25 *If Y is the protection matrix of a $2r$-canonical solution and $(i,j),(u,v)$ are two edges, then*

$$\frac{(\mathbf{Y}_i + \mathbf{Y}_j - \mathbf{Y}_0)_u}{y_i + y_j - y_0} + \frac{(\mathbf{Y}_i + \mathbf{Y}_j - \mathbf{Y}_0)_v}{y_i + y_j - y_0} \ge 1 + 4\varepsilon^3$$

Proof: Without loss of generality, we can assume that j and u are the closer endpoints of the edges (i,j) and (u,v). We first handle the case when $d(j,u) > \ell$. Then $Y_{iu} = y_i y_u, Y_{iv} = y_i y_v, Y_{ju} = y_j y_u$ and $Y_{jv} = y_j y_v$. Hence, the LHS is $y_u + y_v$, which is greater than $1 + 2\varepsilon^2$ since a $2r$-canonical solution is ε^2 saturated.

When $d(j,u) \le \ell$, all the four vertices are within distance $\ell + 2$ of each other. Now, in any subgraph H of diameter 3ℓ, we may think of the restriction of $\mathbf{y}$ to H as the probabilities of the vertices being present in a distribution over vertex covers of H. Notice that if $\mathbf{y}$ is a $2r$-canonical solution, H may contain vertices close to (within distance ℓ of) at most one fixed vertex. In case there is such a vertex r, $\forall i \in H$ $y_i = \mathbb{P}[i = 1 | r = 1]$. If there is no such vertex, all vertices in H have $y_i = 1/2 + \varepsilon$ and we can these as probabilities for a distribution which chooses a 1-splash with probability $1/2 + \varepsilon$ and 0-splash with probability $1/2 - \varepsilon$ around any arbitrary vertex in H (Claim 3.8). Also, we can interpret Y_{pq} as $\mathbb{P}[p = 1, q = 1]$ for the same distribution as above.

Consider the distribution over the subgraph within a radius $\ell + 2$ from i. We first note that since $(\mathbf{Y}_0 - \mathbf{Y}_i)/(1 - y_i)$ is a valid vertex cover solution and $(\mathbf{Y}_0 - \mathbf{Y}_i)_i = 0$, $(\mathbf{Y}_0 - \mathbf{Y}_i)_j/(1 - y_i) = 1$ which gives $y_i + y_j - 1 = Y_{ij}$. Using this and the fact that $\mathbb{P}[(i = 1) \vee (j = 1)|u = 1] = 1$, we have

$$\begin{aligned}\frac{(\mathbf{Y}_i + \mathbf{Y}_j - \mathbf{Y}_0)_u}{y_i + y_j - y_0} &= \frac{y_u(\mathbb{P}[i = 1|u = 1] + \mathbb{P}[j = 1|u = 1] - 1)}{\mathbb{P}[i = 1, j = 1]} \\ &= \frac{y_u \mathbb{P}[(i = 1) \wedge (j = 1)|u = 1]}{\mathbb{P}[i = 1, j = 1]} \\ &= \mathbb{P}[u = 1|i = 1, j = 1]\end{aligned}$$

Therefore, we get

$$\begin{aligned}\frac{(\mathbf{Y}_i + \mathbf{Y}_j - \mathbf{Y}_0)_u}{y_i + y_j - y_0} + \frac{(\mathbf{Y}_i + \mathbf{Y}_j - \mathbf{Y}_0)_v}{y_i + y_j - y_0} - 1 &= \mathbb{P}[u = 1|i = 1, j = 1] + \mathbb{P}[v = 1|i = 1, j = 1] - 1 \\ &= \mathbb{P}[(u = 1) \wedge (v = 1)|i = 1, j = 1] \\ &= \mathbb{P}[(u = 1) \wedge (v = 1)|j = 1]\end{aligned}$$

The last equality following from the fact that it is sufficient to condition on the closer of the two vertices i and j. Also,

$$\begin{aligned}\mathbb{P}[(u = 1) \wedge (v = 1)|j = 1] &= \mathbb{P}[u = 1|j = 1] + \mathbb{P}[v = 1|j = 1] - 1 \\ &= \frac{Y_{uj}}{y_j} + \frac{Y_{vj}}{y_j} - 1 \\ &\geq 4\varepsilon^3 \quad \text{(by Lemma 3.19)}\end{aligned}$$

■

We now want to express $\mathbf{w} = (\mathbf{Y}_i + \mathbf{Y}_j - \mathbf{Y}_0)/(y_i + y_j - 1)$ as a convex combination of $(2r + 2)$-canonical solutions. Let S be the witness to $\mathbf{y}$ being $2r$-canonical. We now find a set $T \supseteq S$ such that $\mathbf{w}$ is a convex combination of solutions $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(m)}$ which take 0/1 values over T and which are $(2r + 2)$-canonical, with T being the witness. There are two cases:

Case 1: $i \notin S$ *and* $\exists r \in S$ *s.t.* $d(i, r) \leq 5\ell$ (with $d(i, r)$ being length of the shortest path not passing through S)
By the proof in the previous section, we know that the vector $z = \mathbf{Y}_i/y_i$ is a convex combination of $(2r + 1)$-canonical solutions with a set S_1 being the witness for all of them. Also, $j \in S_1$ as it includes every vertex within distance ℓ of i. We take $T = S_1$.

Case 2: $i \in S$ *or* $d(i, r) > 5\ell \; \forall r \in S$
In this case $\mathbf{z} = \mathbf{Y}_i/y_i$ is $(2r + 1)$-canonical with $S \cup \{i\}$ the witness. We now look at the protection matrix Z for $\mathbf{z}$ and consider the vector $\mathbf{z}' = \mathbf{Z_j}/z_j$. This is a convex combination of $(2r+2)$-canonical solutions having a common witness S_2 which contains $S \cup \{i\}$. Take $T = S_2$.

In both cases $|T| \leq (2r + 2)C + (4r + 4)D$. We now employ Lemma 3.14 to T to obtain vertex covers $T^1, \ldots, T^m$ on T such that $\lambda_1 T^1 + \ldots + \lambda_m T^m = \mathbf{w}_{|T}$ with $\sum_{l=1}^{m} \lambda_l = 1$. We can extend them to create $(2r + 2)$-canonical solutions $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(m)}$ as in the previous section. By the arguments in the previous section, all these have T as the witness. This completes the proof. ■

Chapter 4

Lasserre Gaps for Constraint Satisfaction Problems

An instance of the MAX k-CSP problem consists of a set of constraints over n variables. Each constraint is a boolean function depending on the values of at most k variables and is said to be satisfied when the value of the function is 1. The objective is to find an assignment to the variables from a specified domain, which satisfies as many constraints as possible. The most common setting is the one where the domain of the variables is $\{0, 1\}$ or $\{\mathsf{true}, \mathsf{false}\}$ and the constraints are disjunctions of variables or their negations. If each constraint involves at most k variables, this gives the MAX k-SAT problem. Other examples involve Max-Cut, MAX All-Equal etc. In the case when the variables take values in some larger domain $[q]$ for $q > 2$, we denote the problem as MAX k-CSP$_q$.

Algorithms and hardness results for MAX k-CSP

It is obvious from the the fact that MAX k-SAT is a special case of MAX k-CSP, that the general MAX k-CSP problem is NP-hard to solve exactly. We will say that an algorithm achieves an (ρ)-approximation if the algorithm produces an assignment satisfying at least v/ρ of the constraints, when the best assignment satisfies a v fraction of all the constraints. Note that in keeenping with our notation of the integrality gaps, we also always define the approximation ratio to be always greater than 1. For the case when the domain of the variables is $\{0, 1\}$ the most trivial approximation is to assign all variables randomly. Since every constraint involves at most k variables and has at least one satisfying assignment, the algorithm satisfies $1/2^k$ fraction of all the constraints in expectation. Since the optimum fraction v is at most 1, this is a 2^k-approximation.

The first non-trivial approximation for the problem in the case of binary domain was obtained by Trevisan [Tre98] who obtained a $2^k/2$-approximation algorithm using linear programming. Using semidefinite programs, Hast[Has05] achieved an approximation ratio of $O((\log k \cdot 2^k)/k)$. The best known algorithm is due to Charikar, Makarychev and Makarychev [CMM07b], which achieves an approximation ratio of $2^k/ck$, where $c > 0.44$ is an absolute constant. For the case when each variable has domain size q, which we denote by MAX k-CSP$_q$, their algorithm can be extended to give an approximation ratio of q^k/ck. Independently, Guruswami and Raghavendra [GR08] obtained a $q^k/C(q)k$-approximation algorithm, where $C(q)$ is a small constant which decreases with q.

It is also known that it is NP-hard to approximate the maximum fraction of constraints satisfiable, within some factor ρ depending on k and the domain size of each variable. For the setting of boolean variables, Samorodnitsky and Trevisan [ST00] showed that it is hard to approximate MAX k-CSP within a factor better than $2^k/2^{2\sqrt{k}}$, which was later improved to $2^k/2^{\sqrt{2k}}$ by Engebretsen and Holmerin [EH05]. Assuming the Unique Games Conjecture (UGC) of Khot [Kho02], Samorodnitsky and Trevisan [ST06] later showed that MAX k-CSP is hard to approximate within a factor of $2^k/2^{\lceil \log(k+1) \rceil} \geq 2^k/2k$. For MAX k-CSP$_q$, Guruswami and Raghavendra [GR08] extended the techniques of [ST06] to obtain a hardness ratio of q^k/q^2k when q is a prime. The best known results are due to Austrin and Mossel [AM08] who obtained a ratio of $q^k/q(q-1)k$ when q is a prime power, and $q^k/q^{\lceil \log_2 k+1 \rceil}$ for general q. They also improved the Samorodnitsky-Trevisan lower bound to $2^k/(k + O(k^{0.525}))$ from $2k/2^k$ in the worst case. The results of [GR08] and [AM08] are also proved assuming the UGC.

Also under the UGC, Håstad [Hås07] showed a lower bound for generic problems of the type MAX k-CSP (P), where each constraint is of the form of a fixed predicate $P : \{0,1\}^k \to \{0,1\}$ applied to k variables or their negations. He shows that if we pick a predicate P at random, then it is NP-hard to approximate MAX k-CSP (P) within a factor better than $2^k/|P^{-1}|$ with high probability over the choice of P ($2^k/|P^{-1}|$ can be achieved by just randomly assigning each variable). Here $P^{-1}(1)$ denotes the set of inputs that P accepts.

Lower bounds for semidefinite hierarchies

Arora, Alekhnovich and Tourlakis showed an integrality gap of $2^k/(2^k-1)$ for MAX k-SAT after $\Omega(n)$ levels of the Lovász-Schrijver hierarchy. Also [STT07a] reasoned about the inapproximability of 3-XOR although they were interested in proving results for Vertex Cover. Recently, a beautiful result of Schoenebeck [Sch08] gave the first construction of integrality gaps in the Lasserre hierarchy. He proves an integrality gap of 2 for MAX k-XOR and this implies gaps for other CSP problems which can be easily "reduced" to k-XOR (we discuss more on these reductions later).

Our contribution

We show that the integrality gap for MAX k-CSP is at least $2^k/2^{\lceil \log k+1 \rceil}$ even after $\Omega(n)$ levels of the Lasserre hierarchy. This is optimal in terms of the number of levels since the Lasserre relaxation is tight after n levels. Also, this matches the UGC hardness results of Samorodnitsky and Trevisan. For MAX k-CSP$_q$, we handle the case of prime q. Here we can prove optimal integrality gaps equal to $q^k/(kq(q-1))$ which match the lower bounds of Austrin and Mossel. The lower bounds also match the results obtained by the algorithm of Charikar, Makarychev and Makarychev upto a constant independent of k (but depending on q).

As an added bonus, we get integrality gaps for *random predicates*. This is an artifact of the property that the Lasserre hierarchy is very amenable to "reductions by inclusion" i.e. if one proves a gap for a predicate P_1, and if P_1 (or some simple transformation of it) implies P_2 (i.e. every satisfying assignment of P_1 satisfies P_2), then it is very easy to also prove a gap for P_2. For example, Schoenebeck [Sch08] shows a gap for MAX k-XOR, but k-XOR (or its negation if k is even) implies k-SAT and hence his result immediately implies a result for MAX k-SAT as well. By a theorem of Håstad, for almost every randomly chosen predicate P, there is a simple transformation of the predicates we consider, that implies P. Hence, we immediately get integrality gaps for almost every randomly chosen predicate.

This is the first instance of integrality gaps which are upto $\Omega(n)$ levels, ruling out exponential time algorithms in the hierarchy, and at the same time are better than the best known NP-hardness results. The hardness results corresponding to the above integrality gaps, as well as the random predicates result, are known only when assuming the UGC. The only other instance (for superconstant number of levels) is that of Vertex Cover, where an integrality gap of $2 - \varepsilon$ was proved by Georgiou et al. [GMPT07], but it holds only for $\Omega(\sqrt{\log n / \log \log n})$ levels of the Lovász-Schrijver (SDP) hierarchy which is weaker than the Lasserre hierarchy.

On the technical side, we extend the technique of constructing vectors from linear equations used in [FO06], [STT07a], [Sch08] in two ways. The technique was used in [FO06] to show a gap for the Lovász θ-function for an instance created by reduction of 3-XOR to Independent Set. They looked at how two variables were correlated by the clauses of the XOR formula and created coordinates in the vectors according to these correlations. [STT07a] proved gaps for such instances for $\Omega(n)$ levels of LS+ and looked at more variables at every round, but this fact was not explicit in the proof. Schoenebeck [Sch08] made it explicit and used it to prove strong gaps in the Lasserre hierarchy.

We generalize the technique to deal with equations over $\mathbb{F}_q$ for any prime q, instead of just $\mathbb{F}_2$. This actually follows just by a simple trick of constructing the vectors out of small "building blocks", each of which is a $q - 1$ dimensional vector corresponding to a vertex of the $q - 1$ dimensional simplex. More importantly, we "decouple" the equations from the constraints to some extent. Previous works worked by combining constraints as a whole to study correlations and worked in the setting when each constraint was a single equation. The independence of equations in a random instance was used to show the gaps. However, as we explain in Section 4.3, it turns out to be more useful to simply think of vectors corresponding to a system of linear equations which satisfy certain properties, as it allows more freedom in the choice of the constraints.

4.1 Preliminaries and notation

4.1.1 Constraint Satisfaction Problems

For an instance Φ of MAX k-CSP$_q$, we denote the variables by $\{x_1, \ldots, x_n\}$, their domain $\{0, \ldots, q-1\}$ by $[q]$ and the constraints by $C_1, \ldots, C_m$. 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 an ordered tuple[1] T_i with $|T_i| \leq k$. We denote the number of constraints satisfied by the best assignment by OPT(Φ).

For a given set $S \subseteq [n]$, we denote by $[q]^S$ the set of all mappings from the set S to $[q]$. In context of variables, these mappings can be understood as partial assignments to a given subset of

[1]We will often ignore the order of the variables in T_i and also refer to T_i as a set of variables.

variables. For $\alpha \in [q]^S$, we denote its projection to $S' \subseteq S$ as $\alpha(S')$. Also, for $\alpha_1 \in [q]^{S_1}, \alpha_2 \in [q]^{S_2}$ such that $\alpha_1(S_1 \cap S_2) = \alpha_2(S_1 \cap S_2)$, we denote by $\alpha_1 \circ \alpha_2$ the assignment over $S_1 \cup S_2$ defined by α_1 and α_2. Hence, $(\alpha_1 \circ \alpha_2)(j)$ equals $\alpha_1(j)$ for $j \in S_1$ and $\alpha_2(j)$ for $j \in S_2 \setminus S_1$. We only use the notation $\alpha_1 \circ \alpha_2$ when it is well defined for $\alpha_1, \alpha_2, S_1, S_2$.

We shall prove results for constraint satisfaction problems where every constraint is specified by the same predicate $P : [q]^k \to \{0,1\}$. We denote the set of inputs which the predicate accepts (outputs 1 on) by $P^{-1}(1)$. To generate an instance of the problem each constraint is of the form of P applied to a k-tuple of *literals*. For variable x over domain $[q]$, we can generalize the notion of a literal as $x + a$ (computed modulo q) for $a \in [q]$.

Definition 4.1 *For a given $P : [q]^k \to \{0,1\}$, an instance Φ of* MAX k-CSP *(P) is a set of constraints $C_1, \ldots, C_m$ where each constraint C_i is over a k-tuple of variables $T_i = \{x_{i_1}, \ldots, x_{i_k}\}$ and is of the form $P(x_{i_1} + a_{i_1}, \ldots, x_{i_k} + a_{i_k})$ for some $a_{i_1}, \ldots, a_{i_k} \in [q]$.*

Given a predicate P, we will consider a random instance Φ of the MAX k-CSP (P) problem. To generate a random instance with m constraints, for every constraint C_i, we randomly select a k-tuple of distinct variables $T_i = \{x_{i_1}, \ldots, x_{i_k}\}$ and $a_{i_1}, \ldots, a_{i_k} \in [q]$, and put $C_i \equiv P(x_{i_1} + a_{i_1}, \ldots, x_{i_k} + a_{i_k})$. It is well known and used in various works on integrality gaps and proof complexity (e.g. [BOGH+03], [AAT05], [STT07a] and [Sch08]), that random instances of CSPs are highly unsatisfiable and at the same time highly expanding i.e. for every set of constraints which is not too large, most variables occur only in one constraint. We capture the properties we need in the lemma below. A proof is provided in the appendix for the sake of completeness.

Lemma 4.2 *Let $\varepsilon, \delta > 0$ and a predicate $P : [q]^k \to \{0,1\}$ be given. Then there exist $\beta = O(q^k \log q/\varepsilon^2)$, $\eta = \Omega((1/\beta)^{5/\delta})$ and $N \in \mathbb{N}$, such that if $n \geq N$ and Φ is a random instance of* MAX k-CSP *(P) with $m = \beta n$ constraints, then with probability $1 - o(1)$*

1. $\mathsf{OPT}(\Phi) \leq \frac{|P^{-1}(1)|}{q^k}(1+\varepsilon) \cdot m.$
2. *For all $s \leq \eta n$, every set of s constraints involves at least $(k-1-\delta)s$ variables.*

The instances we will mostly be concerned with, are described by systems of equations over finite fields. Let q be prime and $A \in (\mathbb{F}_q)^{d \times k}$ be a matrix with $rank(A) = d \leq k$. We define the predicate $P_A : [q]^k \to \{0,1\}$ such that

$$P_A(x_1, \ldots, x_k) = 1 \quad \Leftrightarrow \quad A \cdot (x_1 \ldots x_k)^\mathsf{T} = 0$$

To generate a constraint C_i in an instance of MAX k-CSP (P_A), we consider $P_A(x_{i_1} + a_{i_1}, \ldots, x_{i_k} + a_{i_k})$ which is 1 iff $A \cdot (x_{i_1} \ldots x_{i_k})^\mathsf{T} = b^{(i)}$ for $b^{(i)} = -A \cdot (a_{i_1} \ldots a_{i_k})^\mathsf{T}$. We define the problem MAX k-CSP (P_A) below which is a special case of the MAX k-CSP (P) problem.

Definition 4.3 *For a given $A \in (\mathbb{F}_q)^{d \times k}$, an instance Φ of* MAX k-CSP *(P_A) is a set of constraints $C_1, \ldots, C_m$ where each constraint C_i is over a k-tuple $T_i = \{x_{i_1}, \ldots, x_{i_k}\}$ and is of the form $A \cdot (x_{i_1}, \ldots, x_{i_k})^\mathsf{T} = b^{(i)}$ for some $b^{(i)} \in (\mathbb{F}_q)^d$.*

4.1.2 Linear Codes

Recall that a linear code of distance 3 and length k over $\mathbb{F}_q$ is a subspace of $(\mathbb{F}_q)^k$ such that every non-zero vector in the subspace has at least 3 non-zero coordinates. We shall prove our results for predicates P_A where A is a matrix whose rows form a basis for such a code. Such a matrix is called the generator matrix of the code. To get the optimal bounds, we shall use Hamming codes which have the largest dimension for the fixed distance 3. We refer to the code below as Hamming code of length k.

Fact 4.4 *Let* $(q^{r-1} - 1)/(q - 1) < k \le (q^r - 1)/(q - 1)$*. Then there exists a linear code of length* k *and distance* 3 *over* $\mathbb{F}_q$*, with dimension* $k - r$*.*

Proof: Let $l = (q^r - 1)/(q - 1)$. We can first construct a code of length equal to l and dimension $l - r$ by specifying the $r \times l$ check matrix (a matrix whose rows span the space orthogonal to the code). The requirement that the code must have distance 3 means that no two columns of the check matrix should be linearly dependent. We can choose (for example) all non-zero vectors in $(\mathbb{F}_q)^r$ having their first nonzero element as 1 to get a matrix with r rows having this property. It is easy to check that there are $l = (q^r - 1)/(q - 1)$ such columns. To reduce the code length, we delete the last $l - k$ columns of the matrix to get the check matrix of a code with length k, distance 3 and dimension $k - r$. ■

4.1.3 The Lasserre relaxation for MAX k-CSP$_q$

We describe below the Lasserre relaxation for the MAX k-CSP$_q$ problem. Note that an integer solution to the problem will be given by a single mapping $\alpha_0 \in [q]^{[n]}$, which is an assignment to all the variables. Using this, we can define 0/1 variables $X_{(S,\alpha)}$ for each $S \subseteq [n]$ such that $|S| \le t$ and $\alpha \in [q]^S$. The intended solution is $X_{(S,\alpha)} = 1$ if $\alpha_0(S) = \alpha$ and 0 otherwise. As before, we introduce $X_{(\emptyset,\emptyset)}$ which is intended to be 1. Replacing them by vectors gives the SDP relaxation. We denote the vectors corresponding to a set of variables and a partial assignment by $\mathbf{V}_{(S,\alpha)}$.

$$
\begin{array}{lll}
\text{maximize} & \displaystyle\sum_{i=1}^{m} \sum_{\alpha \in [q]^{T_i}} C_i(\alpha) \left|\mathbf{V}_{(T_i,\alpha)}\right|^2 & \\
\text{subject to} & \langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)} \rangle = 0 & \forall\ \alpha_1(S_1 \cap S_2) \neq \alpha_2(S_1 \cap S_2) \\
& \langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)} \rangle = \langle \mathbf{V}_{(S_3,\alpha_3)}, \mathbf{V}_{(S_4,\alpha_4)} \rangle & \forall\ S_1 \cup S_2 = S_3 \cup S_4,\ \alpha_1 \circ \alpha_2 = \alpha_3 \circ \alpha_4 \\
& \displaystyle\sum_{j \in [q]} \left|\mathbf{V}_{(\{i\},j)}\right|^2 = 1 & \forall i \in [n] \\
& \langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)} \rangle \ge 0 & \forall S_1, S_2, \alpha_1, \alpha_2 \\
& \left|\mathbf{V}_{(\emptyset,\emptyset)}\right| = 1 &
\end{array}
$$

Figure 4.1: Lasserre SDP for MAX k-CSP$_q$

For any set S with $|S| \le t$, the vectors $\mathbf{V}_{(S,\alpha)}$ induce a probability distribution over $[q]^S$ such that the assignment $\alpha \in [q]^S$ appears with probability $\left|\mathbf{V}_{(S,\alpha)}\right|^2$. The constraints can be understood by thinking of valid solution as coming from a distribution of assignments for all the variable and of $\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)} \rangle$ as the probability of the event that variables in S_1 get value according to α_1 and those in S_2 according to α_2. The last two constraints simply state the fact that a probability must be positive and the probability of the event which does not restrict any variable to anything must be 1.

The first constraint says that the probability of a variable simultaneously getting two different values must be 0. The second one says that if we calculate the probability that all variables in $(S_1 \cup S_2)$ $(= S_3 \cup S_4)$ get values according to $\alpha_1 \circ \alpha_2$ $(= \alpha_3 \circ \alpha_4)$ in two different ways, it must still be the same. In the third constraint, $\mathbf{V}_{(\{i\},j)}$ denotes the map which assigns value to j to variable x_i. The constraint can be understood as saying that the variable x_i must take exactly one value. This constraint is sometimes also written as $\sum_{i\in[q]} \mathbf{V}_{(\{i\},j)} = \mathbf{V}_{(\emptyset,\emptyset)}$, which is equivalent as can be seen by noting that $\left|\sum_{i\in[q]} \mathbf{V}_{(\{i\},j)} - \mathbf{V}_{(\emptyset,\emptyset)}\right|^2 = \sum_{j\in[q]} \left|\mathbf{V}_{(\{i\},j)}\right|^2 - 1$.

For a CSP instance Φ we denote the optimum of the above SDP by $\mathsf{FRAC}(\Phi)$. The integrality gap of both the SDPs is then given by (SDP optimum)/(integer optimum).

4.2 Results

The main result is an integrality gap for predicates $P_A : [q]^k \to \{0, 1\}$, whenever q is a prime and A is the matrix of a distance 3 code over $\mathbb{F}_q$. The result for general q-ary predicates then simply follows by optimizing parameters.

Theorem 4.5 *Let $A \in (\mathbb{F}_q)^{d\times k}$ be the generator matrix of a distance 3 code and let $\zeta > 0$ be given. Then there is a constant $c = c(q, k, \zeta)$ such that for large enough n, the integrality gap for the Lasserre SDP relaxation of* MAX k-CSP *(P_A) on n variables obtained by cn levels is at least $q^d - \zeta$.*

Corollary 4.6 *Let $k \geq 3$, q be a prime number and let $\zeta > 0$ be given. Then there exists $c = c(k, q, \zeta) > 0$ such that the for sufficiently large n, the integrality gap for the Lasserre relaxation of the* MAX k-CSP *problem on n variables with domain size q is at least $\dfrac{q^k}{kq(q-1) - q(q-2)} - \zeta$ after cn levels.*

Using the results of Håstad [Hås07], we also get integrality gaps for random predicates. Let $Q(p, q, k)$ denote the distribution over predicates, where we choose $P(x)$ to be 1 with probability p, independently for each $x \in [q]^k$. Then, we have the following result for random predicates:

Theorem 4.7 *Let a prime q and $\zeta > 0$ and $k \geq 3$ be given and let $(q^{r-1} - 1)/(q - 1) < k \leq (q^r - 1)/(q - 1)$ for some r. Then there exist constants $c = kq^{-r}(1 - o(1))$ and $c' = c'(q, k, \zeta)$ such that if P is a random predicate chosen according to $Q(p, q, k)$ with $p \geq k^{-c}$, then with probability $1 - o(1)$ over the choice of P, the integrality gap for* MAX k-CSP *(P) after $c'n$ levels of the Lasserre hierarchy is at least $q^k/|P^{-1}(1)| - \zeta$.*

4.3 Proof Overview

For proving integrality gaps for CSPs, our starting point is the result by Schoenebeck [Sch08]. Even though the result is stated as an integrality gap for MAX-k-XOR, it is useful to view the argument as having the following two parts:

1. Given a system of equations over $\mathbb{F}_2$ with no "small contradictions", it shows how to construct vectors $\mathbf{V}_{(S,\alpha)}$ (satisfying consistency conditions), such that for every set S of variables with $|S| \leq t$, $\left|\mathbf{V}_{(S,\alpha)}\right| > 0$ if and only if α satisfies all equations involving variables from S.

2. If one considers a random instance Φ of a CSP, with every constraint being a linear equation on at most k variables (MAX k-XOR), then the system obtained by combining them has no "small contradictions".

It can be shown that the first step implies $\mathsf{FRAC}(\Phi) = m$ (m is the number of constraints) even after t levels. On the other hand, in a system of equations with each chosen randomly, only about half are satisfiable by any assignment. Hence, one gets an integrality gap of factor 2, for t levels. Here t depends on the size of "small contradictions" for step 2 and can be chosen to be $\Omega(n)$.

We note that every constraint does not *have* to be a single equation for the second step to work. In particular, we consider each constraint to be of the form $A \cdot (x_1, \ldots, x_k)^\mathsf{T} = b$ for some $A \in (\mathbb{F}_2)^{d \times k}$, which is the same for all constraints and $b \in \mathbb{F}_2^d$. The constraint is said to be satisfied only when all these d equations are satisfied. Now, if A is full rank, then one can show that in a random CSP instance Φ, only about $1/2^d$ fraction of the constraints are satisfiable by any assignment.

Note that all equations obtained by combining all the ones in the constraints are no longer independent (the ones in each constraint are correlated because of a fixed choice of A). However, one can still show that if A satisfies some extra properties, like any linear combination of the d equations given by A involves at least 3 variables (i.e. A is the generator matrix of a distance 3 code over $\mathbb{F}_2$), then the conclusion in the second step can still be made. Step 1 still allows us to conclude that $\mathsf{FRAC}(\Phi) = m$, thereby obtaining an integrality gap of factor 2^d. Optimizing over d gives the claimed results for MAX k-CSP. We also generalize the first step, to work for equations over arbitrary prime fields $\mathbb{F}_q$, to obtain a gap for MAX k-CSP_q.

4.4 Lasserre vectors for linear equations

We first prove a generalization of Schoenebeck's result [Sch08] that we shall need. As mentioned before, it is more convenient to view it as constructing vectors for a system of linear equations. We show that if the width-t bounded resolution (defined below) of the system of equations cannot derive a contradiction then there exists vectors for sets of size up to $t/2$ which satisfy all consistency constraints in the Lasserre relaxation. Also these vectors "satisfy" all the equations in the sense that if we think of $\left|\mathbf{V}_{(S,\alpha)}\right|^2$ as the probability that the variables in set S get the assignment α, then for any set S, the only assignments having non-zero probability are the ones which satisfy all the equations involving variables from S.

We write a linear equation in variables $\{x_1, \ldots, x_n\}$ as $\omega \cdot x = r$ where ω is a vector of coefficients and x is the vector containing all variables. We also denote by $\overrightarrow{0}$ the coefficient vector with all coefficients equal to 0. For a system Λ of linear equations over $\mathbb{F}_q$, we formally define resolution as below

Definition 4.8 *Let Λ be a system of linear equations over a prime field $\mathbb{F}_q$. Then $\mathsf{Res}(\Lambda, t)$ is defined as the set of all equations involving at most t variables, each of which can be derived by a linear combination of at most two equations from Λ.*

Also, for a set $S \subseteq [n]$, let A_S denote the set of all partial assignments to variables in S, which satisfy all equations in Λ involving only variables from S. We then have the following theorem

Theorem 4.9 *Let q be a prime. Suppose Λ is a system of linear equations in $\mathbb{F}_q$ such that $(\overrightarrow{0} \cdot x = r) \in \Lambda \Leftrightarrow r = 0$ and $\mathsf{Res}(\Lambda, 2t) = \Lambda$. Then there are vectors $\mathbf{V}_{(S,\alpha)}$, for all S with $|S| < t$ and for all $\alpha \in [q]^S$, such that*

1. $\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)} \rangle \geq 0$ *for all* $S_1, S_2, \alpha_1, \alpha_2$.
2. $\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)} \rangle = 0$ *if* $\alpha_1(S_1 \cap S_2) \neq \alpha_2(S_1 \cap S_2)$.
3. *If* $\alpha_i \in [q]^{S_i}$, $1 \leq i \leq 4$ *are such that* $\alpha_1 \circ \alpha_2$ *and* $\alpha_3 \circ \alpha_4$ *are both defined and equal, then*
$$\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)} \rangle = \langle \mathbf{V}_{(S_3,\alpha_3)}, \mathbf{V}_{(S_4,\alpha_4)} \rangle$$
4. $\mathbf{V}_{(S,\alpha)} = 0$ *for* $\alpha \notin A_S$ *and* $\sum_{\alpha \in A_S} \left|\mathbf{V}_{(S,\alpha)}\right|^2 = 1$

Note that the theorem is stated in terms of a system Λ, which is closed under the applications of the operator $\mathsf{Res}(\cdot, 2t)$. Our proof of the theorem essentially follows Schoenebeck's proof except for one modification, which allows the generalization to the q-ary case.

We shall require some additional notation for the proof. For a linear equation $\omega \cdot x = r$, we denote by $Supp(\omega)$ the set of non-zero coordinates in ω and by Ω_S the set of all coefficient vectors ω such that $Supp(\omega) \subseteq S$. Λ_S denotes all equations $(\omega \cdot x = r) \in \Lambda$ such that $\omega \in \Omega_S$.

Suppose $\exists r_1 \neq r_2$ such that $(\omega \cdot x = r_1) \in \Lambda$ and $(\omega \cdot x = r_2) \in \Lambda$ with $|Supp(w)| < t$. Then, $(\overrightarrow{0} \cdot x = r_1 - r_2) \in \mathsf{Res}(\Lambda, 2t)$. Conversely, if we know that $(\overrightarrow{0} \cdot x = r) \notin \mathsf{Res}(\Lambda, 2t)$ for any $r \neq 0$, then we can assume that for any $\omega \in \Omega$, there is at most one value of r such that $(\omega, r) \in \Lambda$. We will abuse notation to write $\omega \in \Lambda$ when some such r exists and is guaranteed to be unique. Also, for ω such that $(\omega \cdot x = r) \in \Lambda$ we define the function $\lambda(\omega)$, which specifies what the value of $\omega \cdot x$ should be for any satisfying assignment (or partial assignment) according to Λ. We take $\lambda(\omega) = r$ if there exists a unique r such that $(\omega \cdot x = r) \in \Lambda$ and undefined otherwise.

Proof Idea: The idea of the proof is to "encode" the partial assignments in the vectors in such a way so that it is easy to enforce consistency according to the given system of constraints. Since the constraints are in the form of linear equations, it is easiest to specify the value of all the linear forms $\omega \cdot x$.

In the binary case, one can think that in the vector $\mathbf{V}_{(S,\alpha)}$ we have a coordinate for each ω in which we specify $\mathbb{E}(-1)^{\omega\cdot x}$, where the expectation is over all x which agree with the assignment α. This specifies all the Fourier coefficients of the function which is 1 iff x is consistent with α, and hence "encodes" α. When $Supp(\omega) \not\subseteq S$, this expectation vanishes and hence one only needs to bother about ω in the set $\Omega = \cup_{|S|<t}\Omega_S$. Furthermore, because of the nature of dependencies by linear constraints, the value of $(\omega_1-\omega_2)\cdot x$ is either completely determined by Λ (when $(\omega_1-\omega_2) \in \Lambda$) or is completely undetermined and hence $\omega_1 \cdot x$ and $\omega_2 \cdot x$ are independent of each other. We thus partition Ω into various equivalence classes based on the set of equations Λ such that all linear forms within a class are completely determined by fixing the value of any one of them. The vectors we construct will have one coordinate corresponding to each of these classes which will enforce the dependencies due to Λ automatically.

Finally, to generalize this to q-ary equations, the natural analogue would be to consider the powers of the roots of unity i.e. expressions of the form $\exp(\frac{2\pi i(\omega\cdot x)}{q})$. However, this is not an option, since the coordinates of the vectors are not allowed to be complex. It is easy to check though that the proof in the binary case requires only one property of the coordinates that if two vectors disagree on the value in one coordinate (i.e. the product is -1), then the disagreements and agreements are balanced over all the coordinates and hence the inner product of the vectors is zero.

In the q-ary case, it can be proved that if the difference in the value of $\omega \cdot x$ in some coordinate according to two vectors is $\Delta \in [q], \Delta \neq 0$, then over all the coordinates, all values of Δ (including 0) occur equally often. We then choose each "coordinate" to be a small $q-1$ dimensional vector such that the product of the vectors at a coordinate is $-1/(q-1)$ if $\Delta \neq 0$ and 1 if $\Delta = 0$. This, combined with the balance property, still gives the orthogonality of the vectors which correspond to inconsistent assignments, and suffices for our purposes. The details are given in the proof below.

Proof of Theorem 4.9: Let $\Omega = \cup_{|S|<t}\Omega_S$. For $\omega_1, \omega_2 \in \Omega$, we say $\omega_1 \sim \omega_2$ if $(\omega_1 - \omega_2) \in \Lambda$. Since this is an equivalence relation, this partitions Ω into equivalence classes $C_1, \ldots, C_N$. We write $C(\omega)$ to denote the class containing ω. Next, we choose a representative (say the lexicographically first element) for each class. We use $[C]$ to denote the representative for the class C.

For constructing the vector $\mathbf{V}_{(S,\alpha)}$, we assign it a $q-1$ dimensional "coordinate" corresponding to each equivalence class of the above relation i.e. for each class we choose a $q-1$ dimensional vector and the final vector $\mathbf{V}_{(S,\alpha)}$ is the direct sum of all these vectors. Let $e_0, e_1, \ldots, e_{q-1}$ denote the q maximally separated unit vectors in $q-1$ dimensions such that $\left\langle e_i, e_j \right\rangle = -\frac{1}{q-1}$ if $i \neq j$ and 1 if $i = j$. Using $\mathbf{V}_{(S,\alpha)}(C)$ to denote the coordinate corresponding to C, we define $\mathbf{V}_{(S,\alpha)}$ as:

$$\mathbf{V}_{(S,\alpha)}(C) = \begin{cases} 0 & \text{if } \alpha \text{ disagrees with some equation in } \Lambda_S \\ 0 & \text{if } C \cap \Omega_S = \emptyset \\ \frac{1}{|A_S|} \cdot e_{\omega\cdot\alpha+\lambda([C(\omega)]-\omega)} & \text{for any } \omega \in C \cap \Omega_S \end{cases}$$

Here $\omega\cdot\alpha$ is defined as $\sum_{i\in S} \omega_i \alpha(i)$ is the inner product of ω with the partial assignment α, which can be computed since $Supp(\omega) \subseteq S$. The expression $\omega \cdot \alpha + \lambda([C(\omega)] - \omega)$ is computed modulo q. To show that the vector is well defined, we first need to argue that the coordinate $\mathbf{V}_{(S,\alpha)}(C)$ does not depend on which ω we choose from $C \cap \Omega_S$.

Claim 4.10 *If $\alpha \in [q]^S$ satisfies all equations in Λ_S, then for any class C and $\omega_1, \omega_2 \in C \cap \Omega_S$*

$$\omega_1 \cdot \alpha + \lambda([C] - \omega_1) = \omega_2 \cdot \alpha + \lambda([C] - \omega_2)$$

Proof: Since $([C] - \omega_1), ([C] - \omega_2), (\omega_1 - \omega_2) \in \Lambda$ and each have support size at most $2t$, it must be the case that $\lambda([C] - \omega_2) - \lambda([C] - \omega_1) = \lambda(\omega_1 - \omega_2)$ otherwise we can derive $\overrightarrow{0} \cdot x = r$ for $r \neq 0$. Also, since α is consistent with Λ_S, it must satisfy $(\omega_1 - \omega_2) \cdot \alpha = \lambda(\omega_1 - \omega_2)$. The claim follows. ■

The next claim shows that the only way two vectors $\mathbf{V}_{(S_1,\alpha_1)}$ and $\mathbf{V}_{(S_2,\alpha_2)}$ can have non-zero inner product is by having $\langle \mathbf{V}_{(S_1,\alpha_1)}(C), \mathbf{V}_{(S_2,\alpha_2)}(C) \rangle = \frac{1}{|A_{S_1}||A_{S_2}|}$ for each coordinate C in which it is non-zero. The theorem essentially follows from this claim using a simple counting argument.

Claim 4.11 *Let $\alpha_1 \in [q]^{S_1}$ and $\alpha_2 \in [q]^{S_2}$ be two partial assignments. If $\langle \mathbf{V}_{(S_1,\alpha_1)}(C), \mathbf{V}_{(S_2,\alpha_2)}(C) \rangle < 0$ for some class C, then $\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)} \rangle = 0$.*

Proof: For any class C' with $\langle \mathbf{V}_{(S_1,\alpha_1)}(C'), \mathbf{V}_{(S_2,\alpha_2)}(C') \rangle \neq 0$, we have some $\omega_1' \in C' \cap \Omega_{S_1}$ and $\omega_2' \in C' \cap \Omega_{S_2}$. We will work with the quantity $\lambda\left(\left[C(\omega_1')\right] - \omega_1'\right) + \omega_1' \cdot \alpha_1 - \lambda\left(\left[C(\omega_2')\right] - \omega_2'\right) - \omega_2' \cdot \alpha_2$ denoted by $\Delta(C')$, which is the difference of the indices of $\mathbf{V}_{(S_1,\alpha_1)}(C')$ and $\mathbf{V}_{(S_2,\alpha_2)}(C')$. It is easy to see that

$$\langle \mathbf{V}_{(S_1,\alpha_1)}(C'), \mathbf{V}_{(S_2,\alpha_2)}(C') \rangle = \begin{cases} \frac{1}{|A_{S_1}||A_{S_2}|} & \text{if } \Delta(C') = 0 \\ -\frac{1}{(q-1)|A_{S_1}||A_{S_2}|} & \text{otherwise} \end{cases}$$

For any r_1, r_2 we will give an injective map which maps a class C_{i_1} having $\Delta(C_{i_1}) = r_1$ to a class C_{i_2} having $\Delta(C_{i_2}) = r_2$. Hence over all the classes, all values of $\Delta(C')$ must occur equally often. This would imply the claim, since if $\langle \mathbf{V}_{(S_1,\alpha_1)}(C'), \mathbf{V}_{(S_2,\alpha_2)}(C') \rangle \neq 0$ for N_0 classes C', then

$$\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)} \rangle = \frac{N_0}{q} \cdot \frac{1}{|A_{S_1}||A_{S_2}|} + \frac{N_0(q-1)}{q} \cdot \left(-\frac{1}{(q-1)|A_{S_1}||A_{S_2}|}\right) = 0$$

We now construct the above map using the class C. Let $\omega_1 \in C \cap \Omega_{S_1}$ and $\omega_2 \in C \cap \Omega_{S_2}$. If $\langle \mathbf{V}_{(S_1,\alpha_1)}(C), \mathbf{V}_{(S_2,\alpha_2)}(C) \rangle < 0$, then we have that

$$\lambda(\omega_2 - \omega_1) + \omega_1 \cdot \alpha_1 - \omega_2 \cdot \alpha_2 = \Delta(C) \neq 0$$

Here we used the fact that

$$\lambda([C] - \omega_1) - \lambda([C] - \omega_2) = \lambda(\omega_2 - \omega_1)$$

Let C' be any class such that $\langle \mathbf{V}_{(S_1,\alpha_1)}(C'), \mathbf{V}_{(S_2,\alpha_2)}(C') \rangle \neq 0$ Let $\omega_1' \in C \cap \Omega_{S_1}$ and $\omega_2' \in C' \cap \Omega_{S_2}$. Then

$$\lambda\left(\omega_2' - \omega_1'\right) + \omega_1' \cdot \alpha_1 - \omega_2' \cdot \alpha_2 = \Delta(C')$$

where $\Delta(C')$ may now also be 0. Thus, for all $\mu \in [q]$, we get that

$$\mu\lambda(\omega_2 - \omega_1) + \lambda\left(\omega_2' - \omega_1'\right) + (\mu\omega_1 + \omega_1') \cdot a_1 - (\mu\omega_2 + \omega_2') \cdot a_2 \quad = \quad \mu\Delta(C) + \Delta(C') \tag{4.1}$$

Since $(\omega_2 - \omega_1) \in \Lambda$ and $(\omega_2' - \omega_1') \in \Lambda$ and each involves at most t variables, we also have $(\mu\omega_1 + \omega_1') - (\mu\omega_2 + \omega_2') \in \Lambda$. Hence, $(\mu\omega_1 + \omega_1')$ and $(\mu\omega_2 + \omega_2')$ must be in the same class, say C'', and we can write

$$\begin{aligned} \mu\lambda(\omega_2 - \omega_1) + \lambda\left(\omega_2' - \omega_1'\right) &= \lambda\left((\mu\omega_2 + \omega_2') - (\mu\omega_1 + \omega_1')\right) \\ &= \lambda\left([C''] - (\mu\omega_1 + \omega_1')\right) - \lambda\left([C''] - (\mu\omega_2 + \omega_2')\right) \end{aligned}$$

Combining this with equation 4.1, we get that

$$\Delta(C'') = \mu\Delta(C) + \Delta(C')$$

Since $\Delta(C) \neq 0$, for any r_1, r_2, choosing $\mu = (r_2 - r_1)/\Delta(C)$ gives a mapping in which the image of a class C' with $\Delta(C') = r_1$ is the class $C(C'')$ with $\Delta(C'') = r_2$. It is also easy to check that this mapping is injective and hence the claim follows. ■

From the above claim, we get property (1) in the theorem, since if two vectors have non-zero inner product, it must be positive *in every coordinate*. From the above claim it also follows that the inner product is only non-zero for vectors corresponding to partial assignments which are "mutually consistent" in the sense described below. The following also proves property (2) as stated in the theorem.

Claim 4.12 *If* $\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)} \rangle \neq 0$*, then* α_1 *and* α_2 *agree on all the variables in* $S_1 \cap S_2$*. Moreover, the assignment over* $S_1 \cup S_2$ *defined by* α_1 *and* α_2 *satisfies all the equations in* $\Lambda_{S_1 \cup S_2}$*.*

Proof: By Claim 4.11, it suffices to show that when α_1 and α_2 disagree on some variable in $S_1 \cap S_2$ or when $\alpha_1 \circ \alpha_2$ violates some equation in $\Lambda_{S_1 \cup S_2}$, then there exists a class C such that that $\langle \mathbf{V}_{(S_1,\alpha_1)}(C), \mathbf{V}_{(S_2,\alpha_2)}(C) \rangle < 0$. Consider the case when for $i \in S_1 \cap S_2$, $\alpha_1(i) \neq \alpha_2(i)$. Let ω be the vector which has coefficient 1 corresponding to x_i and all has other coefficients as zero. Then $\lambda([C(\omega)] - \omega) + \omega \cdot \alpha_1 - \lambda([C(\omega)] - \omega) - \omega \cdot \alpha_2 = \alpha_1(i) - \alpha_2(i) \neq 0$, which implies that $\langle \mathbf{V}_{(S_1,\alpha_1)}(C(\omega)), \mathbf{V}_{(S_2,\alpha_2)}(C(\omega)) \rangle = \frac{1}{|A_{S_1}||A_{S_2}|} \langle e_{\omega\cdot\alpha_1 + \lambda([C(\omega)]-\omega)}, e_{\omega\cdot\alpha_2+\lambda([C(\omega)]-\omega)} \rangle < 0$.

Next, suppose that α_1 and α_2 agree on $S_1 \cap S_2$, but $\alpha_1 \circ \alpha_2$ violates some equation $(\omega \cdot x = r) \in \Lambda_{S_1 \cup S_2}$ i.e. $(\alpha_1 \circ \alpha_2) \cdot \omega \neq r$. Let ω_1 be the vector which is the same as ω for all coordinates in S_1 and is zero otherwise. It is clear that $\omega_1 \in \Omega_{S_1}$ and $(\omega_1 - \omega) \in \Omega_{S_2}$. Also $\omega_1 \sim (\omega_1 - \omega)$ as their difference is ω which is in Λ. Let both of them be in the class C and consider $\lambda([C] - \omega_1) + \alpha_1 \cdot \omega_1 - \lambda([C] - (\omega_1 - \omega)) - \alpha_2 \cdot (\omega_1 - \omega)$. This is equal to $(\alpha_1 \circ \alpha_2) \cdot \omega - \lambda(\omega)$ which is non-zero by assumption. Hence, $\langle \mathbf{V}_{(S_1,\alpha_1)}(C), \mathbf{V}_{(S_2,\alpha_2)}(C) \rangle < 0$ and the claim follows. ■

Properties (2) and (3) will now follow from the claim below.

Claim 4.13 *If* $\alpha_1(S_1 \cap S_2) = \alpha_2(S_1 \cap S_2)$ *and* $\alpha_1 \circ \alpha_2$ *satisfies all equation in* $\Lambda_{S_1 \cup S_2}$*, then*

$$\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S,\alpha_2)} \rangle = \frac{1}{|A_{S_1 \cup S_2}|}$$

Proof: Let C be any class such that $\langle \mathbf{V}_{(S_1,\alpha_1)}(C), \mathbf{V}_{(S,\alpha_2)}(C)\rangle \neq 0$. Then there exist some elements $\omega_1 \in C \cap \Omega_{S_1}$ and $\omega_2 \in C \cap \Omega_2$. Since $\alpha_1 \circ \alpha_2$ satisfies all equations in $\Lambda_{S_1 \cup S_2}$, we must have that $\lambda(\omega_1 - \omega_2) = (\alpha_1 \circ \alpha_2)\cdot(\omega_1 - \omega_2) = \alpha_1 \cdot \omega_1 - \alpha_2 \cdot \omega_2$. This gives that $\lambda([C] - \omega_1) - \alpha_1 \cdot \omega_1 = \lambda([C] - \omega_2) - \alpha_2 \cdot \omega_2$ and hence $\langle \mathbf{V}_{(S_1,\alpha_1)}(C), \mathbf{V}_{(S,\alpha_2)}(C)\rangle = \frac{1}{|A_{S_1}||A_{S_2}|}$. So to compute the inner product, we only need to know how many classes intersect both Ω_{S_1} and Ω_{S_2}.

Note that Λ_S (as a set of linear combinations without the RHS) is a subgroup of Ω_S under addition. For calculating the inner product $\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)}\rangle$, we consider the quotient group $\Omega_{S_1\cup S_2}/\Lambda_{S_1\cup S_2}$ with $\Omega_{S_1}/\Lambda_{S_1\cup S_2}$ and $\Omega_{S_2}/\Lambda_{S_1\cup S_2}$ being subgroups of it. $\Omega_{S_1}/\Lambda_{S_1}$ (which is the same as $\Omega_{S_1}/\Lambda_{S_1\cup S_2}$) has one element for each class which has non-empty intersection with Ω_{S_1} (similarly for S_2). Hence, the number of classes intersecting both Ω_{S_1} and Ω_{S_2} is equal to $\left|\Omega_{S_1}/\Lambda_{S_1} \cap \Omega_{S_2}/\Lambda_{S_2}\right|$.

For A and B which are subgroups of a group G, we know that $|A+B| = |A||B|/|A\cap B|$. Using this gives

$$\left|\Omega_{S_1\cup S_2}/\Lambda_{S_1\cup S_2}\right| = \left|\Omega_{S_1}/\Lambda_{S_1\cup S_2} + \Omega_{S_2}/\Lambda_{S_1\cup S_2}\right| = \frac{|\Omega_{S_1}/\Lambda_{S_1\cup S_2}||\Omega_{S_2}/\Lambda_{S_1\cup S_2}|}{|\Omega_{S_1}/\Lambda_{S_1\cup S_2} \cap \Omega_{S_2}/\Lambda_{S_1\cup S_2}|}$$

Finally, noting that $\Omega_{S_1}/\Lambda_{S_1\cup S_2}$ is the same as $\Omega_{S_1}/\Lambda_{S_1}$ and $|\Omega_{S_1}/\Lambda_{S_1}| = |A_{S_1}|$ (similarly for S_2), we get that

$$\left|\Omega_{S_1}/\Lambda_{S_1} \cap \Omega_{S_2}/\Lambda_{S_2}\right| = \frac{|A_{S_1}||A_{S_2}|}{|A_{S_1\cup S_2}|}$$

Thus, we have

$$\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)}\rangle = \left|\Omega_{S_1}/\Lambda_{S_1} \cap \Omega_{S_2}/\Lambda_{S_2}\right| \cdot \frac{1}{|A_{S_1}||A_{S_2}|} = \frac{1}{|A_{S_1\cup S_2}|}$$

which proves the claim. ∎

To check property (3) we note that

$$\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)}\rangle = \langle \mathbf{V}_{(S_3,\alpha_3)}, \mathbf{V}_{(S_4,\alpha_4)}\rangle = 0$$

by Claim 4.12 when $\alpha_1 \circ \alpha_2$ violates any constraint in $\Lambda_{S_1\cup S_2}$. When it does not violate any constraints, Claim 4.13 applies and we have

$$\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)}\rangle = \langle \mathbf{V}_{(S_3,\alpha_3)}, \mathbf{V}_{(S_4,\alpha_4)}\rangle = \frac{1}{|A_{S_1\cup S_2}|}$$

From Claims 4.12 and 4.13, it is also immediate that $\left|\mathbf{V}_{(S,\alpha)}\right|^2 = 1/|A_S|$ if $\alpha \in A_S$ and is 0 otherwise. Hence, $\sum_{\alpha\in A_S}\left|\mathbf{V}_{(S,\alpha)}\right|^2 = 1$ which proves property (4). ∎

4.5 Deriving the gaps for MAX k-CSP$_q$

We will prove an integrality gap for general instances of MAX k-CSP (P_A). Specializing these to Hamming codes will then give the claimed results for MAX k-CSP$_q$. Let $A \in (\mathbb{F}_q)^{d\times k}$ be a matrix with linearly independent rows. Recall that an instance of MAX k-CSP (P_A) is specified by constraints $C_1, \ldots, C_m$ where each constraint C_i is a system of linear equations over $\mathbb{F}_q$ of the form $A \cdot (x_{i_1} \ldots x_{i_k})^{\mathsf{T}} = b^{(i)}$. We first show that if an instance Φ of MAX k-CSP (P_A) has good expansion, then FRAC(Φ) $= m$ even after large number of levels. The integrality gap will then follow as an easy consequence.

Theorem 4.14 *Let $A \in (\mathbb{F}_q)^{d \times k}$ be the generator matrix of a distance 3 code. Let Φ be an instance of* MAX k-CSP *(P_A) with $m = \beta n$ constraints such that for $s \leq \eta n$, every set of s constraints contains at least $(k - 1 - \delta)s$ variables for $\delta \leq 1/4$. Then,* FRAC(Φ) $= m$ *for the SDP relaxation of* MAX k-CSP *(P_A) obtained by $\eta n/16$ levels of the Lasserre hierarchy.*

Proof: We take Λ_0 to the set of all linear equations appearing in Φ, and take Λ to be the closure of it under Res($\cdot, t$) for $t = \eta n/8$. It will follow easily from Theorem 4.9 that FRAC(Φ) $= m$ once we establish that Λ satisfies the necessary conditions to apply the theorem i.e. $(\overrightarrow{0} \cdot x = r) \in \Lambda \Leftrightarrow r = 0$. Expansion based arguments have often been used in previous works in proof complexity. For example, [Sch08] used the argument due to Ben-Sasson and Wigderson [BSW01]. We will show that even when the equations are not independent as in [BSW01] and [Sch08] (equations for every constraint come from A), the property that A has distance 3 shall turn out to be sufficient for (a modification of) the argument.

Let us assume that $(\overrightarrow{0} \cdot x = r_0) \in \Lambda$ for some $r_0 \neq 0$. We shall show that any derivation of this must have an intermediate equation involving too many variables, thus deriving a contradiction. Consider a minimal derivation tree of $(\overrightarrow{0} \cdot x = r_0)$ for $r_0 \neq 0$. By definition of Res($\cdot, t$), this will be a binary tree with the equation $(\overrightarrow{0} \cdot x = r_0)$ at the root and equations in Λ_0 at the leaves.

Let $(\omega \cdot x = r)$ be any intermediate equation in the tree. We denote the node by (ω, r). We denote by $\nu(\omega, r)$ the number of constraints of Φ used in deriving (ω, r). It is immediate that if (ω_3, r_3) can be derived from (ω_1, r_1) and (ω_2, r_3), then $\nu(\omega_3, r_3) \leq \nu(\omega_1, r_1) + \nu(\omega_2, r_2)$. We first observe that $\nu(\overrightarrow{0}, r_0)$ must be large.

Claim 4.15 $\nu(\overrightarrow{0}, r_0) \geq \eta n$

Proof: let $\nu(\overrightarrow{0}, r_0) = s$. Then, the derivation for $(\overrightarrow{0}, r_0)$ involves a linear combination of equations from s constraints. Since every constraint comes from a distance 3 code, every linear combination of equations *within* a constraint must involve at least 3 variables. Hence, the linear combination required to derive $(\overrightarrow{0}, r_0)$ must include at least 3 variables from each of the s constraints. However, to derive $\overrightarrow{0}$ each of these variables must occur an even number of times and hence the s constraints can involve at most $ks - 3s/2$ variables in total. Since every set of up to ηn constraints is highly expanding, this is only possible when $s \geq \eta n$. ■

In the spirit of [BSW01], we show that one can find a node (ω, r) with $\eta n/2 \leq \nu(\omega, r) \leq \eta n$. Since for less than ηn constraints we can use expansion, we will be able to argue that this node has an equation involving many variables.

Claim 4.16 *There is a node (ω, r) in the tree such that*

$$\eta n/2 \;\leq\; \nu(\omega, r) \;\leq\; \eta n$$

Proof: Let $(\omega_1, r_1), (\omega_2, r_2)$ be the two children of the root $(\overrightarrow{0}, r_0)$. (ω_j, r_j). Since $\nu(\omega_1, r_1) + \nu(\omega_2, r_2) \geq \nu(\overrightarrow{0}, r_0) \geq \eta n$, at least one of them, say (ω_1, r_1), must require more than $\eta n/2$ constraints to derive it.

If $\nu(\omega_1, r_1) \leq \eta n$ then we are done. Else at least one of the children of (ω_1, r_1) must require more than $\eta n/2$ constraints for its derivation and we can continue the argument on this node. Since we always go down one level in the tree and find a node requiring at least ηn constraints, we must stop at some node as the leaves require only one constraint. The node we stop at can be taken to be (ω, r) as required. ∎

Consider the number of variables in ω. Each one of the constraints used in deriving it, contributes at least 3 variable occurrences. Also, since $\nu(\omega, r) \leq \eta n$, all the constraints must contain at least $(k - 1 - \delta)\nu(\omega, r)$ variables in total, which gives that at most $(1 + \delta)\nu(\omega, r)$ variables occurring in more than one constraint. Out of all the variable occurrences in ω, the ones that can cancel out are the ones occurring in more than one constraint. Hence, ω must have at least $(3 - 2(1 + \delta))\nu(\omega, r)$ variables. For $\delta \leq 1/4$, this is greater than $\eta n/8$ which is a contradiction.

Hence, Λ cannot contain an equation of the form $\overrightarrow{0} \cdot x = r_0$ for $r_0 \neq 0$. Since Λ is closed under $\mathsf{Res}(\cdot, \eta n/8)$ by definition, we can apply Theorem 4.9 to get vectors for all sets of size upto $\eta n/8$. The vectors also satisfy all the required consistency conditions. Finally, we note that

$$\sum_{i=1}^{m} \sum_{\alpha \in [q]^{T_i}} C_i(\alpha) \left|\mathbf{V}_{(T_i,\alpha)}\right|^2 \;=\; \sum_{i=1}^{m} \sum_{\alpha \in A_{T_i}} \left|\mathbf{V}_{(T_i,\alpha)}\right|^2 \;=\; \sum_{i=1}^{m} 1 \;=\; m$$

which shows that $\mathsf{FRAC}(\Phi) = m$. ∎

Since random instances are both unsatisfiable and expanding, it is now easy to derive the integrality gap for MAX k-CSP (P_A).

Theorem 4.17 *Let $A \in (\mathbb{F}_q)^{d \times k}$ be the generator matrix of a distance 3 code and let $\zeta > 0$ be given. Then there is a constant $c = c(q, k, \zeta)$ such that for large enough n, the integrality gap for the Lasserre SDP relaxation of* MAX k-CSP *(P_A) on n variables obtained by cn levels is at least $q^d - \zeta$.*

Proof: We take $\varepsilon = \zeta \cdot \frac{1}{q^d}$, $\delta = 1/4$ and consider a random instance Φ with $m = \beta n$ constraints as in Lemma 4.2, such that the Φ satisfies both the properties in the conclusion of the lemma. Then, by Theorem 4.14 $\mathsf{FRAC}(\Phi) = m$ even after $\eta n/16$ levels of the Lasserre hierarchy. On the other hand, by Lemma 4.2, $\mathsf{OPT}(\Phi) \leq \frac{|P_A^{-1}(1)|}{q^k}(1 + \varepsilon) \cdot m = \frac{1}{q^d}(1 + \varepsilon) \cdot m$. Hence, the integrality gap is at least $q^d/(1 + \varepsilon) \geq 1/q^d - \zeta$. ∎

We now derive near optimal integrality gaps for $\Omega(n)$ levels of the Lasserre relaxation of the binary and q-ary MAX k-CSP problems. Note that the integrality gap becomes larger with the dimension of the code. Thus, to optimize the gap, we consider Hamming codes which have the largest dimension for a given length.

Corollary 4.18 *Let $k \geq 3$, q be a prime number and let $\zeta > 0$ be given. Then there exists $c = c(k, q, \zeta) > 0$ such that the for sufficiently large n, the integrality gap for the Lasserre relaxation of the* MAX k-CSP *problem on n variables with domain size q is at least* $\frac{q^k}{kq(q-1) - q(q-2)} - \zeta$ *after cn levels.*

Proof: Let $(q^{r-1} - 1)/(q - 1) < k \leq (q^r - 1)/(q - 1)$. We take A to be the generator matrix of the Hamming code of length k. Note that the above implies that $q^r - q \leq (k - 1)q(q - 1)$ which gives $q^r \leq kq(q - 1) - q(q - 2)$.

Consider a random instance of MAX k-CSP (P_A). By Theorem 4.17, there exists a c and instances on n variables such that the integrality gap after cn levels for MAX k-CSP (P_A) is at least $q^d - \zeta = q^k/q^r - \zeta$. Finally, using that $q^r \leq kq(q - 1) - q(q - 2)$ gives the gap is at least $\frac{q^k}{kq(q-1) - q(q-2)} - \zeta$ as claimed. ■

Note that in case of $q = 2$, the generator matrix of the binary Hamming code, simply produces the predicate considered by Samorodnitsky and Trevisan [ST06]. We also state the following more precise version of the gap for binary k-CSPs. The constants are arbitrary, but we shall need the nature of the tradeoffs between k, β and c to get the gaps for Independent Set and Chromatic Number.

Corollary 4.19 *Let a number k and $\varepsilon > 0$ be given and let A be the generator matrix for the Hamming code of length k. Then there exist $\beta = O(2^k/\varepsilon^2)$ and $c = \Omega((1/\beta)^{25})$ such that if Φ is a random instance of* MAX k-CSP *(P_A) on $n \gg 1/c$ variables and $m = \beta n$ constraints, then with probability $1 - o(1)$*

1. $\mathsf{OPT}(\Phi) \leq \frac{2k}{2^k}(1 + \varepsilon) \cdot m$
2. *For the SDP given by cn levels of the Lasserre hierarchy,* $\mathsf{FRAC}(\Phi) = m$.

Proof: Invoking Lemma 4.2 with $\delta = 1/5$ gives $\beta = O(2^k/\varepsilon^2)$ and $\eta = \Omega((1/\beta)^{25})$. Theorem 4.14 gives $\mathsf{FRAC}(\Phi) = m$ after cn levels, for $c = \Omega(\eta) = \Omega((1/\beta)^{25})$. The dimension d for a Hamming code is $k - 2^{\lceil \log(k+1) \rceil} \geq k - \log(2k)$. Hence $\mathsf{OPT}(\Phi) \leq \frac{2k}{2^k}(1 + \varepsilon)m$. ■

4.6 Implications for Random Predicates

We now derive integrality gaps for MAX k-CSP (P) where $P : [q]^k \to \{0, 1\}$ is chosen at random by selecting each input to be in $P^{-1}(1)$ independently with probability p. We denote this distribution over q-ary predicates with k inputs as $Q(p, q, k)$. For any predicate P, a random assignment satisfies $|P^{-1}(1)|/q^k$ fraction of constraints in MAX k-CSP (P) and hence the largest integrality gap one can have is $q^k/|P^{-1}(1)|$. We will show that for almost every random predicate P, the integrality gap for MAX k-CSP (P) is *at least* $q^k/|P^{-1}(1)| - \zeta$ even after $\Omega(n)$ levels of the Lasserre hierarchy.

The result will follow quite easily using a theorem of Håstad [Hås07] which basically says that a random predicate "contains a copy of" P_A where A is the generator matrix of the Hamming code over $\mathbb{F}_q$. We first define what it means for a predicate to contain a copy of another.

Definition 4.20 *We say that a predicate P_1 contains a predicate equivalent to P_2 if there exists a permutation $\pi : [k] \to [k]$ of the inputs and $b_1, \ldots, b_k \in [q]$, such that*

$$P_2(x_{\pi(1)} + b_1, \ldots, x_{\pi(k)} + b_k) = 1 \;\Rightarrow\; P_1(x_1, \ldots, x_k) = 1$$

We can now state the theorem of Håstad referred to above. Håstad actually states the result only for random boolean predicates but it is easy to verify that the same proof can be extended to q-ary predicates.

Theorem 4.21 *[Hås07]) Let q be a prime and let $(q^{r-1} - 1)/(q - 1) < k \le (q^r - 1)/(q - 1)$. Let A be the generator matrix of the Hamming code over $\mathbb{F}_q$ with length k. Then there is a value c of the form $c = kq^{-r}(1 - o(1))$, such that, with probability $1 - o(1)$, a random predicate chosen according to $Q(p, q, k)$ with $p \ge k^{-c}$ contains a predicate equivalent to P_A.*

Using this theorem, we can now prove optimal integrality gap for almost every predicate in the distribution $Q(p, q, k)$ with the appropriate value of p.

Theorem 4.22 *Let a prime q and $\zeta > 0$ and $k \ge 3$ be given and let $(q^{r-1} - 1)/(q - 1) < k \le (q^r - 1)/(q - 1)$ for some r. Then there exist constants $c = kq^{-r}(1 - o(1))$ and $c' = c'(q, k, \zeta)$ such that if P is a random predicate chosen according to $Q(p, q, k)$ with $p \ge k^{-c}$, then with probability $1 - o(1)$ over the choice of P, the integrality gap for* MAX k-CSP *(P) after $c'n$ levels of the Lasserre hierarchy is at least $q^k/|P^{-1}(1)| - \zeta$.*

Proof: Using Theorem 4.21 we know that with probability $1-o(1)$, a random P contains a predicate equivalent to P_A, where A is the generator matrix of the Hamming code over $\mathbb{F}_q$ with length k. For the rest of the proof, we fix such a P. For this P there exists a permutation π and literals $b_1, \ldots, b_k$ such that $\mathrm{P}_A(x_{\pi(1)} + b_1, \ldots, x_{\pi(k)} + b_k) = 1 \;\Rightarrow\; P(x_1, \ldots, x_k) = 1$.

With every instance Φ of MAX k-CSP (P), we now associate an instance Φ_A of MAX k-CSP (P_A). For every constraint $C_i \equiv P(x_{i_1} + a_{i_1}, \ldots, x_{i_k} + a_{i_k})$ in Φ, we add a constraint C'_i to Φ_A of the form

$$\mathrm{C}'_i \equiv P_A(x_{i_{\pi(1)}} + a_{i_{\pi(1)}} + b_1, \ldots, x_{i_{\pi(k)}} + a_{i_{\pi(k)}} + b_k)$$

Thus, if the constraint C'_i is satisfied by an assignment, then so is C_i. Also, if Φ is distributed as a random instance of MAX k-CSP (P), then Φ_A is distributed as a random instance of MAX k-CSP (P_A) with the same number of constraints.

Let $\varepsilon = \zeta \cdot |P^{-1}(1)|/q^k$ and $\delta = 1/4$. We consider a random instance Φ of MAX k-CSP (P) with $m = \beta n$ constraints as in Lemma 4.2. By Lemma 4.2 we will have with probability $1 - o(1)$ over Φ that

- $\mathsf{OPT}(\Phi) \le \dfrac{|P^{-1}(1)|}{q^k}(1 + \varepsilon) \cdot m$, and
- Every set of $s \le \eta n$ constraints in Φ_A contains at least $(k - 1 - \delta)s$ variables.

By Theorem 4.14, we have $\mathsf{FRAC}(\Phi_A) = m$ for the SDP obtained by $t = \eta n/16$ levels of the Lasserre hierarchy. Hence, there exist vectors $\mathbf{V}_{(S,\alpha)}$ for all $S \subseteq [n], |S| \leq t$ and $\alpha \in [q]^S$ satisfying all the consistency constraints and such that

$$\sum_{\alpha \in [q]^{T_i}} C_i'(\alpha) \left|\mathbf{V}_{(T_i,\alpha)}\right|^2 = 1 \qquad \forall 1 \leq i \leq m$$

However, the same vectors also show that $\mathsf{FRAC}(\Phi) = m$ after t levels, since

$$\sum_{i=1}^{m} \sum_{\alpha \in [q]^{T_i}} C_i(\alpha) \left|\mathbf{V}_{(T_i,\alpha)}\right|^2 \geq \sum_{i=1}^{m} \sum_{\alpha \in [q]^{T_i}} C_i'(\alpha) \left|\mathbf{V}_{(T_i,\alpha)}\right|^2 = \sum_{i=1}^{m} 1 = m$$

Hence, the integrality gap for MAX k-CSP (P) after $\eta n/16$ levels of the Lasserre hierarchy is at least $\mathsf{FRAC}(\Phi)/\mathsf{OPT}(\Phi) \geq \dfrac{q^k}{|P^{-1}(1)|(1+\varepsilon)} \geq \dfrac{q^k}{|P^{-1}(1)|} - \zeta$. ■

Chapter 5

Sherali-Adams Gaps from Pairwise Independence

In this chapter we study integrality gaps for a very general class of constraint satisfaction problems In a result which captures all the precious results on hardness of approximating MAX $\mathsf{k\text{-}CSP}_q$, Austrin and Mossel [AM08] showed that if $P : [q]^k \to \{0, 1\}$ is a predicate such that the set of accepted inputs $P^{-1}(1)$ contains the support of a balanced pairwise independent distribution μ on $[q]^k$, then MAX k-CSP (P) is UG-hard to approximate better than a factor of $q^k/|P^{-1}(1)|$. Considering that a random assignment satisfies $|P^{-1}(1)|/q^k$ fraction of all the constraints, this is the strongest result one can get for a predicate P. Using appropriate choices for the predicate P, this then implies hardness ratios of $q^k/kq^2(1 + o(1))$ for general $q \geq 2$, $q^k/q(q-1)k$ when q is a prime power, and $2^k/(k + O(k^{0.525})$ for $q = 2$.

We study the inapproximability of such a predicate P (which we call promising) in the hierarchy of linear programs defined by Sherali and Adams. In particular, we show an unconditional analogue of the result of Austrin and Mossel in this hierarchy.

The previous results in the Lasserre hierarchy (and earlier analogues in the Lovász-Schrijver hierarchy) seemed to be heavily relying on the structure of the predicate for which the integrality gap was proven, as being 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 (admittedly weaker) Sherali-Adams hierarchy.

Theorem 5.1 *Let $P : [q]^k \to \{0, 1\}$ be predicate such that $P^{-1}(1)$ contains the support of a balanced pairwise independent distribution μ. 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 relaxation obtained by cn levels of the Sherali-Adams hierarchy applied to the standard LP[1] is at least $\dfrac{q^k}{|P^{-1}(1)|} - \zeta$.*

We note that $\Omega(n^\delta)$-level gaps for these predicates can also be deduced via reductions from the recent result of [CMM09] who obtained $\Omega(n^\delta)$-level gaps for Unique Games, where $\delta \to 0$ as $\zeta \to 0$.

[1] See the resulting LP in section 5.1.2.

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. These distributions should be (a) supported only on satisfying (partial) assignments and (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. The second requirement guarantees that the obtained solution is indeed feasible, while the first implies that the solution achieves objective value that corresponds to satisfying *all* the constraints of the instance.

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 constraints graph. In doing so, 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 such distributions.

5.1 Preliminaries and Notation

We use the same notation for CSPs as in the previous chapter. In particular, an instance Φ of MAX k-CSP$_q$, has variable $\{x_1, \ldots, x_n\}$ with domain $[q]$, and constraints $C_1, \ldots, C_m$. Each constraint is a function of the form $C_i : [q]^{T_i} \to \{0, 1\}$ for some $T_i \subseteq [n]$ with $|T_i| \leq k$. For a set of variables $S \subseteq [n]$, $\alpha \in [q]^S$ denotes a partial assignment and $\alpha(S')$ denotes its projection to $S' \subseteq S$. Also, for $\alpha_1 \in [q]^{S_1}, \alpha_2 \in [q]^{S_2}$ such that $S_1 \cap S_2 = \emptyset$, we denote by $\alpha_1 \circ \alpha_2$ the assignment over $S_1 \cup S_2$ defined by α_1 and α_2. For a predicate $P : [q]^k \to \{0, 1\}$, $P^{-1}(1)$ is the set of assignments on which the predicate evaluates to 1. A constraint defined according to P is taken to be of the form $P(x_{i_1} + a_{i_1}, \ldots, x_{i_k} + a_{i_k})$ for $a_{i_1}, \ldots, a_{i_k} \in [q]$.

5.1.1 Expanding CSP Instances

For an instance Φ of MAX k-CSP$_q$, define its constraint graph G_Φ, as the following bipartite graph from L to 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_Φ by G.

For $C_i \in L$ we denote by $\Gamma(C_i) \subseteq R$ the neighbors $\Gamma(C_i)$ of C_i in R. For a set of constraints

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

Our result relies on set of constraints that are well expanding. We make this notion formal below.

Definition 5.2 *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 subsets of L of size at most r is at least e.*

5.1.2 The Sherali-Adams relaxation for CSPs

Below we present a relaxation for the MAX k-CSP$_q$ problem as it is obtained by applying a level-t Sherali-Adams relaxation of the standard LP formulation of some instance Φ of MAX k-CSP$_q$.

As before, an integer solution to the problem can be given by a single mapping $\alpha_0 \in [q]^{[n]}$, which is an assignment to all the variables. We define 0/1 variables $X_{(S,\alpha)}$ for each $S \subseteq [n]$ such that $|S| \leq t$ and $\alpha \in [q]^S$. The intended solution is $X_{(S,\alpha)} = 1$ if $\alpha_0(S) = \alpha$ and 0 otherwise. We also introduce $X_{(\emptyset,\emptyset)}$ which is intended to be 1. By relaxing the integrality constraint on the variables, we obtain the level-t Sherali-Adams LP relaxation.

$$\begin{array}{llcll}
\text{maximize} & \displaystyle\sum_{i=1}^{m} \sum_{\alpha \in [q]^{T_i}} C_i(\alpha) \cdot X_{(T_i,\alpha)} & & & \\
\text{subject to} & \displaystyle\sum_{j \in [q]} X_{(S \cup \{i\}, \alpha \circ j)} & = & X_{(S,\alpha)} & \forall S \ s.t.\ |S| < t,\ \forall i \notin S, \alpha \in [q]^S \\
& X_{(S,\alpha)} & \geq & 0 & \forall S \ s.t.\ |S| \leq t,\ \forall \alpha \in [q]^S \\
& X_{(\emptyset,\emptyset)} & = & 1 &
\end{array}$$

Figure 5.1: Sherali-Adams LP for MAX k-CSP$_q$

For an LP formulation of MAX k-CSP$_q$, and for a given instance Φ of the problem, we denote by FRAC(Φ) the LP (fractional) optimum, and by OPT(Φ) the integral optimum. For the particular instance Φ, the integrality gap is then defined as FRAC(Φ)/OPT(Φ). The integrality gap of the LP formulation is the supremum of integrality gaps over all instances.

Next we give a sufficient condition for the existence of a solution to the level-t Sherali-Adams LP relaxation for a MAX k-CSP$_q$ instance Φ.

Lemma 5.3 *Consider a family of distributions $\{\mathcal{D}(S)\}_{S \subseteq [n]:|S| \leq t}$, where each $\mathcal{D}(S)$ is defined over $[q]^S$. If for every $S \subseteq T \subseteq [n]$ with $|T| \leq t$, the distributions $\mathcal{D}(S), \mathcal{D}(T)$ are equal on S, then*

$$X_{(S,\alpha)} = \mathop{\mathbb{P}}_{\mathcal{D}(S)}[\alpha]$$

satisfy the above level-t Sherali-Adams relaxation.

Proof: Consider some $S \subseteq [n]$, $|S| < t$, and some $i \notin S$. Note that the distributions $\mathcal{D}(S), \mathcal{D}(S \cup \{i\})$ are equal on S, and therefore we have

$$\begin{aligned}
\sum_{j \in [q]} X_{(S \cup \{i\}, \alpha \circ j)} &= \sum_{j \in [q]} \underset{\beta \sim \mathcal{D}(S \cup \{i\})}{\mathbb{P}} [\beta = \alpha \circ j] \\
&= \sum_{j \in [q]} \underset{\beta \sim \mathcal{D}(S \cup \{i\})}{\mathbb{P}} [(\beta(i) = j) \wedge (\beta(S) = \alpha)] \\
&= \underset{\beta \sim \mathcal{D}(S \cup \{i\})}{\mathbb{P}} [\beta(S) = \alpha] \\
&= \underset{\beta' \sim \mathcal{D}(S)}{\mathbb{P}} [\beta' = \alpha] \\
&= X_{(S, \alpha)}.
\end{aligned}$$

The same argument also shows that if $S = \emptyset$, then $X_{(\emptyset,\emptyset)} = 1$. Finally, it is clear that all linear variables are assigned non negative values completing the lemma. ■

5.1.3 Pairwise Independence and Approximation Resistant Predicates

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

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

A predicate P is called approximation resistant if it is hard to approximate the MAX k-CSP$_q$ (P) problem better than using a random assignment. Assuming the Unique Games Conjecture, Austrin and Mossel [AM08] show that a predicate is approximation resistant if it is possible to define a balanced pairwise independent distribution μ such that P is always 1 on the support of μ.

Definition 5.4 *A predicate $P : [q]^k \to \{0, 1\}$ is called* promising*, if there exist a distribution supported over a subset of $P^{-1}(1)$ that is pairwise independent and balanced. If μ is such a distribution we say that P is promising supported by μ.*

5.2 Towards Defining Consistent Distributions

To construct valid solutions for the Sherali-Adams LP relaxation, we need to define distributions over every set S of bounded size as is required by Lemma 5.3. Since we will deal with promising predicates supported by some distribution μ, in order to satisfy consistency between distributions we will heavily rely on the fact that μ is a balanced pairwise independent distribution.

Consider for simplicity that μ is uniform 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 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 $\mathcal{D}(S)$ as the uniform distribution over all consistent assignments of 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 $S_1 = S_2 \setminus \{j\}$. Assuming that the support of no other constraint is contained in S_2, we get that distribution $\mathcal{D}(S_1)$ maps any variable in S_1 to $\{0,1\}$ with probability $1/2$ independently, but some of these assignments are not even extendible to S_2 meaning that $\mathcal{D}(S_2)$ will assign them with probability zero.

Thus, to define $\mathcal{D}(S)$, we cannot simply sample assignments satisfying all constraints in $C(S)$ with probabilities given by μ. 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 also is "local enough" so that $C(\overline{S})$ is not too large. We require the property of such sets that if we remove $\overline{S}$ and $C(\overline{S})$, then the remaining graph $G|_{-\overline{S}}$ still has good expansion. We deal with this in Section 5.2.1.

2. The distribution $\mathcal{D}(S)$ is going to be the uniform distribution over satisfying assignments in $\overline{S}$. In the case that μ 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 $\mathcal{P}_\mu(S)$, such that for $S_1 \subseteq S_2$, the distributions are guaranteed to be consistent if $G|_{-S_1}$ *has good expansion*. This appears in Section 5.2.2.

We then combine the two techniques and define $\mathcal{D}(S)$ according to $\mathcal{P}_\mu(\overline{S})$. This is done in section 5.3.

5.2.1 Finding Advice-Sets

We now give an algorithm below to obtain a superset $\overline{S}$ for a given set S, which we call the advice-set of S. It is inspired by the "expansion correction" procedure in [BOGH+03].

Theorem 5.5 *Algorithm Advice, with internal parameters e_1, e_2, r, returns $\overline{S} \subseteq R$ such that (a) $G|_{-\overline{S}}$ is (ξ_S, e_2) boundary expanding, (b) $\xi_S \geq r - \frac{|S|}{e_1 - e_2}$, and (c) $|\overline{S}| \leq \frac{e_1|S|}{e_1 - e_2}$.*

Proof: Suppose that the loop terminates with $\xi = \xi_S$. Then $\sum_{j=1}^{t} |M_j| = r - \xi_S$. Since G is (r, e_1) boundary expanding, the set $M = \cup_{j=1}^{t} M_j$ has initially at least $e_1(r - \xi_S)$ boundary neighbors in G. During the execution of the while loop, each set M_j has at most $e_2|M_j|$ boundary neighbors in $G|_{-\overline{S}}$. Therefore, at the end of the procedure M has at most $e_2(r - \xi_S)$ boundary neighbors in $G|_{-S}$. It follows that $|S| + e_2(r - \xi_S) \geq e_1(r - \xi_S)$, which implies (b).

From the bound size of S we know that $\xi_S > 0$. In particular, ξ remains positive throughout the execution of the while loop. Next we identify a loop invariant: $G|_{-\overline{S}}$ is (ξ, e_2) boundary expanding.

Indeed, note that the input graph G is (ξ, e_1) boundary expanding. At step j consider the set $\overline{S} \cup \{x_j\}$, and suppose that $G_{-(\overline{S} \cup \{x_j\})}$ is not (ξ, 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 $G_{-(\overline{S} \cup \{x_j\} \cup \partial M_j)}$ is $(\xi - |M_j|, e_2)$ boundary expanding

Algorithm Advice

The input is an (r, e_1) boundary expanding bipartite graph $G = (L, R, E)$, some $e_2 \in (0, e_1)$, and some $S \subseteq R$, $|S| < (e_1 - e_2)r$, with some order $S = \{x_1, \ldots, x_t\}$.

Initially set $\overline{S} \leftarrow \emptyset$ and $\xi \leftarrow r$
For $j = 1, \ldots, |S|$ **do**
  $M_j \leftarrow \emptyset$
  $\overline{S} \leftarrow \overline{S} \cup \{x_j\}$
  If $G|_{-\overline{S}}$ is not (ξ, e_2) boundary expanding **then**
    Find a maximal $M_j \subset L$ in $G|_{-\overline{S}}$, such that $|M_j| \le \xi$ in $G|_{-\overline{S}}$ and $|\partial M_j| \le e_2|M_j|$
    $\overline{S} \leftarrow \overline{S} \cup \partial M_j$
    $\xi \leftarrow \xi - |M_j|$
Return $\overline{S}$

(recall that since ξ remains positive, $|M_j| < \xi$). Now consider the contrary. Then, there must be $M' \subset L$ such that $|M'| \le \xi - |M_j|$ and such that $|\partial M'| \le e_2|M'|$. Consider then $M_j \cup M'$ and note that $|M_j \cup M'| \le \xi$. More importantly $|\partial(M_j \cup M')| \le e_2|M_j \cup M'|$, and therefore we contradict the maximality of M_j; (a) follows.

Finally note that $\overline{S}$ consists of S union the boundary neighbors of all M_j. From the arguments above, the number of those neighbors does not exceed $e_2(r - \xi_S)$ and hence $|\overline{S}| \le |S| + e_2(r - \xi_S) \le |S| + \frac{e_2|S|}{e_1 - e_2} = \frac{e_1|S|}{e_1 - e_2}$, which proves (c). ■

5.2.2 Defining the Distributions $\mathcal{P}_\mu(S)$

We now define for every set S, a distribution $\mathcal{P}_\mu(S)$ such that for any $\alpha \in [q]^S$, $\mathbb{P}_{\mathcal{P}_\mu(S)}[\alpha] > 0$ only if α satisfies all the constraints in $C(S)$. 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 $\mu_i : [q]^{T_i} \to [0, 1]$ 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 μ_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 $\prod_{C_i \in C(S)} \mu_i(\alpha(T_i))$. Formally,

$$\mathop{\mathbb{P}}_{\mathcal{P}_\mu(S)}[\alpha] \;=\; \frac{1}{Z_S} \cdot \prod_{C_i \in C(S)} \mu_i(\alpha(T_i)) \tag{5.1}$$

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

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

To understand the distribution, it is easier to think of the special case when μ is just the uniform distribution on $P^{-1}(1)$ (like in the case of MAX k-XOR). Then $\mathcal{P}_\mu(S)$ is simply the uniform distribution on assignments satisfying all the constraints in $C(S)$. When μ is not uniform, then the probabilities are weighted by the product of the values $\mu_i(\alpha(T_i))$ for all the constraints [2]. However, we still have the property that if $\mathbb{P}_{\mathcal{P}_\mu(S)}[\alpha] > 0$, then α satisfies all the constraints in $C(S)$.

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 5.6 *Let Φ be a* MAX k-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 $(r, k-2-\delta)$ boundary expanding for some $\delta \in (0,1)$ and $|C(S_2)| \le r$. Then $Z_{S_2} = q^{|S_2|}/q^{k|C(S_2)|}$, and for any $\alpha_1 \in [q]^{S_1}$*

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

Proof: Let $C = C(S_2) \setminus C(S_1)$ be given by the set of t many constraints $C_{i_1}, \ldots, C_{i_t}$ with each C_{i_j} being on the set of variables T_{i_j}. Some of these variables may be fixed by α_1. Also, any α_2 consistent with α_1 can be written as $\alpha_1 \circ \alpha$ for some $\alpha \in [q]^{S_2 \setminus S_1}$. Below, we express these probabilities in terms the product of μ on the constraints in $C(S_2) \setminus C(S_1)$.

Note that the equations below are still correct even if we haven't shown $Z_{S_2} > 0$ (in that case both sides are 0). In fact, replacing S_1 by $\emptyset$ in the same calculation will give the value of Z_{S_2}.

$$\begin{aligned}
Z_{S_2} \cdot \sum_{\substack{\alpha_2 \in [q]^{S_2} \\ \alpha_2(S_1)=\alpha_1}} \mathbb{P}_{\mathcal{P}_\mu(S_2)}[\alpha_2] &= \sum_{\alpha \in [q]^{S_2 \setminus S_1}} \prod_{C_i \in C(S_2)} \mu_i((\alpha_1 \circ \alpha)(T_i)) \\
&= \left(\prod_{C_i \in C(S_1)} \mu_i(\alpha_1(T_i)) \right) \sum_{\alpha \in [q]^{S_2 \setminus S_1}} \prod_{j=1}^{t} \mu_{i_j}((\alpha_1 \circ \alpha)(T_{i_j})) \\
&= \left(Z_{S_1} \cdot \mathbb{P}_{\mathcal{P}_\mu(S_1)}[\alpha_1] \right) \sum_{\alpha \in [q]^{S_2 \setminus S_1}} \prod_{j=1}^{t} \mu_{i_j}((\alpha_1 \circ \alpha)(T_{i_j})) \\
&= \left(Z_{S_1} \cdot \mathbb{P}_{\mathcal{P}_\mu(S_1)}[\alpha_1] \right) \cdot q^{|S_2 \setminus S_1|} \mathop{\mathbb{E}}_{\alpha \in [q]^{S_2 \setminus S_1}} \left[\prod_{j=1}^{t} \mu_{i_j}((\alpha_1 \circ \alpha)(T_{i_j})) \right]
\end{aligned}$$

The following claim lets us calculate this expectation conveniently using the expansion of $G|_{-S_1}$.

Claim 5.7 *Let C be as above. Then there exists an ordering $C_{i'_1}, \ldots, C_{i'_t}$ of constraints in C and a partition of $S_2 \setminus S_1$ into sets of variables $F_1, \ldots, F_t$ such that for all j, $F_j \subseteq T_{i'_j}$, $|F_j| \ge k-2$, and*

$$\forall j \; F_j \cap \left(\cup_{l>j} T_{i'_l} \right) = \emptyset.$$

Proof: (of Claim 5.7) We build the sets F_j inductively using the fact that $G|_{-S_1}$ is $(r, k-2-\delta)$ boundary expanding.

[2] Note however that $\mathcal{P}_\mu(S)$ is not a product distribution because different constraints in $C(S)$ may share variables.

Start with the set of constraints $C_1 = C$. Since $|C_1| = |C(S_2) \setminus C(S_1)| \leq r$, this gives that $|\partial(C_1) \setminus S_1| \geq (k-2-\delta)|C_1|$. Hence, there exists $C_{i_j} \in C_1$ such that $|T_{i_j} \cap (\partial(C_1) \setminus S_1)| \geq k-2$. Let $T_{i_j} \cap (\partial(C_1) \setminus S_1) = F_1$ and $i'_1 = i_j$. We then take $C_2 = C_1 \setminus \{C_{i'_1}\}$ and continue in the same way.

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, the expectation above can be split as

$$\mathop{\mathbb{E}}_{\alpha \in [q]^{S_2 \setminus S_1}} \left[\prod_{j=1}^{t} \mu_{i_j}(\alpha_1 \circ \alpha(T_{i_j}))\right] = \mathop{\mathbb{E}}_{\beta_t \in [q]^{F_t}} \left[\mu_{i'_t} \ldots \mathop{\mathbb{E}}_{\beta_2 \in [q]^{F_2}} \left[\mu_{i'_2} \mathop{\mathbb{E}}_{\beta_1 \in [q]^{F_1}} \left[\mu_{i'_1}\right]\right] \ldots\right]$$

where the input to each $\mu_{i'_j}$ depends on α_1 and $\beta_j, \ldots, \beta_t$ 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,

$$\mathop{\mathbb{E}}_{\beta_1 \in [q]^{F_1}} \left[\mu_{i'_1}\right] = \frac{1}{q^{|F_1|}} \cdot \mathop{\mathbb{P}}_{\mu}[(\alpha_1 \circ \beta_2 \ldots \circ \beta_t)(T_{i'_1} \setminus F_1)] = \frac{1}{q^k}$$

irrespective of the values assigned by $\alpha_1 \circ \beta_2 \circ \ldots \circ \beta_t$ to the remaining (at most 2) variables in $T_{i'_1} \setminus F_1$. Continuing in this fashion from right to left, we get that

$$\mathop{\mathbb{E}}_{\alpha \in [q]^{S_2 \setminus S_1}} \left[\prod_{j=1}^{t} \mu_{i_j}((\alpha_1 \circ \alpha)(T_{i_j}))\right] = \left(\frac{1}{q^k}\right)^t = \left(\frac{1}{q^k}\right)^{|C(S_2) \setminus C(S_1)|}$$

Hence, we get that

$$Z_{S_2} \cdot \sum_{\substack{\alpha_2 \in [q]^{S_2} \\ \alpha_2(S_1) = \alpha_1}} \mathop{\mathbb{P}}_{\mathcal{P}_\mu(S_2)}[\alpha_2] = \left(Z_{S_1} \cdot \frac{q^{|S_2 \setminus S_1|}}{q^{k|C(S_2) \setminus C(S_1)|}}\right) \mathop{\mathbb{P}}_{\mathcal{P}_\mu(S_1)}[\alpha_1]. \tag{5.2}$$

Summing over all $\alpha_1 \in [q]^{S_1}$ on both sides gives

$$Z_{S_2} = Z_{S_1} \cdot \frac{q^{|S_2 \setminus S_1|}}{q^{k|C(S_2) \setminus C(S_1)|}}.$$

Since we know that G is $(r, k-2-\delta)$ boundary expanding, we can replace S_1 by $\emptyset$ in the above equation to obtain $Z_{S_2} = q^{|S_2|}/q^{k|C(S_2)|}$ as claimed. Also note that since $C(S_1) \subseteq C(S_2)$, $Z_{S_2} > 0$ implies $Z_{S_1} > 0$. Hence, using equation (5.2) we get

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

which proves the lemma. ■

5.3 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 random instance Φ of the MAX k-CSP $_q(P)$ problem. To generate a random instance with m constraints, for every constraint C_i, we randomly select a k-tuple of distinct variables $T_i = \{x_{i_1}, \ldots, x_{i_k}\}$ and $a_{i_1}, \ldots, a_{i_k} \in [q]$, and put $C_i \equiv P(x_{i_1} + a_{i_1}, \ldots, x_{i_k} + a_{i_k})$. It is well known and used in various works on integrality gaps and proof complexity (e.g. [BOGH+03], [AAT05], [STT07a] and [Sch08]), that random instances of CSPs are both highly unsatisfiable and highly expanding. We capture the properties we need in the lemma below (a proof is provided in the appendix).

Lemma 5.8 *Let $\varepsilon, \delta > 0$ and a predicate $P : [q]^k \to \{0,1\}$ be given. Then there exist $\gamma = O(q^k \log q/\varepsilon^2)$, $\eta = \Omega((1/\gamma)^{10/\delta})$ and $N \in \mathbb{N}$, such that if $n \geq N$ and Φ is a random instance of* MAX k-CSP *(P) with $m = \gamma n$ constraints, then with probability* $1 - o(1)$

1. $\mathsf{OPT}(\Phi) \leq \frac{|P^{-1}(1)|}{q^k}(1+\varepsilon) \cdot m$.
2. *For any set C of constraints with $|C| \leq \eta n$, we have $|\partial(C)| \geq (k-2-\delta)|C|$.*

Let Φ be an instance of MAX k-CSP$_q$ on n variables for which G_Φ is $(\eta n, k-2-\delta)$ boundary expanding for some $\delta < 1/2$, as in Lemma 5.8. For such a Φ, we now define the distributions $\mathcal{D}(S)$.

For a set S of size at most $t = \eta\delta n/4k$, let $\overline{S}$ be subset of variables output by the algorithm Advice when run with input S and parameters $r = \eta n, e_1 = (k-2-\delta), e_2 = (k-2-2\delta)$ on the graph G_Φ. Theorem 5.5 shows that

$$|\overline{S}| \;\leq\; (k-2-\delta)|S|/\delta \;\leq\; \eta n/4.$$

We then use (5.1) to define the distribution $\mathcal{D}(S)$ for sets S of size at most $\delta\eta n/4k$ as

$$\mathop{\mathbb{P}}_{\mathcal{D}(S)}[\alpha] \quad = \quad \sum_{\substack{\beta \in [q]^{\overline{S}} \\ \beta(S) = \alpha}} \mathop{\mathbb{P}}_{\mathcal{P}_\mu(\overline{S})}[\beta].$$

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

Claim 5.9 *Let the distributions $\mathcal{D}(S)$ be defined as above. Then for any two sets $S_1 \subseteq S_2 \subseteq [n]$ with $|S_2| \leq t = \eta\delta n/4k$, the distributions $\mathcal{D}(S_1), \mathcal{D}(S_2)$ are equal on S_1.*

Proof: The distributions $\mathcal{D}(S_1), \mathcal{D}(S_2)$ are defined according to $\mathcal{P}_\mu(\overline{S}_1)$ and $\mathcal{P}_\mu(\overline{S}_2)$ respectively. To prove the claim, we show that $\mathcal{P}_\mu(\overline{S}_1)$ and $\mathcal{P}_\mu(\overline{S}_2)$ are equal to the distribution $\mathcal{P}_\mu(\overline{S}_1 \cup \overline{S}_2)$ on $\overline{S}_1, \overline{S}_2$ respectively (note that it need not be the case that $\overline{S}_1 \subseteq \overline{S}_2$).

Let $S_3 = \overline{S}_1 \cup \overline{S}_2$. Since $|\overline{S}_1|, |\overline{S}_2| \leq \eta n/4$, we have $|S_3| \leq \eta n/2$ and hence $|C(S_3)| \leq \eta n/2$. Also, by Theorem 5.5, we know that both $G|_{-\overline{S}_1}$ and $G|_{-\overline{S}_2}$ are $(2\eta n/3, k-2-2\delta)$ boundary expanding.

Thus, using Lemma 5.6 for the pairs $(\overline{S}_1, S_3)$ and $(\overline{S}_2, S_3)$, we get that

$$\begin{aligned}
\mathop{\mathbb{P}}_{\mathcal{D}(S_1)}[\alpha_1] &= \sum_{\substack{\beta_1 \in [q]^{\overline{S}_1} \\ \beta_1(S_1)=\alpha_1}} \mathop{\mathbb{P}}_{\mathcal{P}_\mu(\overline{S}_1)}[\beta_1] \\
&= \sum_{\substack{\beta_3 \in [q]^{S_3} \\ \beta_3(S_1)=\alpha_1}} \mathop{\mathbb{P}}_{\mathcal{P}_\mu(S_3)}[\beta_3] \\
&= \sum_{\substack{\beta_2 \in [q]^{\overline{S}_2} \\ \beta_2(S_1)=\alpha_1}} \mathop{\mathbb{P}}_{\mathcal{P}_\mu(\overline{S}_2)}[\beta_2] \\
&= \sum_{\substack{\alpha_2 \in [q]^{S_2} \\ \alpha_2(S_1)=\alpha_1}} \mathop{\mathbb{P}}_{\mathcal{D}(S_2)}[\alpha_2]
\end{aligned}$$

which shows that $\mathcal{D}(S_1)$ and $\mathcal{D}(S_2)$ are equal on S_1. ■

It is now easy to prove the main result.

Theorem 5.10 *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 relaxation obtained by cn levels of the Sherali-Adams hierarchy is at least $\dfrac{q^k}{|P^{-1}(1)|} - \zeta$.*

Proof: We take $\varepsilon = \zeta/q^k, \delta = 1/4$ and consider a random instance Φ of MAX k-CSP (P) with $m = \gamma n$ as given by Lemma 5.8. Thus, $\mathsf{OPT}(\Phi) \leq \frac{|P^{-1}(1)|}{q^k}(1+\varepsilon)\cdot m$.

On the other hand, by Claim 5.9 we can define distributions $\mathcal{D}(S)$ over every set of at most $\delta\eta n/4k$ variables such that for $S_1 \subseteq S_2$, $\mathcal{D}(S_1)$ and $\mathcal{D}(S_2)$ are consistent over S_1. By Lemma 5.3 this gives a feasible solution to the LP obtained by $\delta\eta n/4k$ levels. Also, by definition of $\mathcal{D}(S)$, we have that $\mathbb{P}_{\mathcal{D}(S)}[\alpha] > 0$ only if α satisfies all constraints in $C(S)$. Hence, the value of $\mathsf{FRAC}(\Phi)$ is given by

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

Thus, the integrality gap after $\delta\eta n/4k$ levels is at least

$$\frac{\mathsf{FRAC}(\Phi)}{\mathsf{OPT}(\Phi)} = \frac{q^k}{|P^{-1}(1)|(1+\varepsilon)} \geq \frac{q^k}{|P^{-1}(1)|} - \zeta.$$

■

Chapter 6

Reductions in the Lasserre Hierarchy

Taking the analogy of the hierarchies as a computational model a bit further, in this chapter we study if integrality gaps for one problem can be used to show an integrality gap for another problem using reductions (see proof overview for the structure of reductions for integrality gaps). We remark that (simple) reductions between integrality gaps were considered before, for example by [AAT05] and [Sch08] - we develop techniques to do somewhat more complicated ones. We use the arguments from known reductions in hardness of approximation literature, which usually start from the hardness of a constraint satisfaction problem and use it to conclude the hardness of another problem. Using the integrality gaps proved in Chapter 4, we derive integrality gaps for the following problems:

1. Maximum Independent Set: This is the problem of finding the largest set of vertices in a graph, not containing any edges. The best known approximation algorithm by Boppana and Haldórsson [BH92] achieves an approximation ratio of $O\left(\frac{n}{(\log n)^2}\right)$. Also, Feige [Fei97] showed that the integrality gap for the Lovász ϑ-function, which is an SDP relaxation equivalent to 1 round of Lasserre, is at least $n/2^{c\sqrt{\log n}}$ for some constant c.

2. Approximate Graph Coloring: This is the problem of coloring a graph (with different colors for adjacent vertices) with minimum number of colors, when the graph is *known* to be colorable with a small constant number of colors. The best known algorithm, due to Chlamtac [Chl07] colors a 3-colorable graph with at most $n^{0.2072}$ colors.

3. Chromatic Number: This is the general problem of finding the minimum number of colors for coloring a graph when no guarantee as the one above is given. For this problem, Feige, Langberg and Schechtman [FLS04] show that a gap can be as large as $n/\text{polylog}(n)$ for an SDP relaxation which is weaker than 1 round of Lasserre.

4. Minimum Vertex Cover: As discussed before, in this problem, it is required to find the smallest possible subset of vertices in a graph, which touches every edge. Integrality gap of a factor $7/6$ for $\Omega(n)$ levels of the Lovász-Schrijver SDP hierarchy was discussed in Chapter 2. These were later strengthened to the Lasserre hierarchy by the result of Schoenebeck [Sch08]. An

integrality gap of factor $2 - \varepsilon$ was also shown by for $\Omega(\sqrt{\log n / \log\log n})$ levels of the Lovász-Schrijver hierarchy by Georgiou et. al. [GMPT07].

Our Results

We develop techniques to carry out PCP based reductions between SDPs, to obtain gaps for the problems above. We present a summary of the known NP-hardness results and the integrality gap results we obtain, in the table below. UG-hardness denotes the hardness assuming the Unique Games Conjecture. For Approximate Graph Coloring the hardness mentions the tradeoff for a graph known to be l-colorable.

	NP-hardness	UG-hardness	Integrality Gap	No. of levels
Maximum Independent Set	$\frac{n}{2^{(\log n)^{3/4+\varepsilon}}}$ [KP06]		$\frac{n}{2^{c_1\sqrt{\log n \log\log n}}}$	$2^{c_2\sqrt{\log n \log\log n}}$
Approximate Graph Coloring	l vs. $2^{\frac{\log^2 l}{25}}$ [Kho01]		l vs. $\frac{2^{l/2}}{4l^2}$	$\Omega(n)$
Chromatic Number	$\frac{n}{2^{(\log n)^{3/4+\varepsilon}}}$ [KP06]		$\frac{n}{2^{c_1\sqrt{\log n \log\log n}}}$	$2^{c_2\sqrt{\log n \log\log n}}$
Minimum Vertex Cover	1.36 [DS05]	2 - ε [KR03]	1.36	$\Omega(n^\delta)$

It can be shown that the integrality gap for the relaxations of Maximum Independent Set and Chromatic Number number obtained by t levels of the Lasserre hierarchy is *at most* n/t after t levels. Hence the the number of levels in the above results is optimal up to the difference in the constants c_1 and c_2. We give the proof of this fact for Maximum Independent Set for the sake of illustration.

Claim 6.1 *Let* $\Phi = (V, E)$ *be an instance of the* Maximum Independent Set *problem of size n. Let* OPT(Φ) *denote the size of the largest independent set and let* FRAC(Φ) *denote the value of the* SDP *obtained by t levels of the Lasserre hierarchy. Then,* $\mathsf{FRAC}(\Phi)/\mathsf{OPT}(\Phi) \leq n/t$.

Proof: Let the SDP solution be described by vectors $\mathbf{U}_S$ for $|S| \leq t$. Since $\sum_{i\in[n]} \left|\mathbf{U}_{\{i\}}\right|^2 = \mathsf{FRAC}(\Phi)$, there exists a set S of size t such that $\sum_{i\in S} \left|\mathbf{U}_{\{i\}}\right|^2 \geq (t/n)\cdot\mathsf{FRAC}(\Phi)$. Also, the vectors $\{\mathbf{U}_{S'}\}_{S'\subseteq S}$ define a probability distribution over the independent sets of the subgraph induced by S (see Lemma 6.14 for a formal proof of this and related facts) with the expected size of the independent sets as $\sum_{i\in S} \left|\mathbf{U}_{\{i\}}\right|^2$. The proof follows by observing

$$\frac{t}{n}\cdot\mathsf{FRAC}(\Phi) \leq \sum_{i\in S}\left|\mathbf{U}_{\{i\}}\right|^2 \leq \mathsf{OPT}(\Phi)$$

since the size of the largest independent set can be at most OPT(Φ) . ■

6.1 Overview of proofs

Consider a reduction from a constraint satisfaction problem, to another problem, say Maximum Independent Set for concreteness. Starting from a CSP instance Φ, this reduction creates a graph G_Φ and one needs to argue the following two things:

- *Completeness*: If Φ has an assignment satisfying many constraints, then G_Φ has a large independent set.
- *Soundness*: If Φ has no good assignment, then G_Φ has no large independent sets.

If Φ is an integrality gap instance for an SDP, then Φ has no good assignment but has a good SDP solution. Showing that G_Φ is an integrality gap instance, amounts to making the following two claims simultaneously:

- *Vector Completeness*: Since Φ has a good SDP solution, so does G_Φ.
- *Soundness*: Since Φ has no good assignment, G_Φ has no large independent sets.

Notice that if we are using a known NP-hardness reduction, then the soundness condition is already available. Showing an integrality gap reduces to generalizing "completeness" to "vector completeness". We do this by giving various transformation, that transform an SDP solution for Φ, into one for the problem we are reducing to.

Maximum Independent Set

The transformations for independent set are conceptually the simplest and form a basis for our other results as well. We consider a graph G_Φ obtained from Φ (called the FGLSS graph), which has vertices of the form (C_i, α), where C_i is a constraint and α_i is a partial assignment to variables in C_i. Vertices corresponding to contradicting assignments are connected.

Since Φ has an SDP solution for t levels of Lasserre (say for a large t), we have vectors $\mathbf{V}_{(S,\alpha)}$ where S is a set of at most t variables and α is a partial assignment to those variables. We need to produce vectors $\mathbf{U}_{\mathcal{S}}$ where $\mathcal{S}$ is a set of vertices in the FGLSS graph. However, a set of vertices is simply a set of constraints and partial assignments to all the variables involved. Let S' be the set of all variables in all the constraints in $\mathcal{S}$ and let α' be the joint partial assignment defined by all vertices in $\mathcal{S}$ (assuming for now, that no partial assignments in $\mathcal{S}$ contradict). We take $\mathbf{U}_{\mathcal{S}} = \mathbf{V}_{(S',\alpha')}$.

The reduction (by [BGS98]) proceeds by taking products of the graph G_Φ to get G_Φ^r and randomly sampling a certain vertex-induced subgraph. It turns out to be sufficient however, to create an SDP solution for G_Φ^r. Each vertex in G_Φ^r is an r-tuple of vertices in G_Φ with an edge between two r-tuples if vertices in any of the r coordinates are adjacent in G_Φ. A set $\overline{S}$ of vertices in G_Φ^r is a set of r-tuples and we consider sets $\mathcal{S}_1, \ldots, \mathcal{S}_r$ where $\mathcal{S}_j$ is the projection of $\overline{S}$ to the jth coordinate. For G_Φ^r, we simply take $\mathbf{U}_{\overline{S}} = \mathbf{U}_{\mathcal{S}_1} \otimes \ldots \otimes \mathbf{U}_{\mathcal{S}_r}$. This corresponds to the intuition that every independent set in G_Φ^r corresponds to picking one independent set in each copy of G_Φ.

Approximate Graph Coloring and Chromatic Number

To obtain a gap for Approximate Graph Coloring, we modify the FGLSS reduction slightly. The gap for Chromatic Number is derived from this by taking graph products and tensoring vectors as before. The modified reduction below is in the spirit of randomized PCPs of Feige and Killian [FK98].

Consider an instance Φ when the constraints are known to be of the type $A \cdot (x_1, \ldots, x_k)^{\mathsf{T}} = b$ and consider G_Φ as before. Supposing that we had an assignment satisfying all constraints, this would give a large independent set. For the graph to be l colorable, we need l independent sets covering the graph. Let l be the nullity of the matrix A and consider the vectors $w_1, \ldots, w_l$ such that $A \cdot w_l = 0$. If α is a partial assignment to variables $(x_1, \ldots, x_k)$ which satisfies the above constraint, then so is $\alpha + w_j$ for $1 \le j \le l$. [1]

The problem is this does not give us an independent set. If $x_1, \ldots, x_n$ was an assignment, and we looked at the restriction of this assignment to every constraint and added w_j to every restriction, this does not give a consistent assignment to $x_1, \ldots, x_n$. However, this *is* an independent set if we slightly modify the FGLSS graph. Every constraint C_i is on an *ordered tuple* T_i of vertices. For assignment α_1 to T_{i_i} and α_2 to T_{i_2}, we connect them only if α_1, α_2 differ on a variable in the *same coordinate* in T_{i_1}, T_{i_2}. One can then verify that the transformation above takes us from one independent set to another. Changing the graph may affect the soundness, but it is possible to choose Φ so that the resulting graph still has no large independent sets.

Using this intuition, we now need to produce a vector every for small set of vertices and every assignment of colors to the same. For a vertex (C_i, α) in G_Φ, we take the vectors corresponding to the l colors as $\mathbf{V}_{(T_i, \alpha+w_1)}, \ldots, \mathbf{V}_{(T_i, \alpha+w_l)}$. The vectors for sets can be created similarly.

Minimum Vertex Cover

For Minimum Vertex Cover, we use the reduction by Dinur and Safra [DS05], which is a little complicated to describe. However, an interesting point comes up in analyzing a (small) step which is not "local", as opposed to all the steps in the previous reductions. The step is a simple Chernoff bound in the completeness part, showing how large independent sets intersect certain other sets of vertices in a graph. Generalizing this to "vector-completeness" seems to require analyzing the "local distributions" defined by the vectors, and combining the bounds globally using properties of the SDP solution. Another component of the proof is a transformation of vectors for long-code based reductions.

6.2 Integrality Gap for Maximum Independent Set

To obtain the integrality gaps for Maximum Independent Set we use the reductions by Feige et. al. [FGL+91] and by Bellare, Goldreich and Sudan [BGS98]. However, before getting to the proof of the integrality gap, we describe how to transform vectors for a general FGLSS reduction. This transformation shall be useful for our other results as well.

6.2.1 Vectors for products of the FGLSS graph

Let Φ be an instance of MAX k-CSP with constraints $C_1, \ldots, C_m$ on tuples $T_1, \ldots, T_m$ and the domain of variables as $\{0, 1\}$. Assume that each constraint has exactly l satisfying assignments. We

[1] A reader familiar with [FK98] may recognize this as an attempt to get a zero-knowledge protocol for showing that Φ is satisfiable. However, we do not argue the soundness of this protocol - we simply enforce it in the choice of Φ.

describe below the reduction by [BGS98] from Φ to an independent set problem.

1. Given Φ, create the FGLSS graph $G_\Phi = (V_\Phi, E_\Phi)$ with a vertex for every constraint C_i and every partial assignment to variables in the corresponding tuple T_i which satisfies the constraint C_i. Two vertices $(C_{i_1}, \alpha_1), (C_{i_2}, \alpha_2)$ are connected if α_1 and α_2 assign different values to some variable. Formally

$$\begin{aligned} V_\Phi &= \{(C_i, \alpha) \mid \alpha \in \{0,1\}^{T_i},\ C_i(\alpha) = 1\} \\ E_\Phi &= \{\{(C_{i_1}, \alpha_1), (C_{i_2}, \alpha_2)\} \mid \alpha_1(T_{i_1} \cap T_{i_2}) \neq \alpha_2(T_{i_1} \cap T_{i_2})\} \end{aligned}$$

2. Construct the product graph $G_\Phi^r = (V_\Phi^r, E')$ with vertices of G_Φ^r being r-tuples of vertices in G_Φ. Two vertices $\{(C_{i_1}, \alpha_1), \ldots, (C_{i_r}, \alpha_r)\}$ and $\{(C_{i'_1}, \alpha'_1), \ldots, (C_{i'_r}, \alpha'_r)\}$ are connected if for some j, $\{(C_{i_j}, \alpha_j), (C_{i'_j}, \alpha'_j)\} \in E_\Phi$.

Note that if Φ had m constraints, then G_Φ^r has $l^r \cdot m^r$ vertices, with there being m^r disjoint cliques of l^r vertices, corresponding to every r-tuple of constraints. We denote the clique corresponding to constraints $C_{i_1}, \ldots, C_{i_r}$ as $C(i_1, \ldots, i_r)$. Formally,

$$C(i_1, \ldots, i_r) = \left\{\{(C_{i_1}, \alpha_1), \ldots, (C_{i_r}, \alpha_r)\} \mid \ \wedge_{j=1}^r C_{i_j}(\alpha_j) = 1\right\}$$

The largest independent set in G_Φ^r can have at most m^r vertices. We claim that a good SDP solution for Φ can be transformed into a good solution for the independent set SDP on G_Φ^r.

Lemma 6.2 *Let Φ be an instance of* MAX k-CSP *as above with m constraints. If* $\mathsf{FRAC}(\Phi) = m$ *after t levels of the Lasserre hierarchy, then* $\mathsf{FRAC}(G_\Phi^r) \geq m^r$ *for the independent set SDP obtained after t/k levels. Moreover, the contribution to the SDP value from vertices in each clique $C(i_1, \ldots, i_r)$ is 1.*

Proof: We first define an independent set solution for t/k levels on G_Φ and then show how to extend it to G_Φ^r. Consider a set $\mathcal{S}$ of $h \leq t/k$ vertices in G_Φ. It is specified by h constraints and partial assignments $\{(C_{i_1}, \alpha_1), \ldots, (C_{i_h}, \alpha_h)\}$. Define $\mathbf{U}_\mathcal{S}$ as

$$\mathbf{U}_\mathcal{S} = \begin{cases} 0 & \exists j_1, j_2 \leq h \text{ s.t. } \alpha_{j_1}(T_{i_{j_1}} \cap T_{i_{j_2}}) \neq \alpha_{j_2}(T_{i_{j_1}} \cap T_{i_{j_2}}) \\ \mathbf{V}_{(\cup_j T_{i_j}, \alpha_1 \circ \ldots \circ \alpha_h)} & \text{otherwise} \end{cases}$$

We now consider a set $\overline{S}$ of vertices in G_Φ^r. It is a *set of r-tuples* of vertices in G_Φ. Let $\mathcal{S}_j$ denote the set of vertices of G_Φ which occur in the jth coordinate of the r-tuples in $\overline{S}$. Define the vector $\mathbf{U}_{\overline{S}}$ as

$$\mathbf{U}_{\overline{S}} = \mathbf{U}_{\mathcal{S}_1} \otimes \ldots \otimes \mathbf{U}_{\mathcal{S}_r}$$

Let $\mathbf{U}_{\overline{\emptyset}}$ denote the vector for the empty set of vertices in G_Φ^r. We take $\mathbf{U}_{\overline{\emptyset}} = \mathbf{U}_\emptyset \otimes \ldots \otimes \mathbf{U}_\emptyset$. The vectors $\mathbf{U}_{\overline{S}}$ are defined for all sets $\overline{S}$ with at most t/k vertices. We now show that they satisfy all Lasserre constraints.

Claim 6.3 *The vectors $\mathbf{U}_{\overline{S}}$ satisfy all conditions of the (t/k)-round independent set SDP on G_Φ^r.*

Proof: Since all vectors $\mathbf{U}_{\overline{S}}$ are tensors of valid Lasserre vectors for the SDP for Φ, all inner products are between 0 and 1. We only need to verify that the vectors corresponding to two vertices connected by an edge are orthogonal, and that $\left\langle \mathbf{U}_{\overline{S}_1}, \mathbf{U}_{\overline{S}_2} \right\rangle$ depends only on $\overline{S}_1 \cup \overline{S}_2$.

- Consider two vertices $\{(C_{i_1}, \alpha_1), \ldots, (C_{i_r}, \alpha_r)\}$ and $\{(C_{i'_1}, \alpha'_1), \ldots, (C_{i'_r}, \alpha'_r)\}$ connected by an edge. The corresponding vectors are $\mathbf{V}_{(T_{i_1},\alpha_1)} \otimes \ldots \otimes \mathbf{V}_{(T_{i_r},\alpha_r)}$ and $\mathbf{V}_{(T_{i'_1},\alpha'_1)} \otimes \ldots \otimes \mathbf{V}_{(T_{i'_r},\alpha'_r)}$. The fact that there is an edge between the vertices means that for some $j \le r$, $\alpha_j(T_{i_j} \cap T_{i'_j}) \neq \alpha'_j(T_{i_j} \cap T_{i'_j})$. Hence $\left\langle \mathbf{V}_{(T_{i_j},\alpha_j)}, \mathbf{V}_{(T_{i'_j},\alpha'_j)} \right\rangle = 0$ since the vectors $\mathbf{V}_{(\cdot,\cdot)}$ form a valid Lasserre solution. This gives

$$\left\langle \mathbf{V}_{(T_{i_1},\alpha_1)} \otimes \ldots \otimes \mathbf{V}_{(T_{i_r},\alpha_r)}, \mathbf{V}_{(T_{i'_1},\alpha'_1)} \otimes \ldots \otimes \mathbf{V}_{(T_{i'_r},\alpha'_r)} \right\rangle = \prod_{j=1}^{r} \left\langle \mathbf{V}_{(T_{i_j},\alpha_j)}, \mathbf{V}_{(T_{i'_j},\alpha'_j)} \right\rangle = 0$$

- Next, consider sets $\overline{S}_1, \overline{S}_2, \overline{S}_3, \overline{S}_4$ such that $\overline{S}_1 \cup \overline{S}_2 = \overline{S}_3 \cup \overline{S}_4$. For $1 \le u \le 4$, let $\mathcal{S}_j^{(u)}$ denote the union of elements in the jth coordinate of the r-tuples in $\overline{S}_u$. $\overline{S}_1 \cup \overline{S}_2 = \overline{S}_3 \cup \overline{S}_4$ means that in particular $\mathcal{S}_j^{(1)} \cup \mathcal{S}_j^{(2)} = \mathcal{S}_j^{(3)} \cup \mathcal{S}_j^{(4)}$ for all $1 \le j \le r$. For a fixed j, let $\mathcal{S}_j^{(1)} \cup \mathcal{S}_j^{(2)} = \mathcal{S}_j^{(3)} \cup \mathcal{S}_j^{(4)} = \{(C_{i_1}, \alpha_1), \ldots, (C_{i_h}, \alpha_h)\}$. If the set contains two contradicting partial assignments, then either one of $\mathbf{U}_{\mathcal{S}_1}^{(j)}$ and $\mathbf{U}_{\mathcal{S}_2}^{(j)}$ is 0, or they are equal to Lasserre vectors corresponding to contradicting partial assignments. In either case $\left\langle \mathbf{U}_{\mathcal{S}_1}^{(j)}, \mathbf{U}_{\mathcal{S}_2}^{(j)} \right\rangle = 0$ and similarly $\left\langle \mathbf{U}_{\mathcal{S}_3}^{(j)}, \mathbf{U}_{\mathcal{S}_4}^{(j)} \right\rangle = 0$. If there are no contradicting partial assignments, then the tuples in $\mathcal{S}_1^{(j)} \cup \mathcal{S}_2^{(j)}$ can be extended to a unique partial assignment $\alpha_1 \circ \ldots \circ \alpha_h$ over $\cup_{j=1}^{h} T_{j_j}$. Since the set of all tuples, and hence the assignment, is same for $\mathcal{S}_3^{(j)} \cup \mathcal{S}_4^{(j)}$, and the corresponding CSP vectors are consistent, we get $\left\langle \mathbf{U}_{\mathcal{S}_1}^{(j)}, \mathbf{U}_{\mathcal{S}_2}^{(j)} \right\rangle = \left\langle \mathbf{U}_{\mathcal{S}_3}^{(j)}, \mathbf{U}_{\mathcal{S}_4}^{(j)} \right\rangle$ for all j, which implies $\left\langle \mathbf{U}_{\overline{S}_1}, \mathbf{U}_{\overline{S}_2} \right\rangle = \left\langle \mathbf{U}_{\overline{S}_3}, \mathbf{U}_{\overline{S}_4} \right\rangle$.

■

We show that the value for all the vertices in any clique $C(i_1, \ldots, i_r)$ is 1. Letting $\alpha_1, \ldots, \alpha_r$ range over all satisfying assignments to $C_{i_1}, \ldots, C_{i_j}$, the contribution of vertices in this clique to the SDP objective value is

$$\sum_{\alpha_1,\ldots,\alpha_r} \prod_{j=1}^{r} \left\langle \mathbf{V}_{(T_{i_j},\alpha_j)}, \mathbf{V}_{(\emptyset,\emptyset)} \right\rangle = \prod_{j=1}^{r} \left\langle \sum_{\alpha_j} \mathbf{V}_{(T_{i_j},\alpha_j)}, \mathbf{V}_{(\emptyset,\emptyset)} \right\rangle = \prod_{j=1}^{r} (1) = 1$$

where $\left\langle \sum_{\alpha_j} \mathbf{V}_{(T_{i_j},\alpha_j)}, \mathbf{V}_{(\emptyset,\emptyset)} \right\rangle = 1$ since the contribution of the constraint C_{i_j} to the SDP for MAX k-CSP is 1. ■

6.2.2 Obtaining the Integrality Gap

We can now prove the following integrality gap for Maximum Independent Set.

Theorem 6.4 *There exist constants $c_1, c_2 > 0$ and graphs on N vertices for arbitrarily large N, such that the integrality gap for the SDP for independent set obtained by $2^{c_2 \sqrt{\log N \log\log N}}$ levels of the Lasserre hierarchy, is at least $\dfrac{N}{2^{c_1 \sqrt{\log N \log\log N}}}$.*

Proof: Our integrality gap instance will be a subgraph of G_Φ^r for appropriate choices of Φ and r. We construct the graph $G = (V, E)$ by randomly picking M cliques of the form $C(i_1, \ldots, i_r)$, and taking G to be the subgraph induced by the vertices in these cliques. An easy Chernoff bound shows that if only a small fraction of constraints in Φ were satisfiable, then the size of the largest independent set in G is small.

Claim 6.5 *Let $s = \mathsf{OPT}(\Phi)/m$. Then for $M \geq \frac{100nr}{s^r}$, with probability $1 - o(1)$, all independent sets in G have size at most $2s^r M$.*

Proof: It is easy to see that any independent set in G_Φ can be extended to an assignment to the variables $x_1, \ldots, x_n$ and has size equal to the number of constraints in Φ satisfied by the assignment. Hence, the size of the largest independent set in G_Φ is at most $s \cdot m$. Also, an independent set in G_Φ^r is a set of r-tuples of vertices in G_Φ such that if we consider the set of vertices in the jth coordinate for any j, they form an independent set in G_Φ. Hence, any independent set in G_Φ^r has size at most $(s \cdot m)^r$. Also, note that since an independent set of G_Φ^r can be extended to an assignment to $x_1, \ldots, x_n$ in each of the r coordinates, there are at most 2^{nr} different independent sets.

Any independent set in the sampled graph G also extends to an assignment to $x_1, \ldots, x_n$ in each coordinate and can be thought of as the intersection of an independent set I of G_Φ^r with the sampled blocks. Fix any independent set I of G_Φ^r. We sample M out of the m^r blocks in G_Φ and each block has at most one vertex belonging to I (because each block is a clique). Hence, by Chernoff bounds, the probability that more than $2s^r \cdot M$ vertices in G belong to I is at most $\exp(-s^r \cdot M/50)$. Taking a union bound over all choices of I, we get that with probability at least $1 - \exp(-s^r \cdot M/50 + nr)$, all independent set of G have size at most $2s^r \cdot M$. Choosing $M \geq 100nr/s^r$ ensures that the probability is $1 - o(1)$. ■

We now make the choices for all the parameters involved. For a large n, let $k = \delta \log n$ for some small constant δ, and let $r = \log n/(\log \log n)$. Consider an instance Φ of $\mathsf{MAX\ k\text{-}CSP}$ as given by Corollary 4.19. By choosing $\varepsilon = 1/2$, we can get that $k/2^k \leq s \leq 3k/2^k$. Also, since the constraints are based on the Hamming code, the number of satisfying assignments to each constraint is at most $l \leq 2k$. We pick $M = 100nr \cdot (2^{kr}/k^r)$.

By the previous claim, the size of the maximum independent set in G is at most $2Ms^r$ (w.h.p. over the choice of G). We take the SDP solution to be the same as constructed for G_Φ^r. By Lemma 6.2, the contribution of the vectors in each clique to the SDP value is 1. Hence, the value of the SDP solution for G is M, which gives an integrality gap of $(1/2s^r) \geq (1/2) \cdot (2^k/3k)^r$. On the other hand, the number of vertices in G is

$$N = M \cdot l^r \leq (100nr \cdot 2^{kr}/k^r) \cdot (2k)^r = O(nr \cdot 2^{(k+1)r})$$

With our choice of parameters, the integrality gap is at least $\dfrac{N}{2^{c_1 \sqrt{\log N \log \log N}}}$ for some constant c_1.

To verify the number of levels, note that Corollary 4.19 gives $\beta = O(2^k) = O(n^\delta)$ and $c = \Omega((1/n^\delta)^{25})$. Hence, we have SDP solutions for $cn = \Omega(n^{1-25\delta})$ levels for Φ and consequently for $\Omega(n^{1-25\delta}/k)$ levels for the independent set SDP on G. For $\delta < 1/25$, this is at least $2^{c_2 \sqrt{\log N \log \log N}}$ for some constant c_2. ■

6.3 Gaps for Graph Coloring

In this section, we show that SDPs in the Lasserre hierarchy fail to approximate the chromatic number of a graph. Gaps for chromatic number are syntactically different from the usual integrality gaps for SDPs because the value of the chromatic number is not a linear function of the inner products of vectors in an SDP. Instead for any l, one can write down an SDP for which a *feasible solution* gives a *vector l-coloring* of the graph. We show graphs for which an l-coloring remains feasible even after many levels of the Lasserre hierarchy, even though the actual chromatic number of the graph is much larger than l.

We show that for any constant l, there are graphs with chromatic number at least $\frac{2^{l/2}}{4l^2}$ which admit a vector l-coloring even after $\Omega(n)$ levels of the Lasserre hierarchy. For Chromatic Number, we show that the ratio of the chromatic number of the graph and the number of colors in the vector coloring obtained by $2^{\Omega(\sqrt{\log n \log\log n})}$ levels of the Lasserre hierarchy can be as high as $\frac{n}{2^{O(\sqrt{\log n \log\log n})}}$. We write the Lasserre SDP for l-coloring as a constraint satisfaction problem, with the additional restriction that *all* the constraints which say colors of two adjacent vertices must be different, are satisfied. This formulation is equivalent to the one considered by Chlamtac [Chl07]. [2] To avoid confusion with the SDP for MAX k-CSP we denote the sets of vertices here by $\mathcal{S}$, partial assignments by γ and vectors for coloring by $\overline{\mathbf{V}}_{(\mathcal{S},\gamma)}$.

$$\text{Minimize } l \qquad \text{s.t. there exist vectors } \overline{\mathbf{V}}_{(\mathcal{S},\gamma)} \text{ for all } |\mathcal{S}| \le t, \gamma \in [l]^{\mathcal{S}} \text{ satisfying}$$

$$\begin{aligned}
\left\langle \overline{\mathbf{V}}_{(\{u_1\},\gamma)}, \overline{\mathbf{V}}_{(\{u_2\},\gamma)} \right\rangle &= 0 && \forall (u_1,u_2) \in E, \gamma \in [l] \\
\left\langle \overline{\mathbf{V}}_{(\mathcal{S}_1,\gamma_1)}, \overline{\mathbf{V}}_{(\mathcal{S}_2,\gamma_2)} \right\rangle &= 0 && \forall\ \gamma_1(\mathcal{S}_1 \cap \mathcal{S}_2) \neq \gamma_2(\mathcal{S}_1 \cap \mathcal{S}_2) \\
\left\langle \overline{\mathbf{V}}_{(\mathcal{S}_1,\gamma_1)}, \overline{\mathbf{V}}_{(\mathcal{S}_2,\gamma_2)} \right\rangle &= \left\langle \overline{\mathbf{V}}_{(\mathcal{S}_3,\gamma_3)}, \overline{\mathbf{V}}_{(\mathcal{S}_4,\gamma_4)} \right\rangle && \forall\ \mathcal{S}_1 \cup \mathcal{S}_2 = \mathcal{S}_3 \cup \mathcal{S}_4,\ \gamma_1 \circ \gamma_2 = \gamma_3 \circ \gamma_4 \\
\sum_{j \in [l]} \left| \overline{\mathbf{V}}_{(\{u\},j)} \right|^2 &= 1 && \forall u \in V \\
\left\langle \overline{\mathbf{V}}_{(\mathcal{S}_1,\gamma_1)}, \overline{\mathbf{V}}_{(\mathcal{S}_2,\gamma_2)} \right\rangle &\ge 0 && \forall \mathcal{S}_1, \mathcal{S}_2, \gamma_1, \gamma_2 \\
\left| \overline{\mathbf{V}}_{(\emptyset,\emptyset)} \right| &= 1
\end{aligned}$$

Figure 6.1: Level-t Lasserre SDP for l-coloring of a graph

The reduction we describe in this section is a slightly modified version of the reduction for independent set and is specifically designed to work for problems of the type MAX k-CSP (P_A), where the constraints being linear equations in $\mathbb{F}_2$. It is inspired by what could be a "zero-knowledge protocol" for such predicates, in the sense of Feige and Killian [FK98]. Here, we describe the reduction without going through the protocol, at the cost of defining the following additional (and somewhat unnatural) solution concept for a CSP instance.

Definition 6.6 *Let Φ be an instance of* MAX k-CSP *with constraints $C_1, \ldots, C_m$ on ordered k-tuples $T_1, \ldots, T_m$ of variables. For a constraint C_i on variables $(x_{i_1}, \ldots, x_{i_k})$, we say that the constraint*

[2] In fact, the SDPs in Lasserre hierarchy are fairly independent of the representation used. It is easy to switch between different SDPs for a problem, by losing at most a constant factor in the number of levels.

k-satisfied *by assignments* $\Pi_1, \ldots, \Pi_k$ *if* $C_i(\Pi_1(x_{i_1}), \ldots, \Pi_k(x_{i_k})) = 1$. *We denote by* $\mathsf{OPT}_k(\Phi)$*, the maximum number of constraints in* Φ *that are k-satisfied by any assignments* $\Pi_1, \ldots, \Pi_k$.

Note that the above definition crucially uses the fact that the constraint is defined on an *ordered tuple* of variables as we read the value of the first variable from Π_1, the second from Π_2 and so on. By slightly strengthening the notion of unsatisfiability for a random CSP instance in Lemma 4.2, we can strengthen Corollary 4.19 as below.

Corollary 6.7 *Let a number k and* $\varepsilon > 0$ *be given and let A be the generator matrix for the Hamming code of length k. Then there exist* $\beta = O(k2^k/\varepsilon^2)$ *and* $c = \Omega((1/\beta)^{25})$ *such that if* Φ *is a random instance of* MAX k-CSP (P_A) *on* $n \gg 1/c$ *variables and* $m = \beta n$ *constraints, then with probability* $1 - o(1)$

1. $\mathsf{OPT}_k(\Phi) \leq \frac{2k}{2^k}(1+\varepsilon)\cdot m$
2. *For the SDP given by cn levels of the Lasserre hierarchy,* $\mathsf{FRAC}(\Phi) = m$.

6.3.1 Gaps for Approximate Graph Coloring

We reduce from a CSP instance Φ as in Corollary 6.7. For an instance Φ of MAX k-CSP (P_A), consider the vectors $w \in \{0,1\}^k$ such that $A \cdot w^{\mathsf{T}} = 0$ over $\mathbb{F}_2$. If A is the generator matrix of the Hamming code of length k, there are $2^{\lceil \log(k+1) \rceil}$ such vectors. We shall show that the graph produced by our reduction has a vector coloring with $l = 2^{\lceil \log(k+1) \rceil}$ colors, where we shall identify the domain $[l]$ of the coloring CSP with the vectors $w_1, \ldots, w_l$ satisfying $A \cdot w_j^{\mathsf{T}} = 0$.

We now give the reduction from Φ as above, to Approximate Graph Coloring. Similar to the case of independent set, we create the FGLSS graph with a vertex for every constraint and every satisfying partial assignment to the variables in that constraint. However, we have fewer edges: we connect two vertices (C_{i_1}, α_1) and (C_{i_2}, α_2) iff α_1 and α_2 disagree on a variable that occurs *at the same position* in the ordered tuples T_{i_1} and T_{i_2}. Formally, we create the graph $G_\Phi = (V_\Phi, E_\Phi)$ such that

$$V_\Phi = \{(C_i, \alpha) \mid \alpha \in \{0,1\}^{T_i},\ C_i(\alpha) = 1\}$$
$$E_\Phi = \{\{(C_{i_1}, \alpha_1), (C_{i_2}, \alpha_2)\} \mid \exists 1 \leq j \leq k.\ [T_{i_1,j} = T_{i_2,j}] \wedge [\alpha_1(T_{i_1,j}) \neq \alpha_2(T_{i_2,j})]\}$$

where $T_{i,j}$ is used to denote the variable in the jth position in the ordered tuple corresponding to the ith constraint. To show that G_Φ has large chromatic number, we claim that all independent sets in G_Φ are small.

Claim 6.8 *The size of the maximum independent set in* G_Φ *is* $\mathsf{OPT}_k(\Phi)$.

Proof: Let I be an independent set in Φ. Hence I is a set of pairs of the form (C_i, α), where C_i is a constraint and α is a partial assignment giving values for variables in T_i. Since all vertices corresponding to a single constraint are connected, I can include at most one vertex corresponding to one constraint.

Consider the values given to all the variables $x_1, \ldots, x_k$ by all the partial assignments in I, when the variable is present in the jth position in the tuple. Since all the partial assignments to constraints must be consistent in the values at the jth position, these values can be extended to a unique assignment, say $\Pi_j = (a_1, \ldots, a_n)$ to the variables $x_1, \ldots, x_n$. Similarly, we can define assignments $\Pi_1, \ldots, \Pi_k$ for each of the k positions.

Hence, the independent set corresponds to picking at most one of the satisfying assignments for every constraint, with the jth variable in the tuple set according to Π_j. This gives that the size of the largest independent set is at most $\mathsf{OPT}_k(\Phi)$. ■

Lemma 6.9 *Let Φ be an instance of* MAX k-CSP *(P_A), with m constraints such that each constraint has exactly l satisfying assignments. If* $\mathsf{FRAC}(\Phi) = m$ *after t levels of the Lasserre hierarchy, then there is a feasible solution to the SDP for l-coloring of G_Φ obtained by $t/2k$ levels of the Lasserre hierarchy.*

Proof: We now define the vectors $\overline{\mathbf{V}}_{(\mathcal{S},\gamma)}$ for a set $\mathcal{S} \subseteq V_\Phi, |\mathcal{S}| \leq t/k$ and $\gamma \in [l]^{\mathcal{S}}$. Let $(\mathcal{S}, \gamma) = (\{(C_{i_1}, \alpha_1), \ldots, (C_{i_h}, \alpha_h)\}, \gamma)$ Recall that the domain $[l]$ is identified with the vectors $w_1, \ldots, w_l \in \{0, 1\}^k$ which satisfy $A \cdot w_j^\top = 0$ for $1 \leq j \leq l$. Hence the partial assignment γ assigns a *vector* in $\mathbb{F}_2^k$ to each vertex (C_{i_j}, α_j). We use the vectors given by γ to modify the assignments to each C_{i_j}. This can be viewed as the zero-knowledge step of randomizing over all the satisfying assignments to each constraint. Formally, we change α_j to $\alpha_j + \gamma((C_{i_j}, \alpha_j))$ where $\gamma((C_{i_j}, \alpha_j))$ is the vector in $\mathbb{F}_2^k$ (the "color") assigned by γ to the vertex (C_{i_j}, α_j) and the '+' is over $\mathbb{F}_2$. Let $[\alpha_j, \gamma]$ denote this assignment to T_{i_j} which is shifted by γ.

With this interpretation, we define the vectors as 0 if these shifted partial assignments contradict, and otherwise as the Lasserre vectors corresponding to the assignment defined collectively by all the shifted assignments. For all $|\mathcal{S}| \leq t/k$ and $\gamma \in [l]^{\mathcal{S}}$, we define

$$\overline{\mathbf{V}}_{(\mathcal{S},\gamma)} = \begin{cases} 0 & \exists j_1, j_2.\ [\alpha_{j_1}, \gamma](T_{i_{j_1}} \cap T_{i_{j_2}}) \neq [\alpha_{j_2}, \gamma](T_{i_{j_1}} \cap T_{i_{j_2}}) \\ \mathbf{V}_{(\cup T_{i_j}, [\alpha_1,\gamma]\circ\ldots\circ[\alpha_h,\gamma])} & \text{otherwise} \end{cases}$$

We now need to verify that the vectors satisfy all the SDP conditions.

- For an edge $\{(C_{i_1}, \alpha_1), (C_{i_2}, \alpha_2)\}$, we must have that $\left\langle \overline{\mathbf{V}}_{(\{(C_{i_1},\alpha_1)\},\gamma)}, \overline{\mathbf{V}}_{(\{(C_{i_2},\alpha_2)\},\gamma)} \right\rangle = 0$. Note that if (C_{i_1}, α_1) and (C_{i_2}, α_2) have an edge, then for some j, T_{i_1} and T_{i_2} have the same variable in the jth position and α_1, α_2 disagree on that variable. Then $[\alpha_1, \gamma]$ and $[\alpha_2, \gamma]$, which are equal to $\alpha_1 + w$ and $\alpha_2 + w$ for some w in the null space of A, would also disagree on that variable. Hence, by validity of the Lasserre solution for the CSP
$$\left\langle \overline{\mathbf{V}}_{(\{(C_{i_1},\alpha_1)\},\gamma)}, \overline{\mathbf{V}}_{(\{(C_{i_2},\alpha_2)\},\gamma)} \right\rangle = \left\langle \mathbf{V}_{((T_{i_1},[\alpha_1,\gamma])}, \mathbf{V}_{(T_{i_2},[\alpha_2,\gamma])} \right\rangle = 0$$
- We next verify that $\left\langle \overline{\mathbf{V}}_{(\mathcal{S}_1,\gamma_1)}, \overline{\mathbf{V}}_{(\mathcal{S}_2,\gamma_2)} \right\rangle = 0$ whenever γ_1, γ_2 disagree on $\mathcal{S}_1 \cap \mathcal{S}_2$. The disagreement means that there is some vertex $(C_{i_j}, \alpha_j) \in \mathcal{S}_1 \cap \mathcal{S}_2$ such that $\gamma_1((C_{i_j}, \alpha_j)) \neq \gamma_2((C_{i_j}, \alpha_j))$. If T_{i_j} is the tuple of variables corresponding to C_{i_j}, then $[\alpha_j, \gamma_1](T_{i_j}) \neq [\alpha_j, \gamma_2](T_{i_j})$. Assuming neither of $\overline{\mathbf{V}}_{(\mathcal{S}_1,\gamma_1)}$ and $\overline{\mathbf{V}}_{(\mathcal{S}_2,\gamma_2)}$ is zero, we must have that $\overline{\mathbf{V}}_{(\mathcal{S}_1,\gamma_1)} = \mathbf{V}_{(S'_1,\alpha'_1)}$ and

$\overline{\mathbf{V}}_{(\mathcal{S}_2,\gamma_2)} = \mathbf{V}_{(S'_2,\alpha'_2)}$ for some $S'_1, S'_2 \subseteq [n]$ and partial assignments α'_1, α'_2. Also, we have that $T_{i_j} \subseteq \mathcal{S}_1 \cap \mathcal{S}_2$ and $\alpha'_1(T_{i_j}) = [\alpha_j, \gamma_1](T_{i_j}) \neq [\alpha_j, \gamma_2](T_{i_j}) = \alpha'_2(T_{i_j})$. This gives $\left\langle \mathbf{V}_{(S'_1,\alpha'_1)}, \mathbf{V}_{(S'_2,\alpha'_2)} \right\rangle = 0$.

- We also need to show that $\left\langle \overline{\mathbf{V}}_{(\mathcal{S}_1,\gamma_1)}, \overline{\mathbf{V}}_{(\mathcal{S}_2,\gamma_2)} \right\rangle = \left\langle \overline{\mathbf{V}}_{(\mathcal{S}_3,\gamma_3)}, \overline{\mathbf{V}}_{(\mathcal{S}_4,\gamma_4)} \right\rangle$ whenever $\mathcal{S}_1 \cup \mathcal{S}_2 = \mathcal{S}_3 \cup \mathcal{S}_4$ and $\gamma_1 \circ \gamma_2 = \gamma_3 \circ \gamma_4$. For convenience, we only show this when all sets have size at most $t/2k$ by showing that $\left\langle \overline{\mathbf{V}}_{(\mathcal{S}_1,\gamma_1)}, \overline{\mathbf{V}}_{(\mathcal{S}_2,\gamma_2)} \right\rangle = \left\langle \overline{\mathbf{V}}_{(\mathcal{S}_1 \cup \mathcal{S}_2,\gamma_1 \circ \gamma_2)}, \overline{\mathbf{V}}_{(\emptyset,\emptyset)} \right\rangle$. Again, assuming neither of these vectors are zero, let $\overline{\mathbf{V}}_{(\mathcal{S}_1,\gamma_1)} = \mathbf{V}_{(S'_1,\alpha'_1)}$ and $\overline{\mathbf{V}}_{(\mathcal{S}_2,\gamma_2)} = \mathbf{V}_{(S'_2,\alpha'_2)}$.

  If α'_1 and α'_2 contradict (note that this may happen even when $\gamma_1 \circ \gamma_2$ is defined) then there must be some vertices $(C_{i_1}, \alpha_1) \in \mathcal{S}_1$ and $(C_{i_2}, \alpha_2) \in \mathcal{S}_2$ such that C_{i_1} and C_{i_2} involve a common variable, on which the shifted assignments $[\alpha_1, \gamma_1]$ and $[\alpha_2, \gamma_2]$ disagree. But then both these vertices will also be present in $\mathcal{S}_1 \cup \mathcal{S}_2$ and the assignments shifted according to $\gamma_1 \circ \gamma_2$ will also disagree. Hence, $\overline{\mathbf{V}}_{(\mathcal{S}_1 \cup \mathcal{S}_2,\gamma_1 \circ \gamma_2)} = 0$ which satisfies the condition in this case. If not, we must have that $\overline{\mathbf{V}}_{(\mathcal{S}_1 \cup \mathcal{S}_2,\gamma_1 \circ \gamma_2)} = \mathbf{V}_{(S'_3,\alpha'_3)}$. Since the vectors $\mathbf{V}_{(\cdot,\cdot)}$ for a valid CSP solution, it will be sufficient to show that $S'_3 = S'_1 \cup S'_2$ and $\alpha'_3 = \alpha'_1 \circ \alpha'_2$. Since S'_3 contains all the variables involved in constraints present either in $\mathcal{S}_1$ or $\mathcal{S}_2$, it must include all variables in $S'_1 \cup S'_2$. Finally, for any $(C_{i_j}, \alpha_j) \in \mathcal{S}_1 \cup \mathcal{S}_2$, $\alpha'_3(T_{i_j}) = [\alpha_j, \gamma_1 \circ \gamma_2](T_{i_j}) = (\alpha'_1 \circ \alpha'_2)(T_{i_j})$, which proves the required condition.

- Finally, we need to verify that for every vertex (C_i, α) of G_Φ, $\sum_{j \in [l]} \left| \overline{\mathbf{V}}_{(\{(C_i,\alpha)\},j)} \right|^2 = 1$. Note that $\overline{\mathbf{V}}_{(\{(C_i,\alpha)\},j)} = \mathbf{V}_{(T_i,\alpha+w_j)}$ where w_j is a vector on $\mathbb{F}_2$ such that $A \cdot w_j^\top = 0$. If the constraint C_i is of the form $A \cdot x = b_i$ and if α is a satisfying assignment, then as j ranges from 1 to l, $(\alpha + w_j)$ ranges over all the satisfying assignments to the constraint C_i. Hence, we have that

$$\sum_{j \in [l]} \left| \overline{\mathbf{V}}_{(\{(C_i,\alpha)\},j)} \right|^2 = \sum_{j \in [l]} \left| \mathbf{V}_{(T_i,\alpha+w_j)} \right|^2 = \sum_{\alpha \in \{0,1\}^{T_i}} C_i(\alpha) \left| \mathbf{V}_{(T_i,\alpha)} \right|^2 = 1$$

  where the last equality used the fact that $\mathsf{FRAC}(\Phi) = m$ and hence the contribution to the SDP value, from assignments of each constraint, is 1.

■

This now gives the claimed gap for Approximate Graph Coloring.

Theorem 6.10 *For every constant l there is a $c = c(l)$ and an infinite family of graphs $G = (V, E)$ with chromatic number $\Omega\left(\frac{2^{l/2}}{l^2}\right)$, and such that G has a vector coloring with l colors for the SDP obtained by $c \cdot |V|$ levels of the Lasserre hierarchy.*

Proof: For any l, there is a k such that $l/2 \leq 2^{\lceil \log(k+1) \rceil} \leq l$. For this k, consider an instance Φ of MAX k-CSP with n variables and $m = \beta n$ constraints as given by Corollary 6.7, choosing $\varepsilon = 1/2$. We take our graph G to be G_Φ as defined above. Claim 6.8 shows that the largest independent set has size at most $\mathsf{OPT}_k(\Phi)$, which is at most $(3l/2^{k+1}) \cdot m$ by Corollary 6.7. Since the number of vertices in G (say N) is at least $k \cdot m$, its chromatic number is $\Omega(2^k/l^2) = \Omega(2^{l/2}/l^2)$.

On the other hand, we have SDP solutions for Φ for $c'n$ levels (with $c' = c'(k)$) with $\mathsf{FRAC}(\Phi) = m$. By Lemma 6.9 G has a vector coloring $2^{\lceil \log(k+1) \rceil}$ colors for the SDP obtained by $c'n/k = cN$ levels of Lasserre, where c depends only on k (which depends only on l). ■

6.3.2 Gaps for Chromatic Number

We now modify the graph and the SDP solution constructed in the previous section to get strong gaps for Chromatic Number. As in the case of independent sets, we define the product graph $G_\Phi^r = (V_\Phi^r, E')$ for G_Φ defined above. Two vertices $\{(C_{i_1}, \alpha_1), \ldots, (C_{i_r}, \alpha_r)\}$ and $\{(C_{i'_1}, \alpha'_1), \ldots, (C_{i'_r}, \alpha'_r)\}$ in V_Φ^r are connected if for some j, $\{(C_{i_j}, \alpha_j), (C_{i'_j}, \alpha'_j)\} \in E_\Phi$. Note that the edge set E_Φ is slightly different than it was in the case of independent set. $C(i_1, \ldots, i_r)$ is defined as before

$$C(i_1, \ldots, i_r) = \{\{(C_{i_1}, \alpha_1), \ldots, (C_{i_r}, \alpha_r)\} \mid \wedge_{j=1}^r C_{i_j}(\alpha_j) = 1\}$$

We argue that if G_Φ has a vector coloring with l colors, then G_Φ has a vector coloring with l^r colors. We think of the l^r colors as r-tuples of values in $[l]$. Hence, a partial assignment assigns to each vertex a tuple in $[l]^r$.

Claim 6.11 *If there is a feasible solution for the l-coloring SDP for G_Φ obtained by t levels of the Lasserre hierarchy, then there is also a feasible solution for the SDP for l^r-coloring the graph G_Φ^r obtained by t levels.*

Proof: We define the vector $\overline{\mathbf{V}}_{(\overline{S}, \overline{\gamma})}$ for all $\overline{S} \subseteq V_\Phi^r$, $|\overline{S}| \le t$ and $\overline{\gamma} \in ([l]^r)^{\overline{S}}$. Each vertex $v \in \overline{S}$ is of the form $\{((C_{i_1}, \alpha_1), \ldots, (C_{i_r}, \alpha_r))\}$. For a such vertex v, let $[v]_j$ denote the element in jth coordinate of v i.e. (C_{i_j}, α_j). Also, $\overline{\gamma}(v)$ is an r-tuple $(l_1, \ldots, l_r)$ and we denote the jth coordinate l_j by $[\overline{\gamma}(v)]_j$. Given a pair $(\overline{S}, \overline{\gamma})$, we break it into different projection sets P_j for each $1 \le j \le r$

$$P_j = \{([v]_j, [\overline{\gamma}(v)]_j) \mid v \in \overline{S}\}$$

Each element in P_j corresponds to a vertex in G_Φ (given by $[v]_j$) and a color in $[l]$ for the vertex (given by $[\overline{\gamma}(v)]_j$). Note that there can be two different elements $((C_i, \alpha), l_j)$ and $((C_i, \alpha), l'_j)$ which assign different colors to the same vertex. If this is the case for any set P_j, we take $\overline{\mathbf{V}}_{(\overline{S}, \overline{\gamma})} = 0$. Otherwise, for each set P_j, we can define the set $\mathcal{S}_j$ of vertices of G_Φ that are contained in P_j and also a partial assignment $\gamma_j \in [l]^{\mathcal{S}_j}$, since every vertex of $\mathcal{S}_j$ gets a unique color by assumption. In this case, we define $\overline{\mathbf{V}}_{(\overline{S}, \overline{\gamma})}$ by tensoring assignment vectors in each coordinate. Formally,

$$\overline{\mathbf{V}}_{(\overline{S}, \overline{\gamma})} = \begin{cases} 0 & \exists 1 \le j \le r \ \& \ ((C_i, \alpha), l_j), ((C_i, \alpha), l'_j) \in P_j \ s.t. \ l_j \ne l'_j \\ \overline{\mathbf{V}}_{(\mathcal{S}_1, \gamma_1)} \otimes \ldots \otimes \overline{\mathbf{V}}_{(\mathcal{S}_r, \gamma_r)} & \text{otherwise} \end{cases}$$

It is easy to verify that the vectors satisfy all the required SDP conditions.

- Let $u_1 = \{(C_{i_1}, \alpha_1), \ldots, (C_{i_r}, \alpha_r)\}$ and $u_2 = \{(C_{i_1}, \alpha_1), \ldots, (C_{i_r}, \alpha_r)\}$ be two adjacent vertices, and let $\overline{\gamma} \in [l]^r$ be any color $(l_1, \ldots, l_r)$. Then, by adjacency, we must have that for some $j \le r$, $\{(C_{i_j}, \alpha_j), (C_{i'_j}, \alpha'_j)\} \in E_\Phi$. Hence,

$$\left\langle \overline{\mathbf{V}}_{(\{u_1\}, \overline{\gamma})}, \overline{\mathbf{V}}_{(\{u_2\}, \overline{\gamma})} \right\rangle = \prod_{j=1}^{r} \left\langle \overline{\mathbf{V}}_{(\{(C_{i_j}, \alpha_j)\}, l_j)}, \overline{\mathbf{V}}_{(\{(C_{i'_j}, \alpha'_j)\}, l_j)} \right\rangle = 0$$

- Similarly, if $(\overline{S}_1, \overline{\gamma}_1)$ and $(\overline{S}_2, \overline{\gamma}_2)$ have a contradiction, or $\overline{S}_1 \cup \overline{S}_2 = \overline{S}_3 \cup \overline{S}_4$ and $\overline{\gamma}_1 \circ \overline{\gamma}_2 = \overline{\gamma}_3 \circ \overline{\gamma}_4$, then these conditions will also hold in each of the coordinate-wise projections. Hence, the SDP conditions will be satisfied in these cases.
- To verify that for each $u \in V$, $\sum_{j \in [l]^r} \left|\overline{\mathbf{V}}_{(\{u\},j)}\right|^2 = 1$ we again note that

$$\sum_{l_1,\ldots,l_r} \left|\overline{\mathbf{V}}_{(\{(C_{i_1},\alpha_1),\ldots,(C_{i_r},\alpha_r)\},(l_1,\ldots,l_r))}\right|^2 = \sum_{l_1,\ldots,l_r} \prod_{j=1}^{r} \left\langle \overline{\mathbf{V}}_{(\{(C_{i_j},\alpha_j)\},l_j)}, \overline{\mathbf{V}}_{(\{(C_{i_j},\alpha_j)\},l_j)} \right\rangle$$
$$= \prod_{j=1}^{r} \sum_{l_j} \left|\overline{\mathbf{V}}_{(\{(C_{i_j},\alpha_j)\},l_j)}\right|^2 = 1$$

■

We now prove the integrality gap for Chromatic Number by similar arguments as in Theorem 6.4.

Theorem 6.12 *There exist constants $c_1, c_2, c_3 > 0$ and graphs G on N vertices, for arbitrarily large N such that*

1. *The chromatic number of G is $\Omega\left(\frac{N}{2^{c_1\sqrt{\log N \log\log N}}}\right)$.*
2. *The SDP for coloring G obtained by $\Omega\left(2^{c_2\sqrt{\log N \log\log N}}\right)$ levels of the Lasserre hierarchy admits a vector coloring with $O\left(2^{c_3\sqrt{\log N \log\log N}}\right)$ colors.*

Proof: We construct the graph G by sampling M cliques of the form $C(i_1, \ldots, i_r)$ from G_Φ^r, and considering the subgraph induces by their vertices. The size of the independent sets is small w.h.p. over the choice of G.

Claim 6.13 *Let $s = \mathsf{OPT}_k(\Phi)/m$. Then for $M \geq \frac{100nr}{s^r}$, with probability $1 - o(1)$, all independent sets in G have size at most $2s^r M$.*

Proof: By Chernoff bound arguments identical to those in Claim 6.5. ■

We again choose $k = \delta \log n$ for some small constant δ, and let $r = \log n/(\log\log n)$ for a large n. Applying Corollary 6.7 with $\varepsilon = 1/2$ gives an instance Φ of MAX k-CSP (P_A) with $k/2^k \leq s \leq 3k/2^k$. (Note that here $s = \mathsf{OPT}_k(\Phi)/m$). The number of assignments to each constraint is exactly $l = 2^{\lceil \log(k+1) \rceil} \leq 2k$. We again pick $M = 100nr \cdot (2^{kr}/k^r)$.

With high probability over the choice of G, the size of the maximum independent set in G is at most $2Ms^r$. The number of vertices in G is

$$N = M \cdot l^r \leq (100nr \cdot 2^{kr}/k^r) \cdot (2k)^r = O(nr \cdot 2^{(k+1)r})$$

and hence the chromatic number of G is at least $(l^r/2s^r)$, which is $\dfrac{N}{2^{c_1\sqrt{\log N \log\log N}}}$ for some constant c_1, with our parameters.

We can again take the Lasserre vectors corresponding to sets of vectors in G, to be the same as the vectors for the corresponding sets in G^r_Φ. By Claim 6.11 the number of colors in the vector coloring is l^r, which is at most $2^{c_2\sqrt{\log N \log\log N}}$ for some constant c_2. Also, the number of levels is $\Omega\left(\frac{n}{k\cdot\beta^{25}}\right)$, which is $\Omega\left(2^{c_3\sqrt{\log N \log\log N}}\right)$ for $c_3 > 0$, if δ (in choosing $k = \delta \log n$) is small enough. ■

6.4 Integrality Gaps for Vertex Cover

In this section, we prove an integrality gap of 1.36 for Minimum Vertex Cover using the reduction by Dinur and Safra [DS05]. One can start with an integer program for Minimum Vertex Cover and obtain the Lasserre SDP for Vertex Cover by introducing vector variables for every set of t vertices in the graph. Equivalently (and this is the form we will be using), one can simply work with the t-round SDP for Maximum Independent Set and modify the objective.

We collect the relation between SDP solutions for Maximum Independent Set and Minimum Vertex Cover, together with some other (known) characterizations of the independent set solution which we shall need, in the lemma below. We provide a proof below for the sake of self-containment.

Lemma 6.14 *Let the vectors $\mathbf{U}_S$ for $|S| \leq t$ form a solution to the t-round Lasserre SDP for* Maximum Independent Set *on a weighted graph $G = (V, E)$, with SDP value* FRAC(G). *Then there exist vectors $\mathbf{V}_{(S,\alpha)}$ for all $|S| \leq t/2, \alpha \in \{0,1\}^S$, determined by the vectors $\mathbf{U}_S$, such that*

1. $\mathbf{U}_\emptyset = \mathbf{V}_{(\emptyset,\emptyset)}$ *and* $\mathbf{U}_S = \mathbf{V}_{(S,1_S)} \forall S$, *where 1_S is the partial assignment which assigns 1 to all elements in S.*
2. *The vectors $\mathbf{V}_{(S,\alpha)}$ satisfy all conditions of the SDP for constraint satisfaction problems.*
3. *For any S, the vectors $\{\mathbf{V}_{(S,\alpha)} \mid \alpha \in \{0,1\}^S\}$ induce a probability distribution over $\{0,1\}^S$. The events measurable in this probability space correspond to all $\alpha' \in \{0,1\}^{S'}$ for all $S' \subseteq S$, and* $\mathbb{P}[\alpha'] = \left|\mathbf{V}_{(S',\alpha')}\right|^2$.
4. *The vectors $\mathbf{V}_{(S,0_S)}$ form a solution to the t-round SDP for* Minimum Vertex Cover *with objective value $\sum_{v\in V} w(v) -$* FRAC(G), *where $w(v)$ denotes the weight of vertex v and 0_S denotes the all-zero assignment.*

Proof: We shall define the vectors for all sets of size upto t. However, for convenience, we shall only prove the SDP conditions for vectors corresponding to sets of size at most $t/2$. For a pair (S, α) where $\alpha \in \{0,1\}^S$, we denote by $\alpha^{-1}(0)$ the set $\{i \in S \mid \alpha(i) = 0\}$ and by $\alpha^{-1}(1)$, the set $S \setminus \alpha^{-1}(0)$. We define the vectors $\mathbf{V}_{(S,\alpha)}$ using inclusion-exclusion, as

$$\mathbf{V}_{(S,\alpha)} = \sum_{T\subseteq\alpha^{-1}(0)} (-1)^{|T|}\mathbf{U}_{T\cup\alpha^{-1}(1)}$$

Also, we define $\mathbf{V}_{(\emptyset,\emptyset)} = \mathbf{U}_\emptyset$.

- Property (1) is then immediate from the definition of the vectors. We also note that for any $i \in \alpha^{-1}(0)$, we can write
$$\mathbf{V}_{(S,\alpha)} = \mathbf{V}_{(S\setminus\{i\},\alpha(S\setminus\{i\}))} - \mathbf{V}_{(S,\alpha')}$$
where $\alpha' \in \{0,1\}^S$ is such that $\alpha'(j) = \alpha(j)\ \forall j \neq i$ and $\alpha'(i) = 1$.

- We now show that the vectors satisfy the SDP conditions. For all S_1, S_2 such that $|S_1 \cup S_2| \leq t$, we will show that
$$\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)}\rangle = \begin{cases} 0 & \text{when } \alpha_1(S_1\cap S_2) \neq \alpha_2(S_1 \cap S_2) \\ \langle \mathbf{V}_{(S_1\cup S_2,\alpha_1\circ\alpha_2)}, \mathbf{V}_{(\emptyset,\emptyset)}\rangle & \text{otherwise}\end{cases}$$
It is easy to check that this will show that all SDP conditions are satisfied for sets of size at most $t/2$. We will proceed by induction on the total number of "zeroes" in the assignments α_1 and α_2 i.e. on $|\alpha_1^{-1}(0)| + |\alpha_2^{-1}(0)|$. The base case is when $\alpha_1 = 1_{S_1}$ and $\alpha_2 = 1_{S_2}$. Then, the product on the left is simply equal to $\langle \mathbf{U}_{S_1}, \mathbf{U}_{S_2}\rangle$, which is equal to $\left|\mathbf{U}_{S_1\cup S_2}\right|^2$ since the vectors $\mathbf{U}_S$ form a valid solution to the independent set SDP.

  For the induction step, first consider the case when α_1 and α_2 disagree on some $i \in S_1 \cap S_2$. Say $\alpha_1(i) = 0$ and $\alpha_2(i) = 1$. Then we can rewrite the inner product as
$$\begin{aligned}\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)}\rangle &= \left\langle \mathbf{V}_{(S_1\setminus\{i\},\alpha_1(S_1\setminus\{i\}))} - \mathbf{V}_{(S_1,\alpha_1')}, \mathbf{V}_{(S_2,\alpha_2)}\right\rangle \\ &= \langle \mathbf{V}_{(S_1\setminus\{i\},\alpha_1(S_1\setminus\{i\}))}, \mathbf{V}_{(S_2,\alpha_2)}\rangle - \left\langle \mathbf{V}_{(S_1,\alpha_1')}, \mathbf{V}_{(S_2,\alpha_2)}\right\rangle\end{aligned}$$
where, as before, α_1' is equal to α_1 for all $j \neq i$ and has $\alpha_1'(i) = 1$. By the induction hypothesis, the terms on the right are either both equal to 0 or both equal $\left\langle \mathbf{V}_{(S_1\cup S_2,\alpha_1'\circ\alpha_2)}, \mathbf{V}_{(\emptyset,\emptyset)}\right\rangle$ depending on whether α_1' and α_2 disagree or not. In either case, their difference is 0.

  Next, we consider $\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)}\rangle$ when $\alpha_1 \circ \alpha_2$ is well defined. For all $i \in S_1 \cap S_2$ such that $\alpha_1(i) = \alpha_2(i) = 0$, we can always write $\mathbf{V}_{(S_2,\alpha_2)} = \mathbf{V}_{(S_2\setminus\{i\},\alpha_2(S_2\setminus\{i\}))} - \mathbf{V}_{(S_2,\alpha_2')}$ and note that $\mathbf{V}_{(S_2,\alpha_2')}$ will be orthogonal to $\mathbf{V}_{(S_1,\alpha_1)}$ by the previous case, since it contradicts α_1 on i. Hence, we can always reduce to the case when there is an i in only one of the sets, say S_1, such that $\alpha_1(i) = 0$. Now we again decompose $\mathbf{V}_{(S_1,\alpha_1)}$, and note that
$$\begin{aligned}\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)}\rangle &= \langle \mathbf{V}_{(S_1\setminus\{i\},\alpha_1(S_1\setminus\{i\}))}, \mathbf{V}_{(S_2,\alpha_2)}\rangle - \left\langle \mathbf{V}_{(S_1,\alpha_1')}, \mathbf{V}_{(S_2,\alpha_2)}\right\rangle \\ &= \langle \mathbf{V}_{(S_1\cup S_2\setminus\{i\},\alpha_1(S_1\setminus\{i\})\circ\alpha_2)}, \mathbf{V}_{(\emptyset,\emptyset)}\rangle - \left\langle \mathbf{V}_{(S_1\cup S_2,\alpha_1'\circ\alpha_2)}, \mathbf{V}_{(\emptyset,\emptyset)}\right\rangle \\ &= \langle \mathbf{V}_{(S_1\cup S_2,\alpha_1\circ\alpha_2)}, \mathbf{V}_{(\emptyset,\emptyset)}\rangle\end{aligned}$$
where we used the induction hypothesis in the second step.

- We claim that to show property (3), it is sufficient to prove that for all S, $\sum_{\alpha\in\{0,1\}^S} \left|\mathbf{V}_{(S,\alpha)}\right|^2 = 1$ (It is clearly necessary if we intend to interpret these values as probabilities). Before proving it, we note that it is equivalent to the condition that $\sum_{\alpha\in\{0,1\}^S} \mathbf{V}_{(S,\alpha)} = \mathbf{V}_{(\emptyset,\emptyset)}$ because it implies
$$\left|\sum_{\alpha\in\{0,1\}^S} \mathbf{V}_{(S,\alpha)} - \mathbf{V}_{(\emptyset,\emptyset)}\right|^2 = \sum_{\alpha\in\{0,1\}^S} \left|\mathbf{V}_{(S,\alpha)}\right|^2 - 1 = 0 \tag{6.1}$$

Using the above equivalence, we can then conclude the consistency of the probability distributions as defined in the lemma. Consider a set S and the distribution over $\alpha \in \{0,1\}^S$ given by $\mathbb{P}[\alpha] = \left|\mathbf{V}_{(S,\alpha)}\right|^2$. In this distribution, if $\alpha' \in \{0,1\}^{S'}$ is an event, then the probability of α' can be calculated by summing the probabilities of all the events $\alpha \in \{0,1\}^S$ such that $\alpha(S') = \alpha'$. This must also equal $\left|\mathbf{V}_{(S',\alpha')}\right|^2$, which would be the probability of α' if we just considered the distribution over S'. This follows from (6.1) because,

$$\begin{aligned}\sum_{\alpha(S')=\alpha} \left|\mathbf{V}_{(S,\alpha)}\right|^2 = \sum_{\alpha(S')=\alpha'} \left\langle \mathbf{V}_{(S,\alpha)}, \mathbf{V}_{(\emptyset,\emptyset)} \right\rangle &= \sum_{\alpha_1 \in \{0,1\}^{S \setminus S'}} \left\langle \mathbf{V}_{(S\setminus S',\alpha_1)}, \mathbf{V}_{(S',\alpha')} \right\rangle \\ &= \left\langle \mathbf{V}_{(\emptyset,\emptyset)}, \mathbf{V}_{(S',\alpha')} \right\rangle = \left|\mathbf{V}_{(S',\alpha')}\right|^2\end{aligned}$$

To prove that $\sum_{\alpha\in\{0,1\}^S} \left|\mathbf{V}_{(S,\alpha)}\right|^2 = 1$ for all S with $|S| \le t/2$, we proceed by induction on $|S|$. The base case (empty set) is trivial. To do the induction step, note that for any $i \in S$

$$\begin{aligned}\sum_{\alpha\in\{0,1\}^S} \left|\mathbf{V}_{(S,\alpha)}\right|^2 &= \sum_{\alpha_1\in\{0,1\}^{S\setminus\{i\}}} \left\langle \mathbf{V}_{(S\setminus\{i\},\alpha_1)}, \mathbf{V}_{(\{i\},0)} \right\rangle + \sum_{\alpha_1\in\{0,1\}^{S\setminus\{i\}}} \left\langle \mathbf{V}_{(S\setminus\{i\},\alpha_1)}, \mathbf{V}_{(\{i\},1)} \right\rangle \\ &= \left\langle \mathbf{V}_{(\emptyset,\emptyset)}, \mathbf{V}_{(\{i\},1)} \right\rangle + \left\langle \mathbf{V}_{(\emptyset,\emptyset)}, \mathbf{V}_{(\{i\},1)} \right\rangle = \left\langle \mathbf{V}_{(\emptyset,\emptyset)}, \mathbf{V}_{(\emptyset,\emptyset)} \right\rangle = 1\end{aligned}$$

- Finally, to show that the vectors $\mathbf{V}_{(S,0_S)}$ form a valid solution to the vertex cover SDP, we note that they satisfy all the consistency conditions by the previous arguments. The only extra condition that the SDP would impose is that for any edge (i,j) , $\left\langle \mathbf{U} - \mathbf{V}_{(\{i\},0)}, \mathbf{U}_\emptyset - \mathbf{V}_{(\{j\},0)} \right\rangle = 0$. But this is immediate because $\mathbf{U}_\emptyset - \mathbf{V}_{(\{i\},0)} = \mathbf{U}_{\{i\}}$ (similarly for j) and $\left\langle \mathbf{U}_{\{i\}}, \mathbf{U}_{\{j\}} \right\rangle = 0$.

■

Through the remaining part of this section, we shall only consider the independent set SDP on all the graphs in the reduction. We will show that the value of the fractional independent set is large for all intermediate graphs obtained in the reduction. On the other hand, it will be possible to conclude directly from the correctness of the Dinur-Safra proof that the size of the actual independent set in these graphs is small. Comparing the corresponding values for vertex cover will give us the required integrality gap.

6.4.1 The starting graphs for the Dinur-Safra reduction

We first describe the graphs required for the reduction by Dinur and Safra [DS05]. They require few key properties of the graph which they use to argue the soundness of the reduction i.e. the graphs produces by the reduction have no large independent set. In fact, graphs of the form G^r_Φ as defined in section 6.2 turn out to satisfy all the required conditions. Also, we already have vector solutions for the independent set SDP on these graphs. We only need to argue that these vectors can be transformed appropriately through the steps of the reduction. Dinur and Safra define the notion of "co-partite" graphs as below.

Definition 6.15 *We say that a graph $G = (M \times L, E)$ is (m,l)* co-partite*, if it is composed of $m = |M|$ disjoint cliques, each of size $l = |L|$. The edges that go between the cliques may be arbitrary. Formally, for all $i \in M$ and $j_1 \neq j_2 \in L$, we require that $\{(i,j_1),(i,j_2)\} \in E$.*

Let Φ be any CSP instance with m constraints and each constraint having exactly l satisfying assignments. Then it is easy to see that the FGLSS graph G_Φ is (m,l) co-partite. Also, the graph G^r_Φ is (m^r, l^r) co-partite. The reduction in [DS05] also requires an (m,l) co-partite graphs such that for some fixed constants $\varepsilon_0, h > 0$ every subset of vertices $I \subseteq M \times L$ with $|I| \geq \varepsilon_0 m$, contains a clique of size h. It also follows from their argument (proof of Theorem 2.1 in [DS05]) [3] that if $\mathsf{OPT}(\Phi) \leq s \cdot m$ for some $s < 1$, then G^r_Φ satisfies this property for an appropriate r.

Theorem 6.16 ([DS05]) *Let Φ be a CSP instance with m constraints, each having l satisfying assignments, and such that any assignment satisfies at most $s < 1$ fraction of the constraints. Also, let $\varepsilon_0, h > 0$ be given. Then there exists an $r = O(\log(h/\varepsilon))$ such that any set of vertices in G^r_Φ, which does not contain an h-clique, has size at most $\varepsilon_0 \cdot m^r$.*

6.4.2 The graphs with block-assignments

The next step of the reduction, which is crucial for the soundness, transforms a graph $G = (V, E)$ which is (m_0, l_0) co-partite into a new graph $G_\mathcal{B}$ which is (m_1, l_1) co-partite and has some additional properties required for the soundness.

We consider the set of *blocks* of d vertices in V, i.e. the set

$$\mathcal{B} = \binom{V}{d} = \{B \subseteq V \mid |B| = d\}$$

Also, for each block B, let L_B denote all "large" partial assignments to vertices in B. Formally, $L_B = \{\alpha \in \{0,1\}^B \mid |\alpha| \geq d_\mathsf{T}\}$, where $d_\mathsf{T} = d/2l_0$ and $|\alpha|$ is the number of 1s in the image of α. The vertex set of the graph $G_\mathcal{B}$ is taken to be set of all pairs of the form (B, α), where $\alpha \in L_B$. To define the edges, we consider a pair of blocks whose symmetric difference is just a pair of vertices (v_1, v_2) such that $(v_1, v_2) \in E$. We connect two partial assignments corresponding to such a pair of blocks, if they form an "obvious contradiction" i.e. they disagree on the intersection or they assign 1 to both the vertices in the symmetric difference. It is important for the soundness analysis in [DS05] that for any such pair of blocks (B_1, B_2), any $\alpha_1 \in L_{B_1}$ is *not connected* to at most two partial assignments in B_2 (and vice-versa). We also add edges between all partial assignments within a block. Thus, we define

$$\begin{aligned}
V_\mathcal{B} &= \{(B,\alpha) \mid B \in \mathcal{B}, \alpha \in L_B\} \\
E^{(1)}_\mathcal{B} &= \bigcup_{\substack{\hat{B} \in \binom{V}{d-1} \\ (v_1,v_2) \in E}} \left\{ (\hat{B} \cup \{v_1\}, \alpha_1), (\hat{B} \cup \{v_2\}, \alpha_2) \mid \alpha_1(\hat{B}) \neq \alpha_2(\hat{B}) \text{ or } \alpha_1(v_1) = \alpha_2(v_2) = 1 \right\} \\
E_\mathcal{B} &= E^{(1)}_\mathcal{B} \cup \left(\bigcup_B \{(B,\alpha_1),(B,\alpha_2) \mid \alpha_1 \neq \alpha_2,\ \alpha_1,\alpha_2 \in L_B\} \right)
\end{aligned}$$

Note that $G_\mathcal{B}$ is also (m_1, l_1) co-partite with $m_1 = |\mathcal{B}|$ and $l_1 = |L_B|$ (which is the same for all B). We now show that if G has a good SDP solution, then so does $G_\mathcal{B}$.

[3] The result in [DS05] is actually stated not for the graph G^r_Φ, but for a graph G' obtained by converting the CSP Φ to a two-player game and considering the graph obtained by parallel repetition. However, the graph G' defined in their paper is a spanning subgraph of G^r_Φ, and hence if a subset of vertices contains an h-clique in G', then it also contains one in G^r_Φ.

Lemma 6.17 *Let $G = (V, E)$ be an (m_0, l_0) co-partite graph such that for the independent set SDP obtained by t levels of the Lasserre hierarchy,* $\mathsf{FRAC}(G) = m_0$. *For given $\varepsilon > 0$ and $d_{\mathsf{T}} > 2/\varepsilon_1$, let $G_{\mathcal{B}}$ be constructed as above. Then, for the independent set SDP obtained by t/d levels,* $\mathsf{FRAC}(G_{\mathcal{B}}) \geq (1 - \varepsilon_1)|\mathcal{B}|$.

Proof: Since G admits a solution for t levels of the Lasserre SDP, we also have vectors $\mathbf{V}_{(S,\alpha)}$ for all $|S| \leq t, \alpha \in \{0,1\}^S$ as described in Lemma 6.14. We now use these vectors to define the vector solution for the SDP on $G_{\mathcal{B}}$. Each vertex of $G_{\mathcal{B}}$ is of the form (B, α) where $B \in \mathcal{B}$ and $\alpha \in L_B$. Consider a set $\mathcal{S}$ of $i \leq t/d$ vertices, $\mathcal{S} = \{(B_1, \alpha_1), \ldots, (B_i, \alpha_i)\}$. As in the Section 6.2, we let the vector corresponding to this set be 0 if any two assignments in the set contradict and equal to the vector for the partial assignment jointly defined by $\alpha_1, \ldots, \alpha_i$ otherwise.

$$\mathbf{U}_{\mathcal{S}} = \begin{cases} 0 & \exists j_1, j_2 \ s.t. \ \alpha_{j_1}(B_{j_1} \cap B_{j_2}) \neq \alpha_{j_2}(B_{j_1} \cap B_{j_2}) \\ \mathbf{V}_{(\cup_j B_j, \alpha_1 \circ \ldots \circ \alpha_i)} & \text{otherwise} \end{cases}$$

If (B_1, α_1) and (B_2, α_2) have an edge between them because α_1 and α_2 contradict, then we must have $\langle \mathbf{U}_{\{(B_1,\alpha_1)\}}, \mathbf{U}_{\{(B_2,\alpha_2)\}} \rangle = \langle \mathbf{V}_{(B_1,\alpha_1)}, \mathbf{V}_{(B_2,\alpha_2)} \rangle = 0$ since CSP vectors corresponding to contradicting partial assignments must be orthogonal. For an edge between $(\hat{B} \cup \{v_1\}, \alpha_1)$ and $(\hat{B} \cup \{v_2\}, \alpha_2)$ where $(v_1, v_2) \in E$, $\alpha_1(v_1) = \alpha_2(v_2) = 1$ and $\alpha_1(\hat{B}) = \alpha_2(\hat{B}) = \alpha$ (say), we have

$$\left\langle \mathbf{U}_{\{(\hat{B}\cup\{v_1\},\alpha_1)\}}, \mathbf{U}_{\{(\hat{B}\cup\{v_2\},\alpha_2)\}} \right\rangle = \left\langle \mathbf{V}_{(\hat{B},\alpha)}, \mathbf{V}_{(\{v_1,v_2\},(1,1))} \right\rangle = \left\langle \mathbf{V}_{(\hat{B},\alpha)}, 0 \right\rangle = 0$$

because in the independent solution on graph G

$$\left|\mathbf{V}_{(\{v_1,v_2\},(1,1))}\right|^2 = \langle \mathbf{V}_{(\{v_1\},1)}, \mathbf{V}_{(\{v_2\},1)} \rangle = 0$$

The proof that the vectors defined above satisfy the other SDP conditions is identical to that in Claim 6.3 and we omit the details.

The interesting part of the argument will be to show that the value of the independent set SDP for $G_{\mathcal{B}}$ will be large. In the completeness part of the Dinur-Safra reduction, one needs to say that if G has a *large independent set* then so does $G_{\mathcal{B}}$, and it follows very easily using a Chernoff bound on how different blocks intersect the large independent set of G. We need to make the same conclusion about the SDP value and the argument is no longer applicable. However, it is possible to get the conclusion by looking at the "local distributions" defined by the vector $\mathbf{V}_{(B,\alpha)}$ as mentioned in Lemma 6.14. We then combine the bounds obtained in each block globally using properties of the vector solution.

For each block B, let D_B denote the distribution over $\{0,1\}^B$ defined by the vectors $\mathbf{V}_{(B,\alpha)}$ for all $\alpha \in \{0,1\}^B$. For each block, we define a random variable Z_B determined by the event $\alpha \in \{0,1\}^B$ with value $Z_B = |\alpha|$. One can then convert the statement about the SDP value to a statement about the local distributions, by noting that

$$\mathsf{FRAC}(G_{\mathcal{B}}) = \sum_{B\in\mathcal{B}} \sum_{\alpha\in L_B} \left|\mathbf{V}_{(B,\alpha)}\right|^2 = \sum_{B\in\mathcal{B}} \left(1 - \sum_{\alpha\notin L_B} \left|\mathbf{V}_{(B,\alpha)}\right|^2\right) = |\mathcal{B}| - \sum_{B\in\mathcal{B}} \mathop{\mathbb{P}}_{D_B} [Z_B < d_{\mathsf{T}}]$$

The problem thus reduces to showing that $\sum_{B\in\mathcal{B}} \mathbb{P}_{D_B}[Z_B < d_{\mathsf{T}}] \leq \varepsilon_1|\mathcal{B}|$. By a second moment analysis on every local distribution, we have

$$\mathop{\mathbb{P}}_{D_B}[Z_B < d_{\mathsf{T}}] \;\leq\; \mathop{\mathbb{P}}_{D_B}[|Z_B - 2d_{\mathsf{T}}| > d_{\mathsf{T}}] \;\leq\; \frac{\mathbb{E}_{D_B}\left[(Z_B - 2d_{\mathsf{T}})^2\right]}{d_{\mathsf{T}}^2} \tag{6.2}$$

The following claim provides the necessary estimates to bound the sum of probabilities.

Claim 6.18

$$\sum_{B\in\mathcal{B}} \mathop{\mathbb{E}}_{D_B}[Z_B] = 2d_T \cdot |\mathcal{B}| \quad \textit{and} \quad \sum_{B\in\mathcal{B}} \mathop{\mathbb{E}}_{D_B}[Z_B^2] \leq (4d_T^2 + 2d_T)\cdot|\mathcal{B}|$$

Proof: For all blocks B and $v \in B$, define the random variable $X_{v,B}$ which is 1 if a random α chosen according to D_B assigns 1 to the vertex v and 0 otherwise. By using the fact that the distribution is defined by the vectors $\mathbf{V}_{(B,\alpha)}$, we get that for $v \in B$

$$\mathop{\mathbb{E}}_{D_B}[X_{v,B}] \;=\; \mathop{\mathbb{P}}_{D_B}[\alpha(v) = 1] \;=\; \left|\mathbf{V}_{(\{v\},1)}\right|^2 \;=\; \left|\mathbf{U}_{\{v\}}\right|^2$$

where $\mathbf{U}_{\{v\}}$ denotes the vector for vertex v in the solution for graph G. Similarly,

$$\mathop{\mathbb{E}}_{D_B}[X_{v_1,B}X_{v_2,B}] \;=\; \mathop{\mathbb{P}}_{D_B}[\alpha(\{v_1,v_2\}) = (1,1)] \;=\; \left|\mathbf{V}_{(\{v_1,v_2\},(1,1))}\right|^2 \;=\; \langle \mathbf{U}_{\{v_1\}}, \mathbf{U}_{\{v_2\}}\rangle$$

With the above relations, and using the facts that each vertex appears in exactly d/m_0l_0 fraction of the blocks and $\mathsf{FRAC}(G) = m_0$, we can compute the sum of expectations of Z_B as

$$\begin{aligned}\sum_{B\in\mathcal{B}} \mathop{\mathbb{E}}_{D_B}[Z_B] \;&=\; \sum_{B\in\mathcal{B}} \mathop{\mathbb{E}}_{D_B}\left[\sum_{v\in B} X_{v,B}\right] \;=\; \sum_{B\in\mathcal{B}}\sum_{v\in B}\left|\mathbf{U}_{\{v\}}\right|^2 \;=\; \frac{d}{m_0l_0}|\mathcal{B}|\left(\sum_{v\in V}\left|\mathbf{U}_{\{v\}}\right|^2\right)\\ &=\; \frac{d}{m_0l_0}|\mathcal{B}|\cdot m_0 \;=\; 2d_{\mathsf{T}}\cdot|\mathcal{B}|\end{aligned}$$

Similarly, for the expectations of the squares, we get

$$\sum_{B\in\mathcal{B}} \mathop{\mathbb{E}}_{D_B}[Z_B^2] \;=\; \sum_{B\in\mathcal{B}}\sum_{v_1,v_2\in B} \mathop{\mathbb{E}}_{D_B}[X_{v_1,B}X_{v_2,B}] \;=\; \sum_{B\in\mathcal{B}}\sum_{v_1,v_2\in B}\langle \mathbf{U}_{\{v_1\}}, \mathbf{U}_{\{v_2\}}\rangle$$

Again, each pair (v_1, v_2) such that $v_1 \neq v_2$ appears in less than $(d^2/m_0^2l_0^2)$ fraction of the blocks, and a pair such that $v_1 = v_2$ appears in d/m_0l_0 fraction. Hence,

$$\begin{aligned}\sum_{B\in\mathcal{B}} \mathop{\mathbb{E}}_{D_B}[Z_B^2] \;&\leq\; |\mathcal{B}|\left(\frac{d^2}{m_0^2l_0^2}\right)\sum_{v_1,v_2}\langle \mathbf{U}_{\{v_1\}}, \mathbf{U}_{\{v_2\}}\rangle + |\mathcal{B}|\left(\frac{d}{m_0l_0}\right)\sum_{v}\left|\mathbf{U}_{\{v\}}\right|^2\\ &=\; |\mathcal{B}|\left(\frac{d^2}{m_0^2l_0^2}\right)\left|\sum_{v}\mathbf{U}_{\{v\}}\right|^2 + |\mathcal{B}|\cdot 2d_{\mathsf{T}}\end{aligned}$$

Finally, to calculate the first term, we shall need the fact that G is (m_0, l_0) co-partite. Let $V = M_0 \times L_0$. We write each v as (i, j) for $i \in M_0, j \in L_0$. Using the fact that all vectors within a single clique are orthogonal, we get

$$\left| \sum_{(i,j) \in M_0 \times L_0} \mathbf{U}_{\{(i,j)\}} \right|^2 \leq m_0 \cdot \sum_{i \in M_0} \left| \sum_{j \in L_0} \mathbf{U}_{\{(i,j)\}} \right|^2 = m_0 \cdot \sum_{i \in M_0} \sum_{j \in L_0} \left| \mathbf{U}_{\{(i,j)\}} \right|^2 = m_0^2.$$

Using this bound we get that $\sum_{B \in \mathcal{B}} \mathbb{E}_{D_B}[Z_B^2] \leq (4d_\mathsf{T}^2 + 2d_\mathsf{T})|\mathcal{B}|$, which proves the claim. ■

Using equation (6.2) and the previous claim, we get that

$$\sum_{B \in \mathcal{B}} \mathop{\mathbb{P}}_{D_B}[Z_B < d_\mathsf{T}] \;\leq\; \frac{1}{d_\mathsf{T}^2} \sum_{B \in \mathcal{B}} (\mathop{\mathbb{E}}_{D_B}[Z_B^2] - 2d_\mathsf{T} \mathop{\mathbb{E}}_{D_B}[Z_B] + 4d_\mathsf{T}^2) \;\leq\; \frac{2}{d_\mathsf{T}}|\mathcal{B}|$$

Hence, for $d > 2/\varepsilon_1$, the SDP value is at least $(1 - \varepsilon_1)|\mathcal{B}|$. ■

6.4.3 The long-code step

The next step of the reduction defines a weighted graph starting from an (m_1, l_1) co-partite graph. Let $G = (V, E)$ be the given graph and $V = M_1 \times L_1$. We then define the graph $G_{\mathcal{LC}}$ which has a vertex for every $i \in M_1$ and every $J \subseteq L_1$. Also, the graph is weighted with each vertex (i, J) having a weight $w(i, J)$ depending on $|J|$.

$$\begin{aligned} V_{\mathcal{LC}} &= \{(i, J) \mid i \in M_1, J \subseteq L_1\} \\ E_{\mathcal{LC}} &= \{\{(i_1, J_1), (i_2, J_2)\} \mid \forall j_1 \in J_1, j_2 \in J_2.\ \{(i_1, j_1), (i_2, j_2)\} \in E\} \\ w(i, J) &= \frac{1}{m_1} p^{|J|}(1 - p)^{|L_1 \setminus J|} \end{aligned}$$

Lemma 6.19 *Let $G = (V, E)$ be an (m_1, l_1) co-partite graph such that independent set SDP for t-levels on G has value at least $\mathsf{FRAC}(G)$. Let $G_{\mathcal{LC}}$ be defined as above. Then there is a solution to the $t/2$-round SDP for independent set on $G_{\mathcal{LC}}$ with value at least $p \cdot \frac{\mathsf{FRAC}(G)}{m_1}$.*

Proof: Let us denote the vectors in G by $\mathbf{U}_S$ and those in $G_{\mathcal{LC}}$ by $\overline{\mathbf{U}}_{\mathcal{S}}$. We define the vectors $\mathbf{U}_{\mathcal{S}}$ for each $\mathcal{S} = \{(i_1, J_1), \ldots, (i_r, J_r)\}$ for $r \leq t$ as

$$\overline{\mathbf{U}}_{\mathcal{S}} = \sum_{j_1 \in J_1, \ldots, j_r \in J_r} \mathbf{U}_{\{(i_1, j_1), \ldots, (i_r, j_r)\}}$$

We now need to verify that they satisfy all the SDP conditions and the SDP value is as claimed.

- We first check that vectors for adjacent vertices are orthogonal. Let $\{(i_1, J_1), (i_2, J_2)\}$ be an edge in $G_{\mathcal{LC}}$. Then

$$\left\langle \overline{\mathbf{U}}_{\{(i_1, J_1)\}}, \overline{\mathbf{U}}_{\{(i_2, J_2)\}} \right\rangle \;=\; \sum_{j_1 \in J_1, j_2 \in J_2} \left\langle \mathbf{U}_{\{(i_1, j_1)\}}, \mathbf{U}_{\{(i_2, j_2)\}} \right\rangle \;=\; 0$$

  since all pairs $\{(i_1, j_1), (i_2, j_2)\}$ form edges in G.

- For convenience, we shall only verify $\left\langle \overline{\mathbf{U}}_{\mathcal{S}_1}, \overline{\mathbf{U}}_{\mathcal{S}_2} \right\rangle = \left\langle \overline{\mathbf{U}}_{\mathcal{S}_1 \cup \mathcal{S}_2}, \overline{\mathbf{U}}_{\emptyset} \right\rangle$ for all $\mathcal{S}_1, \mathcal{S}_2$ with at most $t/2$ elements. This will prove that vectors $\overline{\mathbf{U}}_{\mathcal{S}}$ with $|\mathcal{S}| \leq t/2$ satisfy the SDP conditions. First, we observe that using the orthogonality of vectors in G corresponding to (i, j_1) and (i, j_2) for $j_1 \neq j_2$,
$$\overline{\mathbf{U}}_{\{(i,J_1),(i,J_2)\}} = \sum_{j_1 \in J_1, j_2 \in J_2} \mathbf{U}_{\{(i,j_1),(i,j_2)\}} = \sum_{j \in J_1 \cap J_2} \mathbf{U}_{\{(i,j)\}} = \overline{\mathbf{U}}_{\{(i,J_1 \cap J_2)\}} \tag{6.3}$$
By inducting this argument, it is always possible to assume without loss of generality that for a set $\mathcal{S} = \{(i_1, J_1), \ldots, (i_r, J_r)\}$, the elements $i_1, \ldots, i_r$ are all distinct. We need the above fact and a little more notation to verify the required condition. For $\mathcal{S} = \{(i_1, J_1), \ldots, (i_r, J_r)\}$, let
$$\mathcal{F}(\mathcal{S}) = \{(i_1, j_1), \ldots, (i_r, j_r) \mid j_1 \in J_1, \ldots, j_r \in J_r\}$$
with this notation,
$$\left\langle \overline{\mathbf{U}}_{\mathcal{S}_1}, \overline{\mathbf{U}}_{\mathcal{S}_2} \right\rangle = \sum_{\substack{T_1 \in \mathcal{F}(\mathcal{S}_1) \\ T_2 \in \mathcal{F}(\mathcal{S}_2)}} \langle \mathbf{U}_{T_1}, \mathbf{U}_{T_2} \rangle$$
Also, $\langle \mathbf{U}_{T_1}, \mathbf{U}_{T_2} \rangle \neq 0$ if and only if $\forall (i, j_1) \in T_1, (i, j_2) \in T_2$, we have $j_1 = j_2$ (again using orthogonality of vectors for vertices in a single clique in G). However, this means that $T_1 \cup T_2 \in \mathcal{F}(\mathcal{S}_1 \cup \mathcal{S}_2)$. Also, since all is in $\mathcal{S}_1, \mathcal{S}_2$ are distinct as observed using (6.3), every element $T \in \mathcal{F}(\mathcal{S}_1 \cup \mathcal{S}_2)$ corresponds to a unique pair $T_1 \in \mathcal{F}(\mathcal{S}_1), T_2 \in \mathcal{F}(\mathcal{S}_2)$. This gives
$$\left\langle \overline{\mathbf{U}}_{\mathcal{S}_1}, \overline{\mathbf{U}}_{\mathcal{S}_2} \right\rangle = \sum_{\substack{T_1 \in \mathcal{F}(\mathcal{S}_1) \\ T_2 \in \mathcal{F}(\mathcal{S}_2)}} \langle \mathbf{U}_{T_1}, \mathbf{U}_{T_2} \rangle = \sum_{T \in \mathcal{F}(\mathcal{S}_1 \cup \mathcal{S}_2)} \langle \mathbf{U}_T, \mathbf{U}_{\emptyset} \rangle = \left\langle \overline{\mathbf{U}}_{\mathcal{S}_1 \cup \mathcal{S}_2}, \overline{\mathbf{U}}_{\emptyset} \right\rangle$$

- From the above condition, it follows that for all $\mathcal{S}$, $\left|\overline{\mathbf{U}}_{\mathcal{S}}\right|^2 = \left\langle \overline{\mathbf{U}}_{\mathcal{S}}, \overline{\mathbf{U}}_{\emptyset} \right\rangle \leq \left|\overline{\mathbf{U}}_{\mathcal{S}}\right|$, using $\left|\overline{\mathbf{U}}_{\emptyset}\right| = 1$. Since the length of all vectors is at most 1, and all inner products are positive in G, all inner products for the vector solution above are between 0 and 1.

To verify the SDP value, we simply need to use the fact that all vectors within a single clique in G are orthogonal. Hence the SDP value is equal to
$$\begin{aligned} \frac{1}{m_1} \sum_{i \in M_1, J \subseteq L_1} p^{|J|}(1-p)^{|L_1 \setminus J|} \left|\overline{\mathbf{U}}_{\{(i,J)\}}\right|^2 &= \frac{1}{m_1} \sum_{i \in M_1, J \subseteq L_1} p^{|J|}(1-p)^{|L_1 \setminus J|} \sum_{j \in J} \left|\mathbf{U}_{\{(i,j)\}}\right|^2 \\ &= \frac{1}{m_1} \sum_{i \in M_1, j \in L_1} p \cdot \left|\mathbf{U}_{\{(i,j)\}}\right|^2 = \frac{p}{m_1} \cdot \mathsf{FRAC}(G) \end{aligned}$$

■

6.4.4 Putting things together

For an (m_0, l_0) co-partite graph G, let $DS(G, \varepsilon_1, d_\mathsf{T})$ denote the graph obtained by starting from G and performing the block assignments step and long-code step, where the reduction has parameters d_T and ε_1 for the block assignments step. Let $p_{max} = \frac{3-\sqrt{5}}{2}$. The soundness analysis of the Dinur-Safra reduction can be summarized in the following theorem (stated in a way adapted to our application).

Theorem 6.20 ([DS05]) *For given $\varepsilon_1 > 0, p \in (1/3, p_{max})$, there exist constants $\varepsilon_0, h, d'_{\mathsf{T}}$ such that if G is an m_0, l_0 co-partite graph such that every set of $\varepsilon_0 \cdot m_0$ vertices in G contains an h-clique, then the weight of the maximum independent set in $DS(G, \varepsilon_1, d_{\mathsf{T}})$ for any $d_{\mathsf{T}} \geq d'_{\mathsf{T}}$ is at most $4p^3 - 3p^4 + \varepsilon_1$.*

Using the above theorem and the previous discussion, we can now prove an integrality gap for Minimum Vertex Cover.

Theorem 6.21 *For any given $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ and an infinite family of graphs such that for graphs in the family, with N vertices, the integrality gap for the SDP relaxation for* Minimum Vertex Cover *obtained by $\Omega(N^{\delta})$ levels of the Lasserre hierarchy, remains at least $1.3606 - \varepsilon$.*

Proof: Let $p \in (1/3, p_{max})$ be such that $\frac{1-4p^3+3p^4}{1-p} = 1.3606$ and $\varepsilon_1 = \varepsilon/10$. Let ε_0, h be as given by Theorem 6.20 and let $d_{\mathsf{T}} = \max(d'_{\mathsf{T}}, 2/\varepsilon_1)$. For large enough n, let Φ be an instance of a constraint satisfaction problem as given by Corollary 4.19 (by using $\varepsilon = 1/2$ to invoke the corollary) for $k = 3$. This is simply an instance of MAX 3-XOR on n variables with $m = O(n)$ constraints in which $s < 2/3$ fraction of the constraints are satisfiable and $\mathsf{FRAC}(\Phi) = m$ even after $\Omega(n)$ levels.

By Theorem 6.16, there exists an $r = O(\log(h/\varepsilon_0))$, such that G^r_{Φ}, which is an (m_0, l_0) co-partite graph for $m_0 = m^r$ and $l_0 = 4^r$, has no h-clique-free subset with $\varepsilon_0 m_0$ vertices. Then the weight of the maximum independent set in $DS(G^r_{\Phi}, \varepsilon_1, d_{\mathsf{T}})$ is at most $4p^3 - 3p^4 + \varepsilon_1$, and hence the weight of the minimum vertex cover is at least $1 - 4p^3 + 3p^4 - \varepsilon_1$.

On the other hand, by lemmata 6.2, 6.17 and 6.19, $\mathsf{FRAC}(DS(G^r_{\Phi}, \varepsilon_1, d_{\mathsf{T}})) \geq p(1 - \varepsilon_1)$ for the independent set SDP obtained by $\Omega(n/l_0 d_{\mathsf{T}})$ levels. Hence the gap for Minimum Vertex Cover is at least $\frac{1-4p^3+3p^4-\varepsilon_1}{1-p(1-\varepsilon_1)} \geq 1.3606 - \varepsilon$. It remains to express the number of levels in terms of the number of vertices in $DS(G^r_{\Phi}, \varepsilon_1, d_{\mathsf{T}})$. However, note that at all parameters in the reduction are constants and the size of the graph grows by a polynomial factor at each step of the reduction. Hence, the number of levels equals $\Omega(n) = \Omega(N^{\delta})$ for constant δ depending on p and ε, where N denotes $|DS(G^r_{\Phi}, \varepsilon_1, d_{\mathsf{T}})|$. ■

Chapter 7

Algorithms for Unique Games on Expanding Graphs

In this chapter we give an algorithm for Unique Games, which is a constraint satisfaction problem defined on a graph (with one constraint for each edge), using a semidefinite program. We give a performance guarantee for the algorithm which improves with the expansion of the underlying graph. The analysis works by showing that expansion can be used to translate *local correlations* i.e. correlations between the values assigned to variables with an edge between them [1], to *global correlations* between arbitrary vertices. We then show that having a good bound on the correlation between the values assigned to arbitrary pairs of variables can be used to give a good performance guarantee for the algorithm.

For a special case of Unique Games, in which the constraints are of the form of a linear equation between the pair of variables sharing an edge, we also study the extension of this algorithm to the SDPs given by the Lasserre hierarchy.

Unique Games is a constraint satisfaction problem where one is given a constraint graph $G = (V, E)$, a label set $[q]$ and for each edge $e = (i, j)$, a bijective mapping $\pi_{ij} : [q] \mapsto [q]$. The goal is to assign to each vertex in G a label from $[q]$ so as to maximize the fraction of the constraints that are "satisfied," where an edge $e = (i, j)$ is said to be *satisfied* by an assignment if i is assigned a label l and j is assigned a label l' such that $\pi_{ij}(l) = l'$. The value of a labeling $\Lambda\colon V \to [q]$ is the fraction of the constraints satisfied by it and is denoted by $\mathsf{val}(\Lambda)$. For a Unique Games instance $\mathcal{U}$, we denote by $\mathsf{opt}(\mathcal{U})$ the maximum value of $\mathsf{val}(\Lambda)$ over all labelings. This optimization problem was first considered by Cai, Condon, and Lipton [CCL90].

The Unique Games Conjecture (UGC) of Khot [Kho02] asserts that for such a constraint satisfaction problem, for arbitrarily small constants $\eta, \zeta > 0$, it is NP-hard to decide whether there is a labeling that satisfies $1 - \eta$ fraction of the constraints (called the YES case) or, for every labeling, the fraction of the constraints satisfied is at most ζ (called the NO case) as long as the size of the

[1] Note that this notion of "local" is weaker than the one we use in the rest of this thesis. However, this suffices for the analysis in this chapter.

label set, q is allowed to grow as a function of η and ζ.

Since its origin, the UGC has been successfully used to prove (often optimal) hardness of approximation results for several important NP-hard problems such as Min-2SAT-Deletion [Kho02], Minimum Vertex Cover [KR03], Maximum Cut [KKMO04], Chromatic Number [DMR06], and non-uniform Sparsest Cut [CKK+04, KV05]. However, one fundamental problem that has resisted attempts to prove inapproximability results, even assuming UGC, is the (uniform) Sparsest Cut problem. This problem has a $O(\sqrt{\log n})$ approximation algorithm by Arora, Rao, and Vazirani [ARV04], but no hardness result beyond NP-hardness is known (recently, in [AMS07], a PTAS is ruled out under a complexity assumption stronger than $\mathsf{P} \neq \mathsf{NP}$).

In fact, it seems unlikely that there is a reduction from Unique Games to Sparsest Cut, unless one assumes that the starting Unique Games instance has some expansion property. This is because if the Unique Games instance itself has a sparse cut, then the instance of Sparsest Cut produced by such a reduction also has a sparse cut (this is certainly the case for known reductions, i.e. [CKK+04, KV05]), irrespective of whether the Unique Games instance is a YES or a NO instance. This motivates the following question: is Unique Games problem hard even with the promise that the constraint graph is an expander? A priori, this could be true even with a very strong notion of expansion, leading to a superconstant hardness result for Sparsest Cut and related problems like Minimum Linear Arrangement.

Here we show that the Unique Games problem is actually easy when the constraint graph is even a relatively weak expander. One notion of expansion that we consider in this work is when the second smallest eigenvalue of the normalized Laplacian of a graph G, denoted by $\lambda := \lambda_2(G)$, is bounded away from 0. We note that the size of balanced cuts (relative to the total number of edges) in a graph is also a useful notion of expansion and the results in this work can be extended to work in that setting.

Our main result. We show the following theorem in Section 7.1:

Theorem 7.1 *There is a polynomial time algorithm for* Unique Games *that, given $\eta > 0$, distinguishes between the following two cases:*

- YES *case: There is a labeling which satisfies at least $1-\eta$ fraction of the constraints.*
- NO *case: Every labeling satisfies less than $1 - O(\frac{\eta}{\lambda}\log(\frac{\lambda}{\eta}))$ fraction of the constraints.*

A consequence of the result is that when the Unique Games instance is $(1-\eta)$-satisfiable and $\lambda \gg \eta$, the algorithm finds a labeling to the Unique Games instance that satisfies 99% of the constraints. An important feature of the algorithm is that its performance does not depend on the number of labels q.

Comparison to previous work. Most of the algorithms for Unique Games (which can be viewed as attempts to disprove the UGC) are based on the SDP relaxation proposed by Feige and Lovász [FL92]. Their paper showed that if the Unique Games instance is unsatisfiable, then the value of the SDP relaxation is bounded away from 1, though they did not give quantitative bounds. Khot

[Kho02] gave a SDP-rounding algorithm to find a labeling that satisfies $1 - O(q^2\eta^{1/5}\log(1/\eta))$ fraction of the constraints when there exists a labeling that satisfies $1-\eta$ fraction of the constraints. The SDP's analysis was then revisited by many papers.

On an $(1-\eta)$-satisfiable instance, these papers obtain a labeling that satisfies at least $1 - f(\eta, n, q)$ fraction of the constraints where $f(\eta, n, q)$ is $\sqrt[3]{\eta \log n}$ in Trevisan [Tre05], $\sqrt{\eta \log q}$ in Charikar, Makarychev, and Makarychev [CMM06a], $\eta\sqrt{\log n \log q}$ in Chlamtac, Makarychev, and Makarychev [CMM06b], and $\eta \log n$ via an LP based approach in Gupta and Talwar [GT06]. Trevisan [Tre05] also gave a combinatorial algorithm that works well on expanders. On an $(1-\eta)$-satisfiable instance, he showed how to obtain a labeling satisfying $1 - \eta \log n \log \frac{1}{\lambda}$ fraction of the constraints. All these results require η to go to 0 as either n or q go to infinity in order to maintain their applicability[2].

Our result differs from the above in that under an additional promise of a natural graph property, namely expansion, the performance of the algorithm is independent of q and n. Furthermore, our analysis steps away from the edge-by-edge analysis of previous papers in favor of a more global analysis of correlations, which may be useful for other problems.

Stronger relaxations of expansion. We note that if we impose a certain structure on our constraints, namely if they are of the form ΓMAX2LIN, our results continue to hold when λ is replaced by stronger relaxations for the expansion of G, given by the Lasserre hierarchy [Las01]. In particular, we show that λ can be replaced by the value of such a relaxation for expansion of G after a constant number of levels.

Application to parallel repetition. Since our main result shows an upper bound on the integrality gap for the standard SDP, the analysis of Feige and Lovász [FL92] allows us to prove (see Section 7.2) a *parallel repetition theorem* for unique games with expansion. We show that the r-round parallel repetition value of a Unique Games instance with value at most $1-\varepsilon$ is at most $(1 - \Omega(\varepsilon \cdot \lambda / \log \frac{1}{\varepsilon}))^r$. In addition to providing an alternate proof, when $\lambda \gg \varepsilon^2 \log(1/\varepsilon)$, this is better than the general bound for nonunique games, where the best bound is $(1 - \Omega(\varepsilon^3 / \log q))^r$ by Holenstein [Hol07], improving upon Raz's Theorem [Raz98]. We note that Safra and Schwartz [SS07] also gave an alternate proof of the parallel repetition theorem for games with expansion, and their result works even for general games. Also, Rao [Rao08] proved a better parallel repetition theorem for, so called, projection games, which are more general than unique games. His result does not assume any expansion of the game graph.

Randomly generated games. For many constraint satisfaction problems such as 3SAT, solving randomly generated instances is of great interest. For instance, proving unsatisfiability of formulae on n variables and with dn randomly chosen clauses seems very difficult for $d \ll \sqrt{n}$. Our results suggest that it will be hard to define a model of probabilistic generation for unique games that will result in very difficult instances, since the natural models all lead to instances with high expansion.

[2]On the other hand, the UGC allows q to grow arbitrarily as a function of η, and therefore, all known algorithms fall short of disproving UGC.

7.1 Main result

Let $\mathcal{U} = (G(V,E), [q], \{\pi_{ij}\}_{(i,j)\in E})$ be a Unique Games instance. We use the standard SDP relaxation for the problem, which involves finding a *vector assignment* for each vertex.

For every $i \in V$, we associate a set of q orthogonal vectors $\{\mathbf{V}_{(i,1)}, \ldots, \mathbf{V}_{(i,q)}\}$. The intention is that if $l_0 \in [q]$ is a label for vertex $i \in V$, then $\mathbf{V}_{(i,l_0)} = \mathbf{1}$, and $\mathbf{V}_{(i,l)} = \mathbf{0}$ for all $l \neq l_0$. Here, $\mathbf{1}$ is some fixed unit vector and $\mathbf{0}$ is the zero-vector. Of course, in a general solution to the SDP this may no longer be true and $\{\mathbf{V}_{(i,1)}, \ldots, \mathbf{V}_{(i,q)}\}$ is just any set of orthogonal vectors.

$$
\begin{array}{lll}
\text{maximize} & \mathop{\mathbb{E}}_{e=(i,j)\in E}\left[\sum_{l\in[q]} \left\langle \mathbf{V}_{(i,l)}, \mathbf{V}_{(j,\pi_{ij}(l))}\right\rangle\right] & \\
\text{subject to} & \sum_{l\in[q]} \left|\mathbf{V}_{(i,l)}\right|^2 = 1 & \forall\, i \in V \\
 & \langle \mathbf{V}_{(i,l)}, \mathbf{V}_{(i,l')}\rangle = 0 & \forall\, i \in V,\ \forall\, l \neq l' \\
 & \left\langle \mathbf{V}_{(i,l)}, \mathbf{V}_{(j,l')}\right\rangle \geq 0 & \forall\, i,j \in V,\ \forall\, l,l'
\end{array}
$$

Figure 7.1: SDP for Unique Games

Our proof will use the fact that the objective function above can be rewritten as

$$1 - \frac{1}{2} \mathop{\mathbb{E}}_{e=(i,i)\in E}\left[\sum_{l\in[q]} \left|\mathbf{V}_{(i,l)} - \mathbf{V}_{(j,\pi_{ij}(l))}\right|^2\right] \tag{7.1}$$

7.1.1 Overview

Let $\mathcal{U} = (G(V,E), [q], \{\pi_{ij}\}_{(i,j)\in E})$ be a Unique Games instance, and let $\{\mathbf{V}_{(i,l)}\}_{i\in V, l\in[q]}$ be an optimal SDP solution. Assume wlog that its value is $1-\eta$, since otherwise we know already that the instance is a NO instance. How do we extract a labeling from the vector solution?

The first constraint suggests an obvious way to view the vectors corresponding to vertex i as a distribution on labels, namely, one that assigns label l to i with probability $\left|\mathbf{V}_{(i,l)}\right|^2$. The most naive idea for a rounding algorithm would be to use this distribution to pick a label for each vertex, where the choice for different vertices is made independently. Of course, this doesn't work since all labels could have equal probability under this distribution and thus the chance that the labels l, l' picked for vertices i, j in an edge e satisfy $\pi_e(l) = l'$ is only $1/q$.

More sophisticated roundings use the fact that if the SDP value is $1-\eta$ for some small η, then the vector assignments to the vertices of an average edge $e = (i, j)$ are highly correlated, in the sense that for "many" l, $\left\langle \overline{\mathbf{V}}_{(i,l)}, \overline{\mathbf{V}}_{(j,\pi(l))}\right\rangle > 1 - \Omega(\eta)$ where $\overline{\mathbf{V}}_{(i,l)}$ denotes the unit vector in the direction of $\mathbf{V}_{(i,l)}$. This suggests many rounding possibilities as explored in previous papers [Kho02, Tre05, CMM06a], but counterexamples [KV05] show that this edge-by-edge analysis can only go so far: high correlation for edges does not by itself imply that a good global assignment exists.

The main idea in our work is to try to understand and exploit correlations in the vector assignments for vertices that are not necessarily adjacent. If i, j are not adjacent vertices we can try to identify the correlation between their vector assignments by noting that since the $\mathbf{V}_{(j,l')}$'s are mutually orthogonal, for each $\mathbf{V}_{(i,l)}$ there is at most one $\mathbf{V}_{(j,l')}$ such that $\left\langle \overline{\mathbf{V}}_{(i,l)}, \overline{\mathbf{V}}_{(j,l')} \right\rangle > 1/\sqrt{2}$. Thus we can set up a maximal partial matching among their labels where the matching contains label pairs (l, l') such that $\left\langle \overline{\mathbf{V}}_{(i,l)}, \overline{\mathbf{V}}_{(j,l')} \right\rangle > 1/\sqrt{2}$. The vector assignments to the two vertices can be thought of as highly correlated if the sum of squared ℓ_2 norm of all the $\mathbf{V}_{(i,l)}$'s (resp, all $\mathbf{V}_{(l,l')}$'s) involved in this matching is close to 1. (This is a rough idea; see precise definition later.)

Our main contribution is to show that if the constraint graph is an expander then high correlation over edges necessarily implies high expected correlation between a randomly-chosen pair of vertices (which may be quite distant in the constraint graph). We also show that this allows us to construct a good global assignment. This is formalized below.

7.1.2 Rounding procedure and correctness proof

Now we describe our randomized rounding procedure $\mathcal{R}$, which outputs a labeling $\Lambda_{\text{alg}}\colon V \to [q]$. This uses a more precise version of the greedy matching outlined in the above overview. For a pair i, j of vertices (possibly nonadjacent), let $\sigma_{ij}\colon [q] \to [q]$ be a bijective mapping that maximizes $\sum_{l\in[q]} \left\langle \mathbf{V}_{(i,l)}, \mathbf{V}_{(j,\sigma_{ij}(l))} \right\rangle$; note that it can be efficiently found using max-weight bipartite matching. The procedure is as follows:

1. Pick a random vertex i.
2. Pick a label l for i from the distribution, where every label $l' \in [q]$ has probability $\left|\mathbf{V}_{(i,l')}\right|^2$.
3. Define $\Lambda_{\text{alg}}(j) := \sigma_{ij}(l)$ for every vertex $j \in V$.

(Of course, the rounding can be trivially derandomized since there are only nq choices for i, l.)

To analyze this procedure we define the *distance* $\rho(i, j)$ of a pair i, j of vertices as

$$\rho(i, j) \;:=\; \frac{1}{2}\sum_{l\in[k]} \left|\mathbf{V}_{(i,l)} - \mathbf{V}_{(j,\sigma_{ij}(l))}\right|^2 \;=\; 1 - \sum_{l\in[k]} \left\langle \mathbf{V}_{(i,l)}, \mathbf{V}_{(j,\sigma_{ij}(l))} \right\rangle$$

We think of two vertices i, j as highly correlated if $\rho(i, j)$ is small, i.e., $\sum_{l\in[q]} \left\langle \mathbf{V}_{(i,l)}, \mathbf{V}_{(j,\sigma_{ij}(l))} \right\rangle \approx 1$.

The following easy lemma shows that if the average vertex pair in G is highly correlated, then the above rounding procedure is likely to produce a good a labeling. Here we assume that G is a regular graph. Using standard arguments, all results can be generalized to the case of non-regular graphs.

Lemma 7.2 (Global Correlation ⇒ High Satisfiability)
The expected fraction of constraints satisfied by the labeling Λ_{alg} computed by the rounding procedure is

$$\mathop{\mathbb{E}}_{\Lambda_{\text{alg}}\leftarrow\mathcal{R}}[\,\mathsf{val}(\Lambda_{\text{alg}})\,] \geq 1 - 6\eta - 12 \mathop{\mathbb{E}}_{i,j\in V}[\,\rho(i, j)\,].$$

It is easy to see that if the SDP value is $1-\eta$ then the average correlation on edges is high. For an edge $e=(i,j)$ in G, let $\eta_e := \frac{1}{2}\sum_{l\in[q]}\left|\mathbf{V}_{(i,l)}-\mathbf{V}_{(j,\pi_{ij}(l))}\right|^2$. Note, $\mathbb{E}_e[\eta_e]=\eta$. Then we have

$$\begin{aligned}\rho(i,j) &= \frac{1}{2}\sum_{l\in[q]}\left|\mathbf{V}_{(i,l)}-\mathbf{V}_{(j,\sigma_{ij}(l))}\right|^2 = 1-\sum_{l\in[q]}\left\langle \mathbf{V}_{(i,l)},\mathbf{V}_{(j,\sigma_{ij}(l))}\right\rangle \\ &\leq 1-\sum_{l\in[q]}\left\langle \mathbf{V}_{(i,l)},\mathbf{V}_{(j,\pi_{ij}(l))}\right\rangle = \eta_e \qquad \text{(since } \sigma_{ij} \text{ is the maximum weight matching).}\end{aligned}$$

As mentioned in the overview, we show that high correlation on edges implies (when the constraint graph is an expander) high correlation on the average pair of vertices. The main technical contribution in this proof is a way to view a vector solution to the above SDP as a vector solution for Sparsest Cut. This involves mapping any sequence of q vectors to a *single* vector in a nicely continuous way, which allows us to show that the distances $\rho(i,j)$ essentially behave like squared Euclidean distances.

Lemma 7.3 (Low Distortion Embedding of ρ)
For every positive even integer t and every SDP solution $\{\mathbf{V}_{(i,l)}\}_{i\in V,l\in[k]}$, there exists a set of vectors $\{\mathbf{W}_i\}_{i\in V}$ such that for every pair i,j of vertices

$$\tfrac{1}{2t}\left|\mathbf{W}_i-\mathbf{W}_j\right|^2 \leq \rho(i,j) \leq \left|\mathbf{W}_i-\mathbf{W}_j\right|^2 + O(2^{-t/2})$$

Corollary 7.4 (Local Correlation $\Rightarrow$ Global Correlation)

$$\mathop{\mathbb{E}}_{i,j\in V}[\,\rho(i,j)\,] \leq \tfrac{2t}{\lambda}\mathop{\mathbb{E}}_{(i,j)\in E}[\,\rho(i,j)\,] + O(2^{-t/2}) \leq 2t\eta/\lambda + O(2^{-t/2}).$$

Proof: We use the following characterization of λ for regular graphs G

$$\lambda = \min \frac{\mathbb{E}_{(i,j)\in E}\left|\mathbf{z}_i-\mathbf{z}_j\right|^2}{\mathbb{E}_{i,j\in V}\left|\mathbf{z}_i-\mathbf{z}_j\right|^2}, \tag{7.2}$$

where the minimum is over all sets of vectors $\{\mathbf{z}_i\}_{i\in V}$. This characterization also shows that λ scaled by $n^2/|E|$ is a relaxation for the Sparsest Cut problem $\min_{\emptyset\neq S\subset V}|E(S,\overline{S})|/|S||\overline{S}|$ of G . Now using the previous Lemma we have

$$\begin{aligned}\mathop{\mathbb{E}}_{i,j\in V}[\,\rho(i,j)\,] &\leq \mathop{\mathbb{E}}_{i,j\in V}\left|\mathbf{W}_i-\mathbf{W}_j\right|^2 + O(2^{-t/2}) \\ &\leq \tfrac{1}{\lambda}\mathop{\mathbb{E}}_{(i,j)\in E}\left|\mathbf{W}_i-\mathbf{W}_j\right|^2 + O(2^{-t/2}) \\ &\leq \tfrac{2t}{\lambda}\mathop{\mathbb{E}}_{(i,j)\in E}[\,\rho(i,j)\,] + O(2^{-t/2}).\end{aligned}$$

■

By combining the Corollary 7.4 and Lemma 7.2, we can show the following theorem.

Theorem 7.5 (implies Theorem 7.1)
There is a polynomial time algorithm that computes a labeling Λ *with*

$$\mathsf{val}(\Lambda) \geq 1 - O\left(\tfrac{\eta}{\lambda}\log\left(\tfrac{\lambda}{\eta}\right)\right)$$

if the optimal value of the SDP *in Figure 7.1 for* $\mathcal{U}$ *is* $1-\eta$.

Proof: By Corollary 7.4 and Lemma 7.2, the labeling Λ_{alg} satisfies a $1-O(t\eta/\lambda+2^{-t/2})$ fraction of the constraints of $\mathcal{U}$. If we choose t to be an integer close to $2\log(\lambda/\eta)$, it follows that $\mathsf{OPT}(\mathcal{U}) \geq 1 - O(\frac{\eta}{\lambda}\log(\frac{\lambda}{\eta}))$. Since the rounding procedure $\mathcal{R}$ can easily be derandomized, a labeling Λ with $\mathsf{val}(\Lambda) \geq 1 - O(\frac{\eta}{\lambda}\log(\frac{\lambda}{\eta}))$ can be computed in polynomial time. ■

7.1.3 Proof of Lemma 7.2

We consider the labeling Λ_{alg} computed by the randomized rounding procedure $\mathcal{R}$. Recall that $\Lambda_{\text{alg}}(j) = \sigma_{ij}(l)$ where the vertex i is chosen uniformly at random and the label l is chosen with probability $\left|\mathbf{V}_{(i,l)}\right|^2$. For notational ease we assume that σ_{ii} is the identity permutation and σ_{ij} is the inverse permutation of σ_{ji}. The following claim gives an estimate on the probability that the constraint between an edge $e = \{j,k\}$ is satisfied by Λ_{alg}. Here we condition on the choice of u.

Claim 7.6 *For every vertex u and every edge* $e = (j,k)$*,*

$$\mathop{\mathbb{P}}_{\Lambda_{\text{alg}}}\left[\Lambda_{\text{alg}}(k) \neq \pi_{jk}(\Lambda_{\text{alg}}(j)) \mid i\right] \leq 6 \cdot (\rho(i,j) + \eta_e + \rho(k,i))$$

Proof: We may assume that both σ_{ij} and σ_{ik} are the identity permutation. Let $\pi = \pi_{jk}$. First note $\mathbb{P}_{\Lambda_{\text{alg}}}\left[\Lambda_{\text{alg}}(k) \neq \pi(\Lambda_{\text{alg}}(j)) \mid i\right] = \sum_{l\in[q]}\left[\left|\mathbf{V}_{(i,l)}\right|^2 \chi_{l\neq\pi(l)}\right]$, where $\chi_{\mathcal{E}}$ denotes the indicator random variable for an event $\mathcal{E}$. By orthogonality of the vectors $\{\mathbf{V}_{(i,l)}\}_{l\in[q]}$, it follows that

$$\sum_{l\in[q]}\left[\left|\mathbf{V}_{(i,l)}\right|^2 \chi_{l\neq\pi(l)}\right] \leq \sum_{l\in[q]}\left[\left(\left|\mathbf{V}_{(i,l)}\right|^2 + \left|\mathbf{V}_{(l,\pi(l))}\right|^2\right)\chi_{l\neq\pi(l)}\right] = \sum_{l\in[q]}\left|\mathbf{V}_{(i,l)} - \mathbf{V}_{(i,\pi(l))}\right|^2.$$

By triangle inequality, $\left|\mathbf{V}_{(i,l)} - \mathbf{V}_{(i,\pi(l))}\right| \leq \left|\mathbf{V}_{(i,l)} - \mathbf{V}_{(j,l)}\right| + \left|\mathbf{V}_{(j,l)} - \mathbf{V}_{(k,\pi(l))}\right| + \left|\mathbf{V}_{(k,\pi(l))} - \mathbf{V}_{(i,\pi(l))}\right|$. Now we square both sides of the inequality and take summations, $\sum_{l\in[q]}\left|\mathbf{V}_{(i,l)} - \mathbf{V}_{(i,\pi(l))}\right|^2 \leq 3\sum_{l\in[q]}\left|\mathbf{V}_{(i,l)} - \mathbf{V}_{(j,l)}\right|^2 + 3\sum_{l\in[q]}\left|\mathbf{V}_{(j,l)} - \mathbf{V}_{(k,\pi(l))}\right|^2 + 3\sum_{l\in[q]}\left|\mathbf{V}_{(k,\pi(l))} - \mathbf{V}_{(i,\pi(l))}\right|^2 = 6\rho(i,j) + 6\eta_e + 6\rho(w,u)$. ■

Proof: [of Lemma 7.2] From Claim 7.6 it follows

$$\mathop{\mathbb{E}}_{\Lambda_{\text{alg}}}[\,\mathsf{val}(\Lambda_{\text{alg}})\,] \geq 1 - 6 \mathop{\mathbb{E}}_{i\in V}\mathop{\mathbb{E}}_{e=(jk)\in E}\left[\rho(i,j) + \eta_e + \rho(k,i)\right].$$

Since G is a regular graph, both (i,j) and (k,i) are uniformly distributed over all pairs of vertices. Hence $\mathbb{E}_{\Lambda_{\text{alg}}}[\,\mathsf{val}(\Lambda_{\text{alg}})\,] \geq 1 - 6\eta - 12\,\mathbb{E}_{i,j\in V}[\,\rho(i,j)\,]$. ■

7.1.4 Proof of Lemma 7.3; the tensoring trick

Let t be a positive even integer and $\{\mathbf{V}_{(i,l)}\}_{i\in V, l\in[q]}$ be an SDP solution for $\mathcal{U}$. Define $\overline{\mathbf{V}}_{(i,l)} = \frac{1}{|\mathbf{V}_{(i,l)}|}\mathbf{V}_{(i,l)}$ and $\mathbf{W}_i = \sum_{l\in[q]} \left|\mathbf{V}_{(i,l)}\right| \overline{\mathbf{V}}_{(i,l)}^{\otimes t}$, where $\otimes t$ denotes t-wise tensoring. Notice that the vectors $\mathbf{W}_i$ are unit vectors. Consider a pair i, j of vertices in G. The following claim implies the lower bound on $\rho(i, j)$ in Lemma 7.3.

Claim 7.7 $\left|\mathbf{W}_i - \mathbf{W}_j\right|^2 \leq 2t\rho(i, j) = t \cdot \sum_{l\in[q]} \left|\mathbf{V}_{(i,l)} - \mathbf{V}_{(j,\sigma_{ij}(i))}\right|^2$

Proof: Since $\mathbf{W}_i$ is a unit vector for each i, it suffices to prove $\left\langle \mathbf{W}_i, \mathbf{W}_j \right\rangle \geq t\rho(i, j)$. Let $\sigma = \sigma_{ij}$. By Cauchy-Schwarz,

$$\textstyle\sum_l \left|\mathbf{V}_{(i,l)}\right| \left|\mathbf{V}_{(j,\sigma(l))}\right| \leq \left(\sum_l \left|\mathbf{V}_{(i,l)}\right|^2\right)^{1/2} \left(\sum_i \left|\mathbf{V}_{(j,\sigma(l))}\right|^2\right)^{1/2} \leq 1.$$

Thus there is some $\alpha \geq 1$ such that the following random variable X is well-defined: it takes value $\left\langle \overline{\mathbf{V}}_{(i,l)}, \overline{\mathbf{V}}_{(j,\sigma(l))} \right\rangle$ with probability $\alpha \cdot \left|\mathbf{V}_{(i,l)}\right| \left|\mathbf{V}_{(j,\sigma(i))}\right|$. By Jensen's Inequality, $(\mathbb{E}[X])^t \leq \mathbb{E}[X^t]$. Hence,

$$\begin{aligned} 1 - \rho(i, j)t \leq (1 - \rho(i, j))^t &= \left(\sum_{l\in[q]} \left[\left|\mathbf{V}_{(i,l)}\right| |V j\sigma(l)| \cdot \left\langle \overline{\mathbf{V}}_{(i,l)}, \overline{\mathbf{V}}_{(j,\sigma(l))} \right\rangle \right] \right)^t \\ &= (\mathbb{E}[X/\alpha])^t \leq (\mathbb{E}[X])^t/\alpha \leq \mathbb{E}[X^t/\alpha] = \left\langle \mathbf{W}_i, \mathbf{W}_j \right\rangle. \end{aligned}$$

This proves the claim. ■

Matching between two label sets. In order to finish the proof of Lemma 7.3, it remains to prove the upper bound on $\rho(i, j)$ in terms of the distance $\left|\mathbf{W}_i - \mathbf{W}_j\right|^2$. For this part of the proof, it is essential that the vectors $\mathbf{W}_i$ are composed of (high) tensor powers of the vectors $\mathbf{V}_{(i,l)}$. For a pair i, j of vertices, consider the following set of label pairs

$$M = \left\{ (l, l') \in [q] \times [q] \mid \left\langle \overline{\mathbf{V}}_{(i,l)}, \overline{\mathbf{V}}_{(j,l')} \right\rangle^2 > 1/2 \right\}.$$

Since $\{\overline{\mathbf{V}}_{(i,l)}\}_{l\in[q]}$ and $\{\overline{\mathbf{V}}_{(j,l')}\}_{l'\in[q]}$ are sets of ortho-normal vectors, M as graph on $[q] \times \{L, R\}$ is a (partial) matching. Let σ be an arbitrary permutation of $[q]$ that agrees with the M on the matched labels, i.e., for all $(l, l') \in M$, we have $\sigma(l) = l'$. The following claim shows the upper bound on $\rho(i, j)$ of Lemma 7.3.

Claim 7.8 $\rho(i, j) \;\leq\; \frac{1}{2} \sum_{l\in[q]} \left|\mathbf{V}_{(i,l)} - \mathbf{V}_{(j,\sigma(l))}\right|^2 \;\leq\; \frac{1}{2} \left|\mathbf{W}_i - \mathbf{W}_j\right|^2 + O(2^{-t/2})$.

Proof: Let $\delta = \left|\mathbf{W}_i - \mathbf{W}_j\right|^2$. Note that

$$\sum_{i,j} \left|\mathbf{V}_{(i,l)}\right| \left|\mathbf{V}_{(j,l')}\right| \left\langle \overline{\mathbf{V}}_{(i,l)}, \overline{\mathbf{V}}_{(j,l')} \right\rangle^t = 1 - \delta/2. \tag{7.3}$$

We may assume that σ is the identity permutation. Then, $\rho(i, j)$ is at most

$$\begin{aligned}\tfrac{1}{2}\sum_{l\in[q]}\left|\mathbf{V}_{(i,l)}-\mathbf{V}_{(j,l)}\right|^2 &= 1-\sum_{l\in[q]}\left\langle \mathbf{V}_{(i,l)},\mathbf{V}_{(j,l)}\right\rangle \\ &\le 1-\sum_{i\in[k]}\left|\mathbf{V}_{(i,l)}\right|\left|\mathbf{V}_{(j,l)}\right|\left\langle \overline{\mathbf{V}}_{(i,l)},\overline{\mathbf{V}}_{(j,l)}\right\rangle^t \qquad \left(\text{using } \left\langle \mathbf{V}_{(i,l)},\mathbf{V}_{(j,l)}\right\rangle \ge 0\right) \\ &= \delta/2+\sum_{l\neq l'}\left|\mathbf{V}_{(i,l)}\right|\left|\mathbf{V}_{(j,l')}\right|\left\langle \overline{\mathbf{V}}_{(i,l)},\overline{\mathbf{V}}_{(j,l')}\right\rangle^t \qquad (\text{by (7.3)}) \\ &= \delta/2+\langle \mathbf{p}, A\,\mathbf{q}\rangle\,,\end{aligned}$$

where $p_i = \left|\mathbf{V}_{(i,l)}\right|$, $q_j = \left|\mathbf{V}_{(j,l')}\right|$, $A_{ii} = 0$, and for $i \neq j$, $A_{ij} = \left\langle \overline{\mathbf{V}}_{(i,l)},\overline{\mathbf{V}}_{(j,l')}\right\rangle^t$. Since both $\mathbf{p}$ and $\mathbf{q}$ are unit vectors, $\langle \mathbf{p}, A\,\mathbf{q}\rangle$ is bounded by the largest singular value of A. As the matrix A has only non-negative entries, its largest singular value is bounded by the maximum sum of a row or a column. By symmetry, we may assume that the first row has the largest sum among all rows and columns. We rearrange the columns in such a way that $A_{11} \ge A_{12} \ge \ldots \ge A_{1q}$. Since $\overline{\mathbf{V}}_{(i,1)}$ is a unit vector and $\{\overline{\mathbf{V}}_{(j,l)}\}_{l\in[q]}$ is a set of orthonormal vectors, we have $\sum_l \left\langle \overline{\mathbf{V}}_{(i,1)},\overline{\mathbf{V}}_{(j,l)}\right\rangle^2 \le 1$. Hence, $\left\langle \overline{\mathbf{V}}_{(i,1)},\overline{\mathbf{V}}_{(j,l)}\right\rangle^2 \le 1/l$ and therefore $A_{1l} \le (1/l)^{t/2}$. On the other hand, every entry of A is at most $2^{-t/2}$, since all pairs (l, l') with $\left\langle \overline{\mathbf{V}}_{(i,l)},\overline{\mathbf{V}}_{(j,l')}\right\rangle^2 > 1/2$ participate in the matching M, and hence, $A_{ll'} = 0$ for all l, l' with $\left\langle \overline{\mathbf{V}}_{(i,l)},\overline{\mathbf{V}}_{(j,l')}\right\rangle^2 > 1/2$. It follows that the sum of the first row can be upper bounded by

$$\sum_{l\in[q]} A_{1l} \le A_{11} + \sum_{l\ge 2}^{\infty}(\tfrac{1}{l})^{t/2} \le 2^{-t/2} + \sum_{l\ge 2}^{\infty}(\tfrac{1}{l})^{t/2} = O(2^{-t/2}).$$

We conclude that the largest singular value of A is at most $O(2^{-t/2})$, and thus $\rho(i, j)$ can be upper bounded by $\delta/2 + O(2^{-t/2}) = \frac{1}{2}\left|\mathbf{W}_i - \mathbf{W}_j\right| + O(2^{-t/2})$, as claimed. ■

7.2 Stronger relaxations of expansion

In this section, we consider stronger SDP relaxations for for Sparsest Cut. A systematic way to obtain stronger relaxations is again provided by SDP hierarchies. Here, we state our results in terms of the relaxations given by the Lasserre hierarchy [Las01]. The results in this section apply only to a special case of Unique Games, called ΓMAX2LIN. We say a Unique Games instance $\mathcal{U} = (G(V, E), [q], \{\pi_{ij}\}_{(i,j)\in E})$ has ΓMAX2LIN form, if the label set $[q]$ can be identified with the group $\mathbb{Z}_q$ in such a way that every constraint permutation π_{ij} satisfies $\pi_{ij}(l + s) = \pi_{ij}(l) + s \in \mathbb{Z}_q$ for all $s, l \in \mathbb{Z}_q$. In other words, π_{ij} encodes a constraint of the form $x_i - x_j = c_{ij} \in \mathbb{Z}_q$. The ΓMAX2LIN property implies that we can find an optimal SDP solution $\{\mathbf{V}_{(i,l)}\}_{i\in V, l\in[q]}$ for $\mathcal{U}$ that is *shift-invariant*, i.e., for all $s \in \mathbb{Z}_q$ we have $\left\langle \mathbf{V}_{(i,l+s)},\mathbf{V}_{(j,l'+s)}\right\rangle = \left\langle \mathbf{V}_{(i,l)},\mathbf{V}_{(j,l')}\right\rangle$. In particular, every vector $\mathbf{V}_{(i,l)}$ has norm $1/\sqrt{q}$.

Alternative Embedding for ΓMAX2LIN. The following lemma can be seen as alternative to Lemma 7.3. We emphasize that the lemma only holds for ΓMAX2LIN instances and shift-invariant SDP solutions.

Lemma 7.9 *Let Λ_{opt} be a labeling for $\mathcal{U}$ with $\mathsf{val}(\Lambda_{\text{opt}}) = 1 - \varepsilon$. Then the set of vectors $\{\mathbf{W}_i\}_{i \in V}$ with $\mathbf{W}_i = \mathbf{V}_{(i,\Lambda_{\text{opt}}(i))}$ has the following two properties:*

1. $\rho(i,j) \leq \frac{q}{2} \left|\mathbf{W}_i - \mathbf{W}_j\right|^2$ *for every pair i, j of vertices*
2. $\frac{q}{2}\, \mathbb{E}_{(i,j)\in E}\left[\left|\mathbf{W}_i - \mathbf{W}_j\right|^2\right] \leq \eta + 2\varepsilon$

Together with Lemma 7.2, the above lemma implies that the randomized rounding procedure $\mathcal{R}$ computes a labeling that satisfies at least a $1 - O(\varepsilon/\lambda)$ fraction of the constraints of $\mathcal{U}$, whenever $\mathsf{OPT}(\mathcal{U}) \geq 1 - \varepsilon$. In this sense, the above lemma allows to prove our main result for the special case of ΓMAX2LIN.

Proof: Item 1 holds, since, by shift invariance,

$$\rho(i,j) = \frac{1}{2}\sum_{l\in[q]} \left|\mathbf{V}_{(i,l)} - \mathbf{V}_{(j,\sigma_{ij}(l))}\right|^2 = \frac{q}{2}\left|\mathbf{V}_{(i,\Lambda_{\text{opt}}(i))} - \mathbf{V}_{(j,\sigma_{ij}(\Lambda_{\text{opt}}(i)))}\right|^2 \leq \frac{q}{2}\left|\mathbf{V}_{(i,\Lambda_{\text{opt}}(i))} - \mathbf{V}_{(j,\Lambda_{\text{opt}}(j))}\right|^2.$$

Here we could assume, again by shift invariance, that $\left|\mathbf{V}_{(i,l)} - \mathbf{V}_{(j,\sigma_{ij}(l))}\right|^2 = \min_{l'} \left|\mathbf{V}_{(i,l)} - \mathbf{V}_{(j,l')}\right|^2$ for all l. It remains to verify Item 2. By shift invariance,

$$\eta_{ij} = \frac{1}{2}\sum_{l\in[q]} \left|\mathbf{V}_{(i,l)} - \mathbf{V}_{(j,\pi_{ij}(l))}\right|^2 = \frac{q}{2}\left|\mathbf{V}_{(i,\Lambda_{\text{opt}}(i))} - \mathbf{V}_{(j,\pi_{ij}(\Lambda_{\text{opt}}(i)))}\right|^2.$$

Hence, if Λ_{opt} satisfies the constraint on an edge $(i,j) \in E$, then $\left|\mathbf{W}_i - \mathbf{W}_j\right| = 2\eta_{ij}$. On the other hand, $\left|\mathbf{W}_i - \mathbf{W}_j\right|^2 \leq 4$ because every vector $\mathbf{W}_i$ has unit norm. Finally, since a $1 - \varepsilon$ fraction of the edges is satisfied by Λ_{opt},

$$\mathop{\mathbb{E}}_{(i,j)\in E} \left|\mathbf{W}_i - \mathbf{W}_j\right|^2 \leq (1-\varepsilon) \cdot \mathop{\mathbb{E}}_{(i,j)\in E} [\, 2\eta_{ij}\,] + \varepsilon \cdot 4.$$

■

Stronger Relaxations for Sparsest Cut. Let r be a positive integer. Denote by $\mathcal{S}$ the set of all subsets of V that have cardinality at most r. For every subset $S \in \mathcal{S}$, we have a variable $\mathbf{U}_S$. We consider the strengthening of the spectral relaxation for Sparsest Cut in Figure 7.2.

The variables $\mathbf{U}_S$ are intended to have values $\mathbf{0}$ or $\mathbf{1}$, where $\mathbf{1}$ is some fixed unit vector. If the intended cut is $(T, \overline{T})$, we would assign $\mathbf{1}$ to all variables $\mathbf{U}_S$ with $S \subseteq T$.

Let $z_r(G)$ denote the optimal value of the above SDP. We have

$$\lambda \leq z_1(G) \leq \ldots \leq z_n(G) = \frac{n^2}{|E|} \min_{\emptyset \neq T \subset V} \frac{|E(T,\overline{T})|}{|T||\overline{T}|}.$$

It can also be seen that the relaxation $z_3(G)$ is at least as strong as the relaxation for Sparsest Cut considered in [ARV04]. The relaxations $z_r(G)$ are inspired by the Lasserre SDP hierarchy [Las01].

$$\begin{array}{lll} \text{minimize} & \dfrac{\mathbb{E}_{(i,j)\in E}\left[\left|\mathbf{U}_{\{i\}}-\mathbf{U}_{\{j\}}\right|^2\right]}{\mathbb{E}_{i,j\in V}\left[\left|\mathbf{U}_{\{i\}}-\mathbf{U}_{\{j\}}\right|^2\right]} & \\ \text{subject to} & \langle \mathbf{U}_{S_1},\mathbf{U}_{S_2}\rangle = \langle \mathbf{U}_{S_3},\mathbf{U}_{S_4}\rangle & \forall\, S_1\cup S_2 = S_3\cup S_4 \\ & \langle \mathbf{U}_{S_1},\mathbf{U}_{S_2}\rangle \in [0,1] & \forall\, S_1,S_2 \\ & |\mathbf{U}_\emptyset| = 1 & \end{array}$$

Figure 7.2: Lasserre SDP for Sparsest Cut

The proof of the following theorem is similar to the proof of Theorem 7.5. The main difference is that we use Lemma 7.9, instead of Lemma 7.3, in order to show that local correlation implies global correlation. By strengthening the SDP for Unique Games, the vectors $\mathbf{W}_i$ obtained from Lemma 7.9 can be extended to a solution for the stronger SDP for Sparsest Cut in Figure 7.2. This allows us to replace the parameter λ by the parameter $z_r(G)$ in the below theorem.

Theorem 7.10 *There is an algorithm that computes in time $(nq)^{O(r)}$ a labeling Λ with*

$$\mathsf{val}(\Lambda) \geq 1 - O(\varepsilon/z_r(G))$$

if $\mathsf{OPT}(\mathcal{U}) \geq 1-\varepsilon$ *and* $\mathcal{U}$ *has* ΓMAX2LIN *form.*

Proof: We consider the level-r Lasserre relaxation for Unique Games. Since, Unique Games is a specific constraint satisfaction problem, the SDP can be written in terms of sets S of at most r vertices and partial assignments α. We give the relaxation in Figure 7.3.

$$\begin{array}{lll} \text{maximize} & \mathbb{E}_{(i,j)\in E}\left[\sum_{l\in[q]}\left\langle \mathbf{V}_{(\{i\},l)},\mathbf{V}_{(\{j\},\pi_{ij}(l))}\right\rangle\right] & \\ \text{subject to} & \left\langle \mathbf{V}_{(S_1,\alpha_1)},\mathbf{V}_{(S_2,\alpha_2)}\right\rangle = 0 & \forall\, \alpha_1(S_1\cap S_2)\neq\alpha_2(S_1\cap S_2) \\ & \left\langle \mathbf{V}_{(S_1,\alpha_1)},\mathbf{V}_{(S_2,\alpha_2)}\right\rangle = \left\langle \mathbf{V}_{(S_3,\alpha_3)},\mathbf{V}_{(S_4,\alpha_4)}\right\rangle & \forall\, S_1\cup S_2 = S_3\cup S_4,\ \alpha_1\circ\alpha_2=\alpha_3\circ\alpha_4 \\ & \sum_{j\in[q]}\left|\mathbf{V}_{(\{i\},j)}\right|^2 = 1 & \forall i\in V \\ & \left\langle \mathbf{V}_{(S_1,\alpha_1)},\mathbf{V}_{(S_2,\alpha_2)}\right\rangle \geq 0 & \forall S_1,S_2,\alpha_1,\alpha_2 \\ & \left|\mathbf{V}_{(\emptyset,\emptyset)}\right| = 1 & \end{array}$$

Figure 7.3: Lasserre SDP for Unique Games

An (approximately) optimal solution to the above SDP can be computed in time $(nq)^{O(r)}$. Suppose the value of the solution is $1-\eta$. Let $\mathbf{W}_i := \sqrt{q}\cdot\mathbf{V}_{(\{i\},\Lambda_{\text{opt}}(i))}$ for some labeling Λ_{opt} with $\mathsf{val}(\Lambda_{\text{opt}}) \geq 1-\varepsilon$ (which exists by assumption).

We claim that $\mathbb{E}_{i,j\in V}\left|\mathbf{W}_i-\mathbf{W}_j\right|^2 \leq \frac{1}{z_r(G)}\,\mathbb{E}_{(i,j)\in E}\left|\mathbf{W}_i-\mathbf{W}_j\right|^2$. In order to show the claim it is sufficient to show that the vectors $\mathbf{W}_i$ can be extended to a solution for the SDP in Figure 7.2 with

$\mathbf{U}_{\{i\}} = \mathbf{W}_i \ \forall i$. We can choose the vectors $\mathbf{U}_S$ as $\mathbf{U}_S = \mathbf{V}_{(S,\Lambda_{\text{opt}}(S))}$ and $\mathbf{U}_\emptyset = \mathbf{V}_{(\emptyset,\emptyset)}$. It is easy to verify that these vectors satisfy the required constraints.

Together with Lemma 7.9, this gives that

$$\mathop{\mathbb{E}}_{i,j\in V}[\rho(i,j)] \le \tfrac{q}{2} \mathop{\mathbb{E}}_{i,j\in V} \left|\mathbf{W}_i - \mathbf{W}_j\right|^2 \le \tfrac{q}{2z_r(G)} \mathop{\mathbb{E}}_{(i,j)\in E} \left|\mathbf{W}_i - \mathbf{W}_j\right|^2 \le (\eta + 2\varepsilon)/z_r(G).$$

By Lemma 7.2, the rounding procedure $\mathcal{R}$ from Section 7.1.2 allows us to compute a labeling Λ such that $\mathsf{val}(\Lambda) \ge 1 - 6\eta - 12\,\mathbb{E}_{i,j\in V}[\rho(i,j)] \ge 1 - 50\varepsilon/z_r(G)$. Here it is important to note that the rounding procedure $\mathcal{R}$ does not depend on the vectors $\mathbf{W}_i$. The existence of these vectors is enough to conclude that the rounding procedure succeeds. ■

7.3 Parallel Repetition for expanding Unique Games

In this section, we consider *bipartite* unique games, i.e., Unique Games instances $\mathcal{U} = (G(V,W,E),[q],\{\pi_{ij}\}_{(i,j)\in E})$ such that $G(V,W,E)$ is a bipartite graph with bipartition (V,W). A bipartite unique game can be seen as a 2-prover, 1-round proof system [FL92]. The two parts V, W correspond to the two provers. The edge set E corresponds to the set of questions asked by the verifier to the two provers and π_{ij} is the accepting predicate for the question corresponding to the edge (v,w).

In this section, we give an upper bound on the *amortized value* $\overline{\omega}(\mathcal{U}) = \sup_r \mathsf{OPT}(\mathcal{U}^{\otimes r})^{1/r}$ of bipartite unique game $\mathcal{U}$ in terms of the expansion of its constraint graph. Here $\mathcal{U}^{\otimes r}$ denotes the game obtained by playing the game $\mathcal{U}$ for r rounds in parallel. We follow an approach proposed by Feige and Lovsz [FL92]. Their approach is based on the SDP in Figure 7.4, which is a relaxation for the value of a bipartite unique game.

$$\begin{array}{lll}
\text{maximize} & \mathop{\mathbb{E}}_{e=(i,j)\in E}\left[\sum_{l\in[q]} \left\langle \mathbf{V}_{(i,l)}, \mathbf{V}_{(j,\pi_{ij}(l))}\right\rangle\right] & \\
\text{subject to} & \left\langle \mathbf{V}_{(i,l)}, \mathbf{V}_{(j,l')}\right\rangle \ge 0 & \forall\, i\in V, j\in W,\ \forall\, l,l' \\
& \sum_{l,l'\in[q]} \left|\left\langle \mathbf{V}_{(i,l)}, \mathbf{V}_{(j,l')}\right\rangle\right| \le 1 & \forall\, i,j\in V \\
& \sum_{l,l'\in[q]} \left|\left\langle \mathbf{V}_{(i,l)}, \mathbf{V}_{(j,l')}\right\rangle\right| \le 1 & \forall\, i,j\in W
\end{array}$$

Figure 7.4: Feige-Lovász SDP for Unique Games

Let $\overline{\sigma}(\mathcal{U})$ denote the value of this SDP relaxation. The following theorem is a consequence of the fact that $\overline{\sigma}(\mathcal{U}^{\otimes r}) = \overline{\sigma}(\mathcal{U})^r$.

Theorem 7.11 ([FL92]) *For every bipartite unique game $\mathcal{U}$, $\overline{\omega}(\mathcal{U}) \le \overline{\sigma}(\mathcal{U})$.*

We observe that the SDP in Figure 7.1 cannot be much stronger than the relaxation $\overline{\sigma}(\mathcal{U})$. The proof uses mostly standard arguments.

Lemma 7.12 *If $\overline{\sigma}(\mathcal{U}) = 1 - \eta$ then the value of the* SDP *in Figure 7.1 is at least* $1 - 2\eta$.

Proof: Let $\{\tilde{\mathbf{V}}_{(i,l)}\}_{i \in V \cup W, l \in [q]}$ be an optimal solution to the SDP in Figure 7.4. We have $\mathbb{E}_{(i,j) \in E} \sum_{l \in [q]} \left\langle \tilde{\mathbf{V}}_{(i,l)}, \tilde{\mathbf{V}}_{(j,\pi_{ij}(l))} \right\rangle = 1 - \eta$.

We first show how to obtain a set of vectors $\{\mathbf{V}_{(i,l)}\}_{i \in V \cup W, l \in [q]}$ that satisfies the first two constraints of the SDP in Figure 7.1 and has objective value $\mathbb{E}_{(i,j) \in E} \sum_{l \in [q]} \left\langle \mathbf{V}_{(i,l)}, \mathbf{V}_{(j,\pi_{ij}(l))} \right\rangle = 1 - \eta$. For each vertex i, consider the psd matrix $M^{(i)} \in \mathbb{R}^{q \times q}$ such that $M^{(i)}_{l,l'} = \left\langle \tilde{\mathbf{V}}_{(i,l)}, \tilde{\mathbf{V}}_{(i,l')} \right\rangle$ and define the quantities

$$e_i \;:=\; 1 - \sum_{l,l'} \left| \left\langle \tilde{\mathbf{V}}_{(i,l)}, \tilde{\mathbf{V}}_{(i,l')} \right\rangle \right| \quad \text{and} \quad s_{i,l} \;:=\; \sum_{l' \neq l} \left| \left\langle \tilde{\mathbf{V}}_{(i,l)}, \tilde{\mathbf{V}}_{(i,l')} \right\rangle \right|.$$

We consider the *diagonal* matrix $D^{(i)}$ with entries $D^{(i)}_{l,l} = \left| \tilde{\mathbf{V}}_{(i,l)} \right|^2 + s_{i,l} + (e_i/q)$. We will show how to construct vectors $\mathbf{V}_{(i,l)}$ satisfying $\langle \mathbf{V}_{(i,l)}, \mathbf{V}_{(i,l')} \rangle = D^{(i)}_{l,l'}$. It is easy to check that this suffices to satisfy the first two constraints of the SDP in Figure 7.1.

Notice that the matrix $D^{(i)} - M^{(i)}$ is positive semidefinite, since it is diagonally dominant. Hence, there exist vectors $\{\mathbf{z}_{(i,l)}\}_{l \in [q]}$ such that $\langle \mathbf{z}_{(i,l)}, \mathbf{z}_{(i,l')} \rangle = D^{(i)}_{l,l'} - M^{(i)}_{l,l'}$. Consider a set of vectors $\{\mathbf{W}_{(i,l)}\}_{i \in V \cup W, l \in [q]}$ such that $\langle \mathbf{W}_{(i,l)}, \mathbf{W}_{(i,l')} \rangle = \langle \mathbf{z}_{(i,l)}, \mathbf{z}_{(i,l')} \rangle$ and $\left\langle \mathbf{W}_{(i,l)}, \mathbf{W}_{(j,l')} \right\rangle$ for $i \neq j$ for any l, l'. Then the vectors $\mathbf{V}_{(i,l)} = \tilde{\mathbf{V}}_{(i,l)} \oplus \mathbf{W}_{(i,l)}$ satisfy the required properties. Also, observe that since $\left\langle \mathbf{V}_{(i,l)}, \mathbf{V}_{(j,l')} \right\rangle = \left\langle \tilde{\mathbf{V}}_{(i,l)}, \tilde{\mathbf{V}}_{(j,l')} \right\rangle$ for $i \neq j$, the objective value is unchanged.

It remains to show how to obtain a set of vectors that satisfies the non-negativity constraint and that has objective value at least $1 - 2\eta$. Consider the set of vectors $\{\mathbf{V}'_{(i,l)}\}_{i \in V \cup W, l \in [q]}$ with $\mathbf{V}'_{(i,l)} = \left| \mathbf{V}_{(i,l)} \right| \overline{\mathbf{V}}^{\otimes 2}_{(i,l)}$. The vectors $\mathbf{V}'_{(i,l)}$ still satisfy the first two constraints. They also satisfy the non-negativity constraint because $\left\langle \mathbf{V}'_{(i,l)}, \mathbf{V}'_{(j,l')} \right\rangle = \left| \mathbf{V}_{(i,l)} \right| \left| \mathbf{V}_{(j,l')} \right| \left\langle \overline{\mathbf{V}}_{(i,l)}, \overline{\mathbf{V}}_{(j,l')} \right\rangle^2 \geq 0$. We can use the same reasoning as in the proof of Claim 7.7 to show that the objective value $\mathbb{E}_{(v,w) \in E} \sum_{l \in [q]} \left\langle \mathbf{V}'_{(i,l)}, \mathbf{V}'_{(j,\pi_{ij}(l))} \right\rangle \geq 1 - 2\eta$. ■

We can now show a bound on the amortized value for the parallel repetition of the unique game defined by $\mathcal{U}$.

Theorem 7.13 *If $\mathcal{U}$ is 2-prover 1-round unique game on alphabet $[q]$ with value at most $1 - \varepsilon$, then the value $\mathcal{U}$ played in parallel for r rounds is at most $(1 - \Omega(\varepsilon \cdot \lambda / \log \frac{1}{\varepsilon}))^r$, where G is the graph corresponding to the questions to the two provers. In particular, the amortized value $\overline{w}(\mathcal{U})$ is at most $1 - \Omega(\varepsilon \cdot \lambda / \log \frac{1}{\varepsilon})$.*

Proof: Following the approach in [FL92], it is sufficient to show $\overline{\sigma}(\mathcal{U}) \leq 1 - \Omega(\varepsilon\lambda / \log \frac{1}{\varepsilon})$. Suppose that $\overline{\sigma}(\mathcal{U}) = 1 - \eta$. Then by Lemma 7.12, the value of the SDP in Figure 7.1 is at least $1 - 2\eta$. By Theorem 7.5, it follows that $\mathsf{OPT}(\mathcal{U}) \geq 1 - O(\eta \log \frac{\lambda}{\eta} / \lambda)$. On the other hand, $\mathsf{OPT}(\mathcal{U}) \leq 1 - \varepsilon$. Hence, $\varepsilon = O(\eta \log \frac{\lambda}{\eta} / \lambda)$ and therefore $\eta = \Omega(\lambda \varepsilon \log \frac{1}{\varepsilon})$, as claimed. ■

Chapter 8

Conclusions and Open Problems

The various results on integrality gaps presented in this thesis illustrate different proof techniques for reasoning about the hierarchies. While the intuition of thinking about "local distributions" defined by the solutions does seem quite helpful for reasoning about the integrality gaps, it still does not capture the full power of the semidefinite programs. Extending this intuition to capture the solutions to semidefinite programs will likely require a better (or different) understanding of the matrices containing moments of sets of variables, according to the respective distributions.

From these results, one may also conclude that while reductions between integrality gaps are somewhat difficult in the Lovász-Schrijver hierarchies, they are much easier to reason about in the Lasserre and Sherali-Adams hierarchies. While transforming an LS+ gap for MAX 3-XOR to Minimum Vertex Cover requires almost the full reasoning in chapter 3, the proof in [Sch08] for the Lasserre hierarchy is significantly simpler. Also, more complicated reductions can easily be reasoned about in the Lasserre hierarchy (and by very similar arguments, also in the Sherali-Adams hierarchy) as illustrated in chapter 6.

Below we outline some other problems which might be interesting to investigate in the context of these hierarchies.

Integrality Gaps for Vertex Cover. Even though a gap of factor $7/6$ for Minimum Vertex Cover is known for $\Omega(n)$ levels of the Lasserre hierarchy [Sch08], gaps of a factor $2 - \varepsilon$ are only known for $\Omega(n)$ levels of the LS hierarchy [STT07b]. Gaps for fewer levels, in particular, $\Omega(n^{f(\varepsilon)})$ levels of the Sherali-Adams hierarchy and $\Omega(\sqrt{\log n / \log\log n})$ levels of the LS+ hierarchy were proved by [CMM09] and [GMPT07] respectively. However, proving optimal (factor $2 - \varepsilon$ for $\Omega(n)$ levels) gaps in these hierarchies still remains a very interesting open question, particularly so for the LS+ hierarchy of semidefinite programs.

Also, for the Lasserre hierarchy, no results are known showing a gap of $2-\varepsilon$ even after 2 levels! Any argument which proves a gap of $2 - \varepsilon$ for a superconstant number of levels of the Lasserre hierarchy, would possibly contain some interesting new ideas.

Integrality Gaps for Unique Games. At present, very few integrality gaps are known for the Unique Games problem discussed in chapter 7. Since this problem is of great importance in complexity theory, and integrality gaps for it can be readily translated into gaps for other problems via reductions; integrality gaps for this problem would be of great interest.

An almost optimal integrality gap ($1 - \varepsilon$ vs $1/q^{O(\varepsilon)}$ for alphabet size q) for the basic SDP in Figure 7.1 was exhibited by Khot and Vishnoi [KV05]. Extending this, Khot and Saket [KS09] and Raghavendra and Steurer [RS09] show integrality gaps for a new hierarchy, which essentially adds constraints for existence of "local distributions" to the SDP. Constraints at t^{th} level of this hierarchy are all the LP constraints given by t levels of the Sherali-Adams hierarchy and also the SDP constraints in Figure 7.1. The SDP variables are related to the LP variables by requiring that $\left\langle \mathbf{V}_{(i,l)}, \mathbf{V}_{(j,l')} \right\rangle = X_{(\{i,j\},\{l,l'\})}$ for every pair of vertices i, j and labels l, l'. While [RS09] prove superconstant gaps for $\Omega((\log\log n)^{1/4})$ levels of this hierarchy, the results of [KS09] can be used to obtain a gap for $(\log\log\log n)^{\Omega(1)}$ levels. Also, for the Sherali-Adams LP hierarchy, [CMM09] prove optimal gaps for $n^{\Omega(1)}$ levels.

However, obtaining integrality gaps for a strong SDP relaxation of unique games, such as the ones given by LS+ or Lasserre hierarchies is an important open problem. Using reductions, this would also translate to Lasserre gaps for Sparsest Cut and many other interesting problems.

Generalizing the techniques of [Sch08]. At present the only available technique for obtaining lower bounds in the Lasserre hierarchy is the one used by [Sch08] (which extends the techniques of [FO06] and [STT07a] used in the context of other hierarchies). This technique was generalized to some extent in chapter 4 and combined with reductions in chapter 6, but it still suffers from two limitations:

- The technique may only be used when a CSP with m constraints can be shown to have SDP value m and the assignments in the support of the local distributions can be shown to satisfy *all* the constraints in the corresponding variables. However, in cases when the local distributions can only be shown to satisfy *almost all* the constraints, it is not clear if a modification of the technique can be used to argue (at least for some problems) that the SDP value is close to m.

- More importantly, the technique seems to be heavily dependent on the structure of the constraints as systems of linear equations. An extension of this technique to constraints which cannot be expressed in this form will likely involve a more general construction of vector solutions to the Lasserre SDP which might be useful in other contexts. Some attempts to understand the relevant local distributions for such constraints were made in chapter 5, but the question of obtaining Lasserre gaps for these constraints remains temptingly open.

Rank vs Size gaps for hierarchies. The goal of the studies so far has been to prove lower bounds on the *level* of the hierarchy one has to go to, in order to obtain a good approximation for the problem in question. In the language of proof complexity (where such hierarchies are viewed as proof systems to prove that the solution set of a certain integer program is empty), this corresponds to a lower bound on the "rank" of the proof.

However, the corresponding notion of "size" lower bounds in proof complexity translates to the *number of constraints* generated by these hierarchies that one really needs to obtain a certain approximation ratio for the given problem. More concretely, one can ask questions of the following form: what is the integrality gap for a semidefinite program for MAX 3-SAT, which uses only $n^{O(1)}$ of all the constraints generated by (level n) Lasserre SDP for the same problem. Viewing the hierarchies as a computational model, the number of constraints has an attractive interpretation as the "number of basic operations" in the model.

In fact, it is also not known if such lower bounds would be closed under reductions, which was the case for lower bounds on the number of levels required in the Lasserre hierarchy. In particular, is it possible to show that result proving a lower bound of $n^{\Omega(1)}$ constraints for 2-approximation of MAX 3-XOR can be translated to a lower bound of $n^{\Omega(1)}$ constraints for $7/6$-approximation of Minimum Vertex Cover?

Bibliography

[AAT05] Michael Alekhnovich, Sanjeev Arora, and Iannis Tourlakis. Towards strong nonapproximability results in the Lovasz-Schrijver hierarchy. In *Proceedings of the 37th ACM Symposium on Theory of Computing*, pages 294–303, 2005.

[ABL02] Sanjeev Arora, Béla Bollobás, and László Lovász. Proving integrality gaps without knowing the linear program. In *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science*, pages 313–322, 2002.

[ABLT06] Sanjeev Arora, Béla Bollobás, László Lovász, and Iannis Tourlakis. Proving integrality gaps without knowing the linear program. *Theory of Computing*, 2(2):19–51, 2006.

[AKK+08] Sanjeev Arora, Subhash Khot, Alexandra Kolla, David Steurer, Madhur Tulsiani, and Nisheeth Vishnoi. Unique games on expanding constraint graphs are easy. In *Proceedings of the 40th ACM Symposium on Theory of Computing*, 2008.

[AM08] Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. In *Proceedings of the 23rd IEEE Conference on Computational Complexity*, pages 249–258, Los Alamitos, CA, USA, 2008. IEEE Computer Society.

[AMS07] Christoph Ambhl, Monaldo Mastrolilli, and Ola Svensson. Inapproximability results for sparsest cut, optimal linear arrangement, and precedence constrained scheduling. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*, pages 329–337, 2007.

[ARV04] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows and a $\sqrt{\log n}$-approximation to sparsest cut. In *Proceedings of the 36th ACM Symposium on Theory of Computing*, 2004.

[BGS98] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCP's and non-approximability – towards tight results. *SIAM Journal on Computing*, 27(3):804–915, 1998. Preliminary version in *Proc. of FOCS'95*.

[BH92] R. Boppana and M.M. Halldórsson. Approximating maximum independent set by excluding subgraphs. *Bit*, 32:180–196, 1992.

[BOGH+03] Josh Buresh-Oppenheim, Nicola Galesi, Shlomo Hoory, Avner Magen, and Toniann Pitassi. Rank bounds and integrality gaps for cutting planes procedures. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science*, pages 318–327, 2003.

[BOT02] Andrej Bogdanov, Kenji Obata, and Luca Trevisan. A lower bound for testing 3-colorability in bounded degree graphs. In *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science*, pages 93–102, 2002.

[BSW01] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow: Resolution made simple. *Journal of the ACM*, 48(2), 2001.

[CCL90] Jin-Yi Cai, Anne Condon, and Richard J. Lipton. Playing games of incomplete information. In *Proceedings of the 7th Symposium on Theoretical Aspects of Computer Science*, pages 58–69, 1990.

[Cha02] Moses Charikar. On semidefinite programming relaxations for graph coloring and vertex cover. In *Proceedings of the 13th ACM-SIAM Symposium on Discrete Algorithms*, pages 616–620, 2002.

[Chl07] Eden Chlamtac. Approximation algorithms using hierarchies of semidefinite programming relaxations. In *FOCS*, pages 691–701, 2007.

[CKK+04] Shuchi Chawla, Robert Krauthgamer, Ravi Kumar, Yuval Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. Manuscript, 2004.

[CMM06a] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for unique games. In *Proceedings of the 38th ACM Symposium on Theory of Computing*, 2006.

[CMM06b] Eden Chlamtac, Konstantin Makarychev, and Yury Makarychev. How to play unique games using embeddings. In *Proceedings of the 47th IEEE Symposium on Foundations of Computer Science*, 2006.

[CMM07a] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Local global tradeoffs in metric embeddings. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*, 2007.

[CMM07b] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for maximum constraint satisfaction problems. In *Proceedings of the 18th ACM-SIAM Symposium on Discrete Algorithms*, pages 62–68, 2007.

[CMM09] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality gaps for sherali-adams relaxations. In *Proceedings of the 41st ACM Symposium on Theory of Computing*, 2009.

[dlVKM07] Wenceslas Fernandez de la Vega and Claire Kenyon-Mathieu. Linear programming relaxations of maxcut. In *SODA*, pages 53–61, 2007.

[DMR06] Irit Dinur, Elchanan Mossel, and Oded Regev. Conditional hardness for approximate coloring. In *Proceedings of the 38th ACM Symposium on Theory of Computing*, 2006.

[DS05] Irit Dinur and Shmuel Safra. On the hardness of approximating minimum vertex-cover. *Annals of Mathematics*, 162(1):439–486, 2005.

[EH05] Lars Engebretsen and Jonas Holmerin. More efficient queries in PCPs for NP and improved approximation hardness of maximum CSP. In *Proceedings of the 22th Symposium on Theoretical Aspects of Computer Science*, pages 194–205, 2005.

[Fei97] Uriel Feige. Randomized graph products, chromatic numbers, and the lovász vartheta-funktion. *Combinatorica*, 17(1):79–90, 1997.

[FGL+91] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Approximating clique is almost NP-complete. In *Proceedings of the 32nd IEEE Symposium on Foundations of Computer Science*, pages 2–12, 1991.

[FK98] Uriel Feige and Joe Kilian. Zero knowledge and the chromatic number. *Journal of Computer and System Sciences*, 57(2):187–199, 1998.

[FL92] Uri Feige and László Lovász. Two-prover one round proof systems: Their power and their problems. In *Proceedings of the 24th ACM Symposium on Theory of Computing*, pages 733–741, 1992.

[FLS04] Uriel Feige, Michael Langberg, and Gideon Schechtman. Graphs with tiny vector chromatic numbers and huge chromatic numbers. *SIAM J. Comput.*, 33(6):1338–1368, 2004.

[FO06] Uriel Feige and Eran Ofek. Random 3CNF formulas elude the Lovász theta function. Manuscript, 2006.

[FR87] Peter Frankl and Vojtech Rodl. Forbidden intersections. *Transactions of the American Mathematical Society*, 300(1):259–286, 1987.

[GMPT07] Konstantinos Georgiou, Avner Magen, Toniann Pitassi, and Iannis Tourlakis. Integrality gaps of 2 - o(1) for vertex cover SDPs in the Lovász-Schrijver hierarchy. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*, pages 702–712, 2007.

[GMT09] Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. Optimal sherali-adams gaps from pairwise independence. In *APPROX-RANDOM*, 2009.

[GR08] Venkatesan Guruswami and Prasad Raghavendra. Constraint satisfaction over a non-boolean domain: Approximation algorithms and unique-games hardness. In *APPROX-RANDOM*, pages 77–90, 2008.

[GT06] Anupam Gupta and Kunal Talwar. Approximating unique games. In *Proceedings of the 17th ACM-SIAM Symposium on Discrete Algorithms*, pages 99–106, 2006.

[Has05] Gustav Hast. Approximating Max kCSP - outperforming a random assignment with almost a linear factor. In *Proceedings of the 32nd International Colloquium on Automata, Languages and Programming*, pages 956–968, 2005.

[Hås07] Johan Håstad. On the approximation resistance of a random predicate. In *APPROX-RANDOM*, pages 149–163, 2007.

[Hol07] Thomas Holenstein. Parallel repetition: simplifications and the no-signaling case. In *Proceedings of the 39th ACM Symposium on Theory of Computing*, pages 411–419, 2007.

[KG98] Jon M. Kleinberg and Michel X. Goemans. The Lovász Theta function and a semidefinite programming relaxation of vertex cover. *SIAM Journal on Discrete Mathematics*, 11:196–204, 1998.

[Kho01] Subhash Khot. Improved inaproximability results for maxclique, chromatic number and approximate graph coloring. In *FOCS*, pages 600–609, 2001.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th ACM Symposium on Theory of Computing*, pages 767–775, 2002.

[KKMO04] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for MAX-CUT and other two-variable CSPs? In *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science*, pages 146–154, 2004.

[KP06] Subhash Khot and Ashok Kumar Ponnuswami. Better inapproximability results for maxclique, chromatic number and min-3lin-deletion. In *ICALP (1)*, pages 226–237, 2006.

[KR03] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within $2-\varepsilon$. In *Proceedings of the 18th IEEE Conference on Computational Complexity*, 2003.

[KS09] Subhash Khot and Rishi Saket. Sdp integrality gaps with local l1-embeddability. In *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science*, 2009.

[KV05] Subhash Khot and Nisheeth Vishnoi. The unique games conjecture, integrality gap for cut problems and the embeddability of negative type metrics into ℓ_1. In *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science*, pages 53–63, 2005.

[Las01] Jean B. Lasserre. An explicit exact sdp relaxation for nonlinear 0-1 programs. In *Proceedings of the 8th Conference on Integer Programming and Combinatorial Optimization*, pages 293–303, London, UK, 2001. Springer-Verlag.

[Lau03] Monique Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming. *Math. Oper. Res.*, 28(3):470–496, 2003.

[Lov79] L. Lovasz. On the shannon capacity of a graph. *Information Theory, IEEE Transactions on*, 25(1):1–7, Jan 1979.

[LS91] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM J. on Optimization*, 1(12):166–190, 1991.

[Rao08] Anup Rao. Parallel repetition in projection games and a concentration bound. In *Proceedings of the 40th ACM Symposium on Theory of Computing*, pages 1–10, 2008.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998. Preliminary version in *Proc. of STOC'95*.

[RS09] Prasad Raghavendra and David Steurer. Integrality gaps for strong sdp relaxations of unique games. In *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science*, 2009.

[SA90] Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discrete Math.*, 3(3):411–430, 1990.

[Sch08] Grant Schoenebeck. Linear level Lasserre lower bounds for certain k-csps. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, 2008.

[SS07] Shmuel Safra and Oded Schwartz. On parallel-repetition, unique-game and max-cut. Manuscript, 2007.

[ST00] Alex Samorodnitsky and Luca Trevisan. A PCP characterization of NP with optimal amortized query complexity. In *Proceedings of the 32nd ACM Symposium on Theory of Computing*, pages 191–199, 2000.

[ST06] Alex Samorodnitsky and Luca Trevisan. Gowers uniformity, influence of variables, and PCPs. In *Proceedings of the 38th ACM Symposium on Theory of Computing*, pages 11–20, 2006.

[STT07a] Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. A linear round lower bound for Lovász-Schrijver SDP relaxations of vertex cover. In *Proceedings of the 22nd IEEE Conference on Computational Complexity*, 2007.

[STT07b] Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. Tight integrality gaps for Lovasz-Schrijver LP relaxations of vertex cover and max cut. In *Proceedings of the 39th ACM Symposium on Theory of Computing*, 2007.

[Tou06] Iannis Tourlakis. New lower bounds for vertex cover in the Lovasz-Schrijver hierarchy. In *Proceedings of the 21st IEEE Conference on Computational Complexity*, 2006.

[Tre98] Luca Trevisan. Parallel approximation algorithms by positive linear programming. *Algorithmica*, 21(1):72–88, 1998. Preliminary version in *Proc. of ESA'96*.

[Tre05] Luca Trevisan. Approximation algorithms for unique games. In *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science*, 2005.

[Tul09] Madhur Tulsiani. CSP gaps and reductions in the Lasserre hierarchy. In *Proceedings of the 41st ACM Symposium on Theory of Computing*, 2009.

Appendix A

Deferred Proofs

A.1 Random instances of MAX k-CSP$_q$

The following lemma easily implies lemmata 2.4, 4.2 and 5.8.

Lemma A.1 *Let $\varepsilon, \delta > 0$ and a predicate $P : [q]^k \to \{0,1\}$ be given. Then there exist $\beta = O(q^k/\varepsilon^2)$, $\eta = \Omega((1/\beta)^{5/\delta})$ and $N \in \mathbb{N}$, such that if $n \geq N$ and Φ is a random instance of* MAX k-CSP *(P) with $m = \beta n$ constraints, then with probability* $1 - o(1)$

1. $\mathsf{OPT}(\Phi) \leq \frac{|P^{-1}(1)|}{q^k}(1+\varepsilon) \cdot m$.
2. *For all $s \leq \eta n$, every set of s constraints involves at least $(k-1-\delta)s$ variables.*

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

For showing the next property, we consider the probability that a set of s constraints contains at most cs variables, where $c = k-1-\delta$. This is upper bounded by

$$\binom{n}{cs}\cdot\binom{\binom{cs}{k}}{s}\cdot s!\binom{\beta n}{s}\cdot\binom{n}{k}^{-s}$$

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

$$\left(\frac{s}{n}\right)^{\delta s}\left(e^{2k+1-\delta}k^{1+\delta}\beta\right)^{s} \leq \left(\frac{s}{n}\right)^{\delta s}(\beta^5)^s = \left(\frac{s\beta^{5/\delta}}{n}\right)^{\delta s}$$

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

$$\begin{aligned}\sum_{s=1}^{\eta n}\left(\frac{s\beta^{5/\delta}}{n}\right)^{\delta s} &= \sum_{s=1}^{\ln^2 n}\left(\frac{s\beta^{5/\delta}}{n}\right)^{\delta s} + \sum_{s=\ln^2 n+1}^{\eta n}\left(\frac{s\beta^{5/\delta}}{n}\right)^{\delta s} \\ &\leq O\left(\frac{\beta^5}{n^\delta}\ln^2 n\right) + O\left(\left(\eta\cdot\beta^{5/\delta}\right)^{\delta\ln^2 n}\right)\end{aligned}$$

The first term is $o(1)$ and is small for large n. The second term is also $o(1)$ for $\eta = 1/(100\beta^{5/\delta})$. ■

Proof of Lemma 2.4: If we take P to 3-XOR, the expansion and unsatisfiability of a random instance follow directly from lemma A.1. We only need to ensure that at the same time no two clauses (constraints) share more than one variable.

The probability that there are no two clauses sharing two variables must be at least $\prod_{s=1,\dots,m}(1 - O(s)/n^2)$ because when we choose the sth clause, by wanting it to not share two variables with another previously chosen clause, we are forbidding $O(s)$ pairs of variables to occur together. Each such pair happens to be in the clause with probability $O(1/n^2)$. Now we use that for small enough x, $1 - x > \exp(-O(x))$ the probability is at least $\exp(-O((\sum_{s=1,\dots,m} s)/n^2)) = \exp(-O(m^2/n^2)) = \exp(-O(\beta^2))$ which is some positive constant. ■

Proof of Lemma 5.8: We only need to express the expansion condition in terms of $|\partial(C)|$ for a set C of constraints. We claim that if C involves at least $(k-1-\delta)|C|$ variables, then $|\partial(C)| \geq (k-2-2\delta)|C|$. If we use $\Gamma(C)$ to denote all the variables contained in C, then it is easy to see that

$$k|C| \geq |\partial(C)| + 2(|\Gamma(C)| - |\partial(C)|)$$

since every constraint in C has exactly k variables and each variable in $\Gamma(C) \setminus \partial(C)$ appears in at least two constraints. Rearranging the terms proves the claim. ■

A.2 Proofs from Chapter 3

A.2.1 Proof of Lemma 3.1

Lemma A.2 *For every* $0 < \alpha < 1/125$, $\eta > 0$, *there exists a* $d = d(\alpha) \in \mathbb{N}$, $\delta, \gamma > 0$, *and* $N \in \mathbb{N}$ *such that for* $n \geq N$ *there exists an* $(\alpha, \delta, \gamma, \eta)$ *graph with max cut less than* $\frac{1}{2}|E|(1+\alpha)$ *and maximum degree at most* d *on* n *vertices. Here* $d(\alpha)$ *is an explicit function that depends only on* α.

We use the following lemma from [ABLT06]

Lemma A.3 *For every $1 < \alpha < 1/250$, $\eta > 0$, there exists a $\delta, \gamma > 0$ such that a random graph from the $G_{n,p}$ distribution where $p = \alpha^{-2}/n$ has the following properties with probability $1 - o(n)$:*

- *after $O(\sqrt{n})$ edges are removed, the girth is $\delta \log n$.*
- *the minimum vertex cover contains at least $(1 - \alpha)n$ vertices*
- *every induced subgraph on a subset S of at most γn vertices has at most $(1 + \eta)|S|$ edges.*

Proof: [of Lemma A.2] Given $\alpha, \eta > 0$, set $\alpha' = \alpha/2$. Use Lemma A.3 with inputs α', η to randomly pick a graph on n vertices. Set $p = (\alpha')^{-2}/n$ as in Lemma A.3. Now, with high probability, we can remove set of edges R to obtain a $(\alpha/2, \delta, \gamma, \eta)$-graph on n vertices. Do not yet remove edges.

Also, it is well known that w.h.p. the max-cut in a random $G_{n,p}$ has size less than $\frac{1}{2}|E|(1+1/\sqrt{d})$, where d is the average degree. The average degree of a vertex in this model is $\lambda = pn = 4\alpha^{-2}$. Hence the size of the max-cut is at most $\frac{1}{2}|E|(1 + \alpha/2)$. The probability that some fixed vertex v_0 has degree greater than 2λ is less than $\exp(-\lambda/3)$ by a Chernoff bound. So by Markov's inequality the probability that more than $\exp(-\lambda/6)n$ vertices have degree greater than 2λ is at most $\exp(-\lambda/6) \leq exp(-10000)$.

If this is the case, then first remove the edge set R. By removing edges we could only decrease the maximum degree. Then simply remove all vertices with degree more than 2λ from the graph and any other subset to obtain a graph G' with $n(1 - exp(-d/6))$ vertices. Now, it is easy to check that G' is a $(\alpha, \delta, \gamma, \eta)$-graph with maximum degree at most $d(\alpha) = 2\lambda = 8/\alpha^2$. Removing the edges and vertices changes the max cut to $\frac{1}{2}|E|(1 + \alpha/2 + o(1)) < \frac{1}{2}|E|(1 + \alpha)$. ■

A.2.2 Proofs of claims about splashes

We use the following notation for the proofs in this appendix. We denote $\mathbb{P}[i = 1|r = b]$ and $\mathbb{P}[i = 1, j = 1|r = b]$ by $P_r^b(i)$ and $P_r^b(i, j)$ respectively. $\mathbb{P}[i = 0|r = b]$ and $\mathbb{P}[i = 0, j = 1|r = b]$ are expressed as $1 - P_r^b(i)$ and $P_r^b(j) - P_r^b(i, j)$ respectively. Also, in cases where $\mathbb{P}[j = 1|i = b]$ depends only on $d(i, j)$, we denote it by $Q^b(d(i, j))$.

Claim A.4 *Consider a b-Splash around a vertex i such that all vertices upto distance ℓ are labeled $\frac{1}{2} + \varepsilon$. Then,*

1. $Q^1(k) = (1/2 + \varepsilon)\left[1 + (-1)^k \left(\frac{1/2-\varepsilon}{1/2+\varepsilon}\right)^{k+1}\right]$ *for* $0 \leq k \leq \ell$

   $Q^0(0) = 0$ *and* $Q^0(k) = Q^1(k-1)$ *for* $1 \leq k \leq \ell$
2. $\left|Q^0(\ell/2) - (1/2 + \varepsilon)\right| \leq \varepsilon^4$
3. $\forall 0 \leq k \leq \ell$, $Q^1(k) + Q^1(k+1) \geq 1 + 4\varepsilon^2$

Proof: We prove the formula for $Q^1(k)$ by induction. For $k = 0$,

$$(1/2 + \varepsilon)\left[1 + (-1)^k \left(\frac{1/2 - \varepsilon}{1/2 + \varepsilon}\right)^{k+1}\right] = (1/2 + \varepsilon)\left[\frac{1}{1/2 + \varepsilon}\right] = 1 = Q^1(0)$$

Assuming the correctness of the formula for $k = n$, we start with the recurrence

$$Q^1(n+1) = (1 - Q^1(n)) + \left(\frac{2\varepsilon}{1/2+\varepsilon}\right) Q^1(n) = 1 - \left(\frac{1/2-\varepsilon}{1/2+\varepsilon}\right) Q^1(n)$$

since the vertex at distance n (in the same path) might not be present with probability $1 - Q^1(n)$ in which case the one at distance $n+1$ is present with probability 1, and it is present with probability $Q^1(n)$ in which case the one at distance $n+1$ is included with probability $\left(\frac{2\varepsilon}{1/2+\varepsilon}\right)$. Therefore, we have

$$\begin{aligned} Q^1(n+1) &= 1 - \left(\frac{1/2-\varepsilon}{1/2+\varepsilon}\right)(1/2+\varepsilon)\left[1 + (-1)^n \left(\frac{1/2-\varepsilon}{1/2+\varepsilon}\right)^{n+1}\right] \\ &= 1 - (1/2-\varepsilon) + (-1)^{n+1}(1/2+\varepsilon)\left(\frac{1/2-\varepsilon}{1/2+\varepsilon}\right)^{n+2} = (1/2+\varepsilon)\left[1 + (-1)^{n+1}\left(\frac{1/2-\varepsilon}{1/2+\varepsilon}\right)^{n+2}\right] \end{aligned}$$

Also note that if i is labeled 0, then all its neighbors must be set to 1. Hence $Q^0(0) = 0$ and $Q^0(1) = 1$. The rest of the induction works exactly as above.

Note that

$$\left|Q^0(\ell/2) - (1/2+\varepsilon)\right| = (1/2+\varepsilon)\left(\frac{1/2-\varepsilon}{1/2+\varepsilon}\right)^{\ell/2} < (1-2\varepsilon)^{\ell/2} = (1-2\varepsilon)^{(\frac{4}{\varepsilon}\log\frac{1}{\varepsilon})} \le \varepsilon^4$$

Finally for $0 \le k < \ell$,

$$\begin{aligned} Q^1(k) + Q^1(k+1) &= (1/2+\varepsilon)\left[2 + (-1)^k\left(\frac{1/2-\varepsilon}{1/2+\varepsilon}\right)^{k+1}\left(1 - \frac{1/2-\varepsilon}{1/2+\varepsilon}\right)\right] \\ &= (1/2+\varepsilon)\left[2 + (-1)^k\left(\frac{1/2-\varepsilon}{1/2+\varepsilon}\right)^{k+1}\left(\frac{2\varepsilon}{1/2+\varepsilon}\right)\right] \\ &\ge (1/2+\varepsilon)\left[2 - \left(\frac{2\varepsilon}{1/2+\varepsilon}\right)\left(\frac{1/2-\varepsilon}{1/2+\varepsilon}\right)^2\right] = 1 + 2\varepsilon - 2\varepsilon\left(\frac{1/2-\varepsilon}{1/2+\varepsilon}\right)^2 \ge 1 + 4\varepsilon^2 \end{aligned}$$

The claim for $k = \ell$ follows from part 2 and the fact that $Q^1(d) = 1/2 + \varepsilon$ for $d > \ell$. ■

Claim A.5 *If we pick a 0-Splash with probability* $1/2 - \varepsilon$ *and a 1-Splash with probability* $1/2 + \varepsilon$*, then all vertices have probability* $1/2 + \varepsilon$*. Furthermore, vertices at distance* $\ell + 1$ *or more from i have weight* $1/2 + \varepsilon$ *in the 0-Splash as well as 1-Splash around i.*

Proof: We prove it by induction on the length of the path from i to j. Let $P_i(j) = (1/2-\varepsilon)P_i^0(j) + (1/2+\varepsilon)P_i^1(j)$. The base case, when the path is of length 0 is clear. If the path between i and j is $i = v_0, v_1, \ldots, v_{m-1}, v_m = j$, then there are two cases. In the first case v_{m-1} and v_m are both within distance ℓ of i. Then

$$P_i(j) = 1 - (1 - \frac{2\varepsilon}{1/2+\varepsilon})P_i(v_{m-1})$$

because v_m is only excluded with probability $\frac{2\varepsilon}{1/2+\varepsilon}$ when v_{m-1} is present and this event is independent of whether or not each vertex $i = v_0, v_1, \ldots, v_{m-1}$ is included in the cover. By induction, $P_i(v_{m-1}) = 1/2 + \varepsilon$, and so $1 - (1 - \frac{2\varepsilon}{1/2+\varepsilon})P_i(v_{m-1}) = 1/2 + \varepsilon$.

In the second case v_{m-1} is at distance ℓ. However,

$$P_i^b(j) = 1 - (1 - \frac{P_i^b(v_{m-1}) - (1/2 - \varepsilon)}{P_i^b(v_{m-1})})P_i^b(v_{m-1}) = 1/2 + \varepsilon$$

because the probability v_{m-1} is included in a b-Splash is $P_i^b(v_{m-1})$ and the probability of including v_m when v_{m-1} is present is $\frac{P_i^b(v_{m-1})-(1/2-\varepsilon)}{P_i^b(v_{m-1})}$. ■

Claim A.6 *Let $i = v_0, v_1, \ldots, v_{m-1}, v_m = j$ be the path to j, $m \leq \ell$, and let u be the vertex on this path which is closest to r. Then*

1. $P_r^b(i, j) = P_r^b(u) \cdot P_u^1(i)P_u^1(j) + [1 - P_r^b(u)] \cdot P_u^0(i)P_u^0(j)$
2. *If $P_r^b(u) = 1/2 + \varepsilon$, then $P_r^b(i, j) = (1/2 + \varepsilon)P_i^1(j)$*

Proof:

1. Let E be the event that both i and j are in a vertex cover and $r = b$. Then $P_r^b(i, j) = \mathbb{P}[E \mid r = b]$. We can also condition on whether u is in the vertex cover.

$$\begin{aligned} P_r^b(i, j) &= \mathbb{P}[u \in VC \mid r = b] \cdot \mathbb{P}[E \mid r = b \text{ and } u \in VC] \\ &+ \mathbb{P}[u \notin VC \mid r = b] \cdot \mathbb{P}[E \mid r = b \text{ and } u \notin VC] \end{aligned}$$

But $\mathbb{P}[E \mid r = b \text{ and } u \in VC] = \mathbb{P}[E \mid u \in VC]$. Because given that u is in or out of the vertex cover, we can determine if i and j are in the vertex cover by following the edges from u to each of them. But this information is independent of whether r is in the vertex cover. For the same reason $\mathbb{P}[E \mid r = b \text{ and } u \in VC] = \mathbb{P}[E \mid u \in VC]$. Therefore

$$P_r^b(i, j) = P_r^b(u) \cdot P_u^1(i)P_u^1(j) + [1 - P_r^b(u)]\dot{P}_u^0(i)P_u^0(j)$$

as claimed.

2. The probability that i and j are in a vertex cover (assume r is not yet fixed) is just $(1/2 + \varepsilon)P_i^1(j)$. Now, we can just condition on l, and rewrite this as

$$\mathbb{P}[u \in VC] \cdot P_u^1(i, j) + \mathbb{P}[u \notin VC] \cdot P_u^0(i, j)$$

We can also not condition on $r = b$ because once l is fixed, that does not affect anything, and in addition, $\mathbb{P}[u \in VC] = 1/2 + \varepsilon = P_r^b(u)$. So this becomes

$$P_r^b(u) \cdot P_u^1(i, j) + [1 - P_r^b(u)] \cdot P_u^0(i, j)$$

Finally, if we note that $P_u^b(i, j) = P_u^b(i)P_u^b(j)$, we see that we get

$$P_r^b(l) \cdot P_u^1(i)P_u^1(j) + [1 - P_r^b(u)] \cdot P_u^0(i)P_u^0(j)$$

which by 1) is simply $P_r^b(i, j)$ as claimed.

■

Claim A.7 *Let i be a vertex and (j,k) be an edge in a b-Splash around r. Then if j and k are not already fixed*

$$P_r^b(i,j) + P_r^b(i,k) \geq P_r^b(i)(1 + 4\varepsilon^3)$$

and

$$[P_r^b(j) - P_r^b(i,j)] + [P_r^b(k) - P_r^b(i,k)] \geq (1 - P_r^b(i))(1 + 4\varepsilon^3)$$

Proof: We consider separately the cases when (j,k) lies on or outside the path between r and i.

***Case 1*: (j,k) lies outside the path connecting r and i**

Without loss of generality, let j be the vertex closer to the path from r to i. Let u be the vertex in the path closest to j. Then by Claim A.6

$$P_r^b(i,j) = P_r^b(u) \cdot P_u^1(i)P_u^1(j) + [1 - P_r^b(u)] \cdot P_u^0(i)P_u^0(j)$$
$$P_r^b(i,k) = P_r^b(u) \cdot P_u^1(i)P_u^1(k) + [1 - P_r^b(u)] \cdot P_u^0(i)P_u^0(k)$$

Therefore,

$$P_r^b(i,j) + P_r^b(i,k) = P_r^b(u)P_l^1(i) \cdot \left[P_u^1(j) + P_u^1(k)\right] + [1 - P_r^b(u)]P_u^0(i) \cdot \left[P_u^0(j) + P_u^0(k)\right]$$

Also by Claim A.4 we know that $P_u^b(j) + P_u^b(k) \geq 1 + 4\varepsilon^2$, if j and k are not already fixed, which gives

$$P_r^b(i,j) + P_r^b(i,k) \geq \left[P_r^b(u)P_u^1(i) + [1 - P_r^b(u)]P_u^0(i)\right](1 + 4\varepsilon^2) = P_r^b(u)(1 + 4\varepsilon^2)$$

***Case 2*: (j,k) lies on the path connecting r and i**

Let j be the vertex closer to r. Also, let $\alpha = P_r^b(j)$ and $\beta = P_j^1(i)$. Then,

$$\begin{aligned}
P_r^b(i,j) &= P_r^b(j)P_j^1(i) = \alpha\beta \\
P_r^b(i,k) &= P_r^b(k)P_k^1(i) = \left[1 - \alpha + \frac{2\varepsilon}{1/2 + \varepsilon}\alpha\right]\left[(1-\beta)\frac{1/2 + \varepsilon}{1/2 - \varepsilon}\right] \\
&= (1-\alpha)(1-\beta)\left(\frac{1/2 + \varepsilon}{1/2 - \varepsilon}\right) + \alpha(1-\beta)\left(\frac{2\varepsilon}{1/2 - \varepsilon}\right)
\end{aligned}$$

where the second equation follows from the recurrence $Q^1(n+1) = (1 - Q^1(n)) + \left(\frac{2\varepsilon}{1/2+\varepsilon}\right) Q^1(n)$ used in Claim A.4. Also,

$$\begin{aligned}
P_r^b(i) &= P_r^b(j)P_j^1(i) + (1 - P_r^b(j))P_j^0(i) = P_r^b(j)P_j^1(i) + (1 - P_r^b(j))P_k^1(i) \\
&= \alpha\beta + (1-\alpha)(1-\beta)\left(\frac{1/2 + \varepsilon}{1/2 - \varepsilon}\right)
\end{aligned}$$

This gives

$$\frac{P_r^b(i,j) + P_r^b(i,j)}{P_r^b(i)} = 1 + \frac{\alpha(1-\beta)\left(\frac{2\varepsilon}{1/2-\varepsilon}\right)}{\alpha\beta + (1-\alpha)(1-\beta)\left(\frac{1/2+\varepsilon}{1/2-\varepsilon}\right)} \geq 1 + 4\varepsilon^3$$

since $\alpha, (1-\beta) > 2\varepsilon$ (all probabilities in a splash are at least 2ε, unless one is 0 and the other is 1, but then both are fixed).

The proof of the second statement follows similarly. ■

Claim A.8 *Let i and j be two vertices in a b-Splash around r, such that $d(i,j) \geq \ell$. Then*

$$\left|P_r^b(i,j) - P_r^b(i)P_r^b(j)\right| \leq 2\varepsilon^4$$

and

$$\left|[P_r^b(j) - P_r^b(i,j)] - (1 - P_r^b(i))P_r^b(j)\right| \leq 2\varepsilon^4$$

Proof: Let u be the vertex closest to r on the path from i to j. Without loss of generality, assume that $d(i,u) \geq \ell/2$. Then

$$\begin{aligned}\left|P_r^b(i,j) - P_r^b(i)P_r^b(j)\right| &= \left|P_r^b(u)\cdot P_u^1(i)P_u^1(j) + [1 - P_r^b(u)]\cdot P_u^0(i)P_u^0(j) - P_r^b(i)P_r^b(j)\right| \\ &\leq \left|(1/2+\varepsilon)\left[P_r^b(u)\cdot P_u^1(j) + [1-P_r^b(u)]\cdot P_u^0(j)\right] - P_r^b(i)P_r^b(j)\right| + \varepsilon^4 \\ &= \left|(1/2+\varepsilon)P_r^b(j) - P_r^b(i)P_r^b(j)\right| + \varepsilon^4 \leq 2\varepsilon^4\end{aligned}$$

where the two inequalities follow from the fact that $|P_r^b(i) - (1/2+\varepsilon)| \leq \varepsilon^4$ if $d(i,r) \geq \ell/2$ as proved in Claim A.4.

The second statement can be proven in a similar fashion. ■