Approximation Algorithms for Constraint Satisfaction Problems

Yury Makarychev

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
By the Department of
Computer Science

April 2008

Abstract

Constraint satisfaction problems (CSP) are very basic and natural combinatorial optimization problems. Given a set of variables and constraints on them, our goal is to satisfy as many constraints as possible.

In this dissertation, we study two constraint satisfaction problems, the Unique Games Problem and MAX 2CSP Problem. Both problems behave very differently depending on what fraction of all constraints (as a function of other parameters) is satisfiable. Thus we design different algorithms for different ranges of parameters.

In the first part of the thesis, we study approximation algorithms for Unique Games. Our algorithms satisfy (roughly)

$$k^{-\varepsilon/(2-\varepsilon)}, \qquad 1 - O(\sqrt{\varepsilon \log n}), \qquad \text{and} \qquad 1 - O(\varepsilon\sqrt{\log n \log k})$$

fraction of all constraints if a $1 - \varepsilon$ fraction of all constraints is satisfiable (where n is the number of variables, k is the domain size). The first two algorithms are near optimal assuming the Unique Games Conjecture. Our algorithms have better approximation guarantees than the previously best known algorithms in all ranges of parameters.

In the second part of the thesis, we present approximation algorithms for MAX 2CSP that satisfy $1-O(\sqrt{\varepsilon})$ and $1-O(\varepsilon\sqrt{\log n})$ fraction of all constraints if a $1-\varepsilon$ fraction of all constraints is satisfiable. Our first algorithm is near optimal assuming the Unique Games Conjecture.

Acknowledgments

First and foremost, I want to thank my advisor Moses Charikar for his support and guidance during my PhD study. I am very grateful to Moses for everything I learned from him. This dissertation would not have been possible without his help.

I would like to thank the members of my PhD committee, Boaz Barak, Bernard Chazelle, Assaf Naor and Robert Schapire for their time and for their helpful comments and suggestions.

I thank my coauthors Amit Agarwal, Noga Alon, Eden Chlamtac, Konstantin Makarychev, Assaf Naor and, of course, Moses Charikar. I benefited greatly from working with them.

I thank faculty and staff of the Department of Computer Science at Princeton University. I thank Melissa Lawson and Mitra Kelly for their administrative help.

I am grateful to my teachers at Moscow Math High School 57: Lev D. Altshuler, Boris M. Davidovich, Sergey Dorichenko, Arkadiy Skopenkov, Yevgeniy A. Vyrodov, and Ivan V. Yashchenko. I thank my undergraduate advisors at Moscow University, Alexander Shen and Nikolai K. Vereshchagin.

I was fortunate to spend two wonderful summers at Microsoft Research. I thank the members of the Theory Group and especially my mentor Assaf Naor.

I gratefully acknowledge a Gordon Wu Fellowship and Moses Charikar's grants CCR-0205594, NSF CAREER award CCR-0237113, MSPA-MCS award 0528414 that supported my research at Princeton.

My special thanks are due to my brother and best friend Konstantin. Most of my research has been done together with him. Finally, I would like to thank my parents, Marina and Sergey, for their love, support and encouragement. This dissertation is dedicated to them.

To my parents, Marina and Sergey.

Contents

Abstract . . . iii

List of Figures . . . viii

1 Introduction **1**

1.1 Unique Games . . . 4

1.1.1 Our Results . . . 6

1.2 MAX 2CSP . . . 8

2 Unique Games **11**

2.1 Overview . . . 11

2.1.1 Semidefinite Programming . . . 11

2.1.2 Techniques . . . 14

2.2 Semidefinite Relaxation . . . 16

2.3 First Algorithm . . . 19

2.3.1 Algorithm . . . 19

2.3.2 Analysis: Technical Details . . . 26

2.4 Second Algorithm . . . 31

2.4.1 Algorithm . . . 31

2.4.2 Analysis: Technical Details . . . 37

2.5 Third Algorithm 50
2.5.1 Overview: Orthogonal Separators 50
2.5.2 Approximation Algorithm 51
2.6 Producing Orthogonal Separators 55
2.6.1 Normalization: Embedding into $L_2[0,\infty]$ 56
2.6.2 Embedding into ℓ_1 and ℓ_2 58
2.6.3 Generating Orthogonal Separators via ℓ_1 62
2.6.4 Generating Orthogonal Separators via ℓ_2 65
2.7 d to 1 Games 70

3 MAX 2 CSP **75**

3.1 Semidefinite Relaxation 75
3.2 First Algorithm 77
3.3 Second Algorithm 82
3.3.1 Preliminaries 82
3.3.2 Separation Theorem 84
3.3.3 Algorithm and Analysis 88

4 Conclusion and Future Work **92**

4.1 Conclusions 92
4.2 Future Work 93

A Properties of Normal Distribution **99**

A.1 Bounds on Normal Distribution Function 99
A.2 Šidák Theorem 101

List of Figures

1.1 Summary of Results . . . 3

1.2 Previously Known Results for MAX 2CSP . . . 9

2.1 Comparison of Approximation Guarantees for Unique Games . . . 15

2.2 First Algorithm for Unique Games . . . 21

2.3 Second Algorithm for Unique Games . . . 32

2.4 Third Algorithm for Unique Games . . . 52

2.5 Generating Orthogonal Separators via ℓ_1 . . . 64

2.6 Generating Orthogonal Separators via ℓ_2 . . . 66

3.1 First Algorithm for MAX 2SAT . . . 78

3.2 Finding $(S, R, -S)$ Partitioning . . . 88

3.3 Second Algorithm for MAX 2SAT . . . 90

Chapter 1

Introduction

In this dissertation, we study approximation algorithms for constraint satisfaction problems. In a constraint satisfaction problem we have a set of variables and constraints on them, and our goal is to find an assignment that satisfies as many constraints as possible.

Our interest in constraint satisfaction problems is two-fold. First of all, CSPs are good model problems: they are well suited for developing new algorithmic techniques, which can potentially be applied to other problems of interest. Another reason to study CSPs is their importance to complexity theory. Many problems arising in complexity theory such as 3SAT, Unique Games, and Label Cover (which is closely related to PCPs) are constraint satisfaction problems. Despite significant progress, many important problems are still open.

The general problem — when we consider arbitrary constraints that depend on at most k boolean variables — is very hard: the optimal approximation algorithm gives only $\Theta(k/2^k)$ approximation. (The upper bound was proved assuming the Unique Games Conjecture by Samorodnitsky and Trevisan [29], and the lower bound was proved by Charikar, Makarychev, and Makarychev [7].) In this dissertation, we focus our attention on two constraint satisfaction problems for which much better approximation guarantees exist: Unique Games and MAX 2SAT. Both problems generalize MAX CUT. Recall that MAX CUT asks to partition vertices of a given graph into two parts so as to maximize the number of cut edges, i.e. edges going from one to part to the other. The problem can be also phrased as a CSP.

Definition 1.0.1 (MAX CUT as a CSP)**.** *We are given a graph $G = (V, E)$. For every vertex $u \in V$, we have a variable $x_u \in \{0, 1\}$; for every edge $(u, v) \in E$, we have the constraint $x_u + x_v = 1$. Our goal is to find an assignment that maximizes the number of satisfied constraints.*

Note that every assignment of variables corresponds to a partitioning of vertices: $V = \{u : x_u = 0\} \cup \{u : x_u = 1\}$ *(and vice versa).*

MAX CUT has two properties that distinguish it from other CSPs:

Binary Domain. The domain size is two (every variable takes two values).

Uniqueness Property. Every constraint (edge) defines a permutation between two variables. That is, for every edge (u, v) the value of x_u *uniquely determines* the value of x_v (assuming that the constraint is satisfied).

Each of the problems we study has only one of these two properties: in MAX 2SAT, the domain size is two but constraints can be arbitrary predicates of two variables; in Unique Games, the domain size can be arbitrary, but all the constraints have the uniqueness property.

Range of Parameters	**Unique Games**	**MAX 2CSP**
$\varepsilon \geq \Omega(1/\log k)$	$\sim k^{-\frac{\varepsilon}{2-\varepsilon}}$ $\star$	
$\Omega(1/\log n) \leq \varepsilon \leq O(1/\log k)$	$1 - O(\sqrt{\varepsilon \log k})$ $\star$	$1 - O(\sqrt{\varepsilon})$ $\star$
$\varepsilon \leq O(1/\log n)$	$1 - O(\varepsilon\sqrt{\log n \log k})$	$1 - O(\varepsilon\sqrt{\log n})$

Figure 1.1: Summary of results. The guarantee represents the fraction of constraints satisfied for instances where $OPT = 1 - \varepsilon$, n is the number of variables, k is the domain size of the unique game. The results marked with $\star$ are near optimal assuming the Unique Games Conjecture.

Our results are presented in Figure 1.1. Remarkably the approximation guarantees for the problems are very similar to each other and the approximation guarantees for MAX 2CSP match the best known approximation guarantees for MAX CUT up to a constant factor in the O notation. (The approximation algorithm for MAX CUT with the guarantee $1-O(\sqrt{\varepsilon})$ was developed by Goemans and Williamson [17]; the approximation $1 - O(\varepsilon\sqrt{\log n})$ for MAX CUT follows from our general result for MAX 2CSP.)

Note that if almost all constraints are satisfiable, it is also possible to satisfy almost all constraints in polynomial time. However, Unique Games and MAX 2CSP are apparently the only generalizations of MAX CUT with this property. For example, even if the domain size is 3 and every constraint is of the form $x_i \neq x_j$ then satisfying all the constraints in a satisfiable instance is NP-hard [31] (this CSP is just another formulation of the Graph 3-Coloring Problem).

The results presented in this dissertation appeared in the following publications:

- A. Agarwal, M. Charikar, K. Makarychev, and Y. Makarychev. $O(\sqrt{\log n})$ *approximation algorithms for MIN UnCut, MIN 2CNF Deletion, and directed cut problems.* In Proceedings of the 37th ACM Symposium on Theory of Computing, pp. 573–581, 2005.

- M. Charikar, K. Makarychev, and Y. Makarychev. *Near-Optimal Algorithms for Unique Games.* In Proceedings of the 38th ACM Symposium on Theory of Computing, pp. 205–214, 2006.

- M. Charikar, K. Makarychev, and Y. Makarychev. *Near-Optimal Algorithms for Maximum Constraint Satisfaction Problems.* In Proceedings of the 18th ACM-SIAM Symposium on Discrete Algorithms, pp. 62–68, 2007.

- E. Chlamtac, K. Makarychev, and Y. Makarychev. *How to Play Unique Games Using Embeddings.* In Proceedings of the 47th IEEE Symposium on Foundations of Computer Science pp. 687–696, 2006.

1.1 Unique Games

The first problem we study is the Unique Games Problem. It is a generalization of many constraint satisfaction problems. Particularly important special cases are MAX CUT and MAX 2-LIN (systems of linear equations $\bmod p$ with at most two equations in each equation).

Definition 1.1.1 (Unique Games Problem)**.** *Given a constraint graph $G = (V, E)$ and a set of permutations π_{uv} on $[k] \equiv \{1, \ldots, k\}$ (for all edges (u, v)), assign a value (state) x_u from the set $[k]$ to each vertex u so as to satisfy the maximum number of constraints of the form $\pi_{uv}(x_u) = x_v$.*

Remark 1.1.2. *Let us explain why the problem is called "a game". The problem was first introduced by Feige and Lovász [15] as a two-prover one round game. Two "provers" play the following cooperative game. The "verifier" (the referee) selects an edge (u, v) of G uniformly at random and sends u to one prover and v to the* *other prover (provers cannot communicate with each other). Each prover has to reply with a number in* $[k]$*. The provers win (the verifier accepts their answers) if*

$$\pi_{uv}(\text{number the 1st prover replied with}) = \text{number the 2nd prover replied with.}$$

It turns out that the value of this game equals the fraction of the constraints satisfied by the optimal solution of the unique game. We refer the reader to [15] for more background on multi-prover games and proof systems.

In any instance of Unique Games if all constraints are satisfiable then it is easy to find a satisfying assignment. However, even if almost all constraints are satisfiable, it is NP-hard to find the optimal solution. Moreover, Khot [20] conjectured that for every positive ε and δ, there exists k such that it is NP-hard to distinguish between the case where a $(1-\varepsilon)$ fraction of all constraints is satisfiable and the case where at most a δ fraction of all constraints is satisfiable. This conjecture, known as the Unique Games Conjecture, implies many inapproximability results for fundamental problems — MAX CUT [23, 28], MIN 2CSP [11, 20], MultiCut and Sparsest Cut [11, 22], Vertex Cover [21] — which are not known to follow from more standard complexity assumptions. Thus it is interesting to determine what fraction of constraints can be satisfied for such instances as a function of ε, k and n (where n is the number of vertices, and there exists an assignment satisfying a $(1-\varepsilon)$ fraction of all constraints).

Note that a random assignment satisfies a $1/k$ fraction of the constraints in a unique game. Andersson, Engebretsen, and Håstad [2] considered semidefinite program (SDP) based algorithms for systems of linear equations mod p (with two variables per equation) and gave an algorithm that performs (very slightly) better than a random assignment. The first approximation algorithm for general Unique

Games was given by Khot [20], and satisfies a $1-O(k^2\varepsilon^{1/5}\sqrt{\log(1/\varepsilon)})$ fraction of all constraints if a $1-\varepsilon$ fraction of all constraints is satisfiable. Recently Trevisan [32] developed an algorithm that satisfies a $1-O(\sqrt[3]{\varepsilon\log n})$ fraction of all constraints (this can be improved to $1-O(\sqrt{\varepsilon\log n})$ [19]), and Gupta and Talwar [19] developed an algorithm that satisfies a $1-O(\varepsilon\log n)$ fraction of all constraints. The result of [19] is based on rounding an LP relaxation for the problem, while previous results use SDP relaxations for Unique Games.

There are very few results that show hardness of Unique Games. Feige and Reichman [13] showed that for every positive ε there is c s.t. it is NP-hard to distinguish between the case where a c fraction of all constraints is satisfiable and the case where only an εc fraction is satisfiable.

1.1.1 Our Results

We present three new approximation algorithms for Unique Games. We state our guarantees for instances where a $1-\varepsilon$ fraction of constraints is satisfiable. The first algorithm satisfies an

$$\Omega\left(\min(1,\frac{1}{\sqrt{\varepsilon\log k}})\cdot(1-\varepsilon)^2\cdot\left(\frac{k}{\sqrt{\log k}}\right)^{-\varepsilon/(2-\varepsilon)}\right) \tag{1.1}$$

fraction of all constraints; the second algorithm satisfies a $1-O(\sqrt{\varepsilon\log k})$ fraction of all constraints; and the third algorithm satisfies a $1-O(\varepsilon\sqrt{\log n\log k})$ fraction of all constraints. We also present an approximation algorithm (based on the same techniques) for d-to-1 games[1].

[1]Constraints in d-to-1 games can be more general than in unique games. Specifically, for every value of u there can be up to d values of v that satisfy the constraint between u and v; but for every value of v there should be only one value of u that satisfies the constraint. See Section 2.7

In order to understand the complexity theoretic implications of our results, it is useful to keep in mind that inapproximability reductions from Unique Games typically use the "Long Code", which increases the size of the instance by a 2^k factor. Thus, such applications of Unique Games usually have domain size $k = O(\log n)$.

Our results show limitations on the hardness bounds achievable using the UGC and stronger versions of it. Chawla, Krauthgamer, Kumar, Rabani, and Sivakumar [11] proposed a strengthened form of the UGC, conjecturing that it holds for $k = \log n$ and $\varepsilon = \delta = \frac{1}{(\log n)^{\Omega(1)}}$. This was used to obtain an $\Omega(\log\log n)$ hardness for Sparsest Cut. Our results refute this strengthened conjecture[2].

The performance of our algorithms is naturally constrained by the integrality gap of the SDP relaxation, i.e. the smallest possible value of an integer solution for an instance with SDP solution of value $(1-\varepsilon)|E|$. Khot and Vishnoi [22] constructed a gap instance for the semidefinite relaxation[3] for the Unique Games Problem where the SDP satisfies a $(1-\varepsilon)$ fraction of constraints, but the optimal solution can satisfy at most $O(k^{-\varepsilon/9})$ (one may show that their analysis can yield $O(k^{-\varepsilon/4+o(\varepsilon)})$). This shows that our results are almost optimal for the standard semidefinite program.

Khot, Kindler, Mossel and O'Donnell [23] proved lower bounds for Unique Games that almost match the upper bounds we obtain in the first and second algorithms. Specifically, they established the following hardness results:

Theorem 1.1.3 ([23], Corollary 13)**.** *The Unique Games Conjecture implies that for every fixed $\varepsilon > 0$, for all $k > k(\varepsilon)$, it is NP-hard to distinguish between instances*

---

for the formal definition.

[2]An updated version of [11] proposes a different strengthened form of the UGC, which is still plausible given our algorithms. They use a modified analysis to account for the asymmetry in ε and δ to obtain an $\Omega(\sqrt{\log\log n})$ hardness for sparsest cut based on this.

[3]We use a slightly stronger SDP than they used, but their integrality gap construction works for our SDP as well.

of Unique Games with domain size k where at least $1-\varepsilon$ fraction of constraints are satisfiable and those where $1/k^{\varepsilon/(2-\varepsilon)}$ fraction of constraints are satisfiable.

Theorem 1.1.4 ([23], Corollary 14)**.** *The Unique Games Conjecture implies that for every fixed $\varepsilon > 0$, for all k, it is NP-hard to distinguish between instances of Unique Games with domain size k where at least $1-\varepsilon$ fraction of constraints are satisfiable and those where $1 - \sqrt{2/\pi}\sqrt{\varepsilon \log k} + o(1)$ fraction of constraints are satisfiable.*

Thus, two of our bounds are near optimal if the UGC is true — even a slight improvement of the results $1/k^{\varepsilon/(2-\varepsilon)}$ or $1 - O(\sqrt{\varepsilon \log k})$ (beyond low order terms) would disprove the Unique Games Conjecture!

1.2 MAX 2CSP

In the second part of the dissertation, we study the MAX 2CSP problem.

Definition 1.2.1 (MAX 2CSP)**.** *We are given a set of boolean variables $\{x_i\}$ and a set of predicates (constraints), each predicate depends on at most two variables. Our goal is to assign a value to each x_i so as to satisfy the maximum number of predicates.*

MAX 2CSP is a very basic constraint satisfaction problem, which generalizes several important graph partitioning problems such as MAX CUT or MAX DICUT (the directed version of MAX CUT).

The problem and its variants have been well studied in the literature (see Figure 1.2). In particular, current lower and upper bounds for the approximation ratio are almost tight. Lewin, Livnat and Zwick [26] developed an algorithm with approximation ratio 0.87401 (improving the previous results of Goemans and

	Approximation Guarantee		Hardness Result	
Approximation ratio	0.87401	[26]	0.87435	[6]
Almost satisfiable instances, $OPT = 1 - \varepsilon$				
• $\varepsilon > 1/\log^{3/2} n$	$1 - O(\varepsilon^{1/3})$	[34]	$1 - O(\sqrt{\varepsilon})$	[23]
• $\varepsilon < 1/\log^{3/2} n$	$1 - \tilde{O}(\varepsilon \log n)$	[24]	$1 - O(\varepsilon\sqrt{\log\log n})$	[11]

Figure 1.2: Previously Known Approximation Results for MAX 2CSP. Note that the approximation ratio almost matches the hardness result from [6]. The hardness results in [6] and [23] are based on the Unique Games Conjecture; the hardness result in [11] is based on a stronger version of the Unique Games Conjecture.

Williamson [17], Feige and Goemans [14], Zwick [33], Matuura and Matsui [27]). Recently Austrin [6] proved an almost matching upper bound of 0.87435 assuming the Unique Games Conjecture (improving the result of Khot, Kindler, Mossel and O'Donnell [23]).

However, the tight bounds are still not known for almost satisfiable instances. Note that even if all constraints are satisfiable the result of [26] only guarantees that the algorithm satisfies a 0.874 fraction of all constraints. The best previously known lower bounds for almost satisfiable instances of MAX 2SAT were proved by Zwick [34] and Klein, Plotkin, Rao, Tardos [24]. If the optimal solution satisfies an $OPT = 1 - \varepsilon$ fraction of all constraints, the algorithm by Zwick [34] satisfies a $1 - O(\varepsilon^{1/3})$ fraction of all constraints; the algorithm by Klein, Plotkin, Rao, and Tardos [24] satisfies a $1 - O(\varepsilon \log n \log\log n)$ fraction of all constraints. The latter algorithm gives a better approximation when ε is very small.

Interestingly, algorithms with better approximation guarantees were known for MAX CUT. The algorithm of Goemans and Williamson [17] satisfied $1 - O(\sqrt{\varepsilon})$ fraction of all constraints; the algorithm of Garg, Vazirani, and Yannakakis satisfied

$1 - O(\varepsilon \log n)$ fraction of all constraints.

We present new approximation algorithms for MAX 2CSP that satisfy $1-O(\sqrt{\varepsilon})$ and $1 - O(\varepsilon\sqrt{\log n})$ fraction of all constraints. Our first guarantee for arbitrary MAX 2CSP matches the guarantee of Goemans and Williamson for MAX CUT. It is optimal assuming the Unique Games Conjecture [23]. Our second result improves the approximation guarantee not only for MAX 2CSP but also for MAX CUT.

Chapter 2

Unique Games

2.1 Overview

2.1.1 Semidefinite Programming

Our algorithms for Unique Games are based on Semidefinite Programming (SDP). Semidefinite Programming is a powerful technique, which has been recently used to construct many combinatorial optimization algorithms. SDP solves the following problem:

$$\text{minimize (or maximize) } f(X) \tag{2.1}$$

subject to

$$\forall i \in \{1, \dots, m\} \quad A_i(X) \geq b_i$$
$$X \succeq 0$$

where $X \in M_n(\mathbb{R})$ is an $n \times n$ matrix; f, A_1, $\dots$, A_m are linear functionals on the linear space $M_n(\mathbb{R})$; $X \succeq 0$ denotes that the matrix X is positive semidefinite.

Every positive semidefinite matrix $X = (x_{ij})$ is a Gram matrix (the matrix of pairwise inner products) of some n vectors $v_1, \dots v_n$ in Euclidean space $\mathbb{R}^{n-1}$:

$$x_{ij} = \langle v_i, v_j \rangle.$$

Moreover, the matrix X determines the vectors v_i up to isometry. Similarly, the Gram matrix of any n vectors is positive semidefinite. Thus we can write semidefinite program (2.1) as an equivalent *vector program.*

$$\text{minimize (or maximize)} \ \sum_{j=1}^{n} \sum_{k=1}^{n} f^{jk} \langle v_j, v_k \rangle$$

subject to

$$\forall i \in \{1, \dots, m\} \quad \sum_{j=1}^{n} \sum_{k=1}^{n} A_i^{jk} \langle v_j, v_k \rangle \geq b_i$$

$$v_1, \dots, v_n \in \mathbb{R}^{n-1}$$

where $A_i^{jk} = A_i(e_j \otimes e_k)$, $f^{jk} = f(e_j \otimes e_k)$, and $e_1, \dots, e_n$ is the standard basis of $\mathbb{R}^n$. The now-standard approach to solving combinatorial optimization problems is as follows.

1. We write the problem as an equivalent integer quadratic problem with 0-1 variables.

2. We *relax* the integer problem. For every 0-1 variable x_i we introduce a vector valued variable v_i; we replace each product $x_i \cdot x_j$ with the inner product

$\langle v_i, v_j \rangle$ in the objective function and constraints of the integer program. We obtain *a semidefinite relaxation* of the problem. Note that for every feasible solution x_i of the original quadratic problem there is a corresponding feasible solution (*intended solution*) $v_i = x_i e$ (where e is a fixed unit vector) of the relaxation, which has the same value. Thus the cost of the optimal solution of the semidefinite program (*SDP value*) is a lower bound on the cost of the combinatorial solution if the problem is a minimization problem; and it is an upper bound if the problem is a maximization problem. We describe the semidefinite relaxation for Unique Games in Section 2.2.

3. We solve the semidefinite relaxation. This can be done efficiently (in polynomial time) with an arbitrary precision as was first shown by Grötschel, Lovász, and Schrijver[18]. The solution is a set of vectors $\{v_i\}$

4. Finally, we need to *round* (transform) the semidefinite solution to a feasible solution of the original combinatorial problem and then estimate the cost of the obtained solution. Rounding is the most interesting step of the algorithm. There is often no easy way to round an SDP solution to a good combinatorial solution.

This approach is well suited for solving Constraint Satisfaction Problems with boolean variables since such problems are naturally cast as integer quadratic programs. Every 0-1 assignment of variables x_i corresponds to a feasible solution to the boolean CSP. Thus the rounding algorithm, roughly speaking, can process vectors v_i one by one and "round" them to zeros and ones. However, developing rounding techniques for problems with large domain size is more challenging. For each variable x_u that takes k values $1, \dots, k$, we have to introduce k indicator variables for the events "$x_u = 1$", ..., "$x_u = k$". These variables are not independent: exactly one of them must be equal to 1 and the remaining must be equal to 0. We denote vector variables corresponding to these indicator variables by $u_1, \dots, u_k$. The rounding algorithm has to assign a value to x_u based on the spatial configuration of the vectors $u_1, \dots, u_k$. It makes the algorithm and analysis considerably more complex.

In this dissertation, we develop new rounding techniques that enable us to get near optimal algorithms for Unique Games (assuming UGC). We hope that these techniques would be useful for rounding other CSPs and assignment problems with large domain.

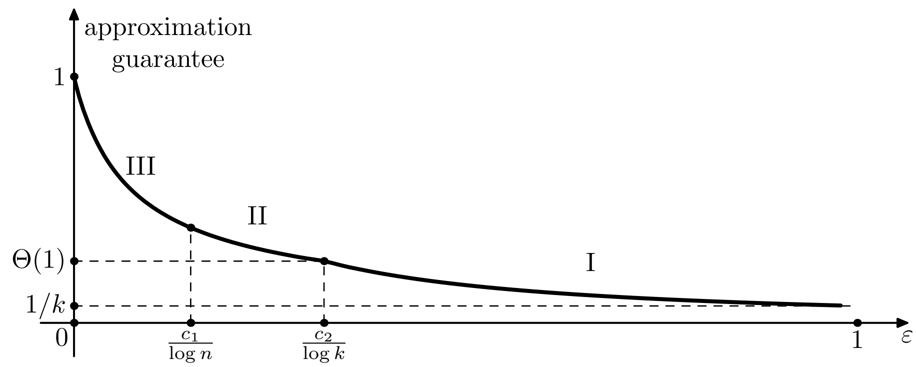
2.1.2 Techniques

In this section, we briefly overview how we round the solution of the SDP relaxation for Unique Games in each of our three algorithms.

The SDP solution is a set of vectors. For every vertex u, we have k vectors: u_1, ..., u_k. Our SDP program (see Section 2.2) enforces that vectors corresponding to one vertex are orthogonal.

We interpret the SDP solution as a probability distribution on assignments of values to variables and the goal of our rounding algorithm is to pick an assignment to variables by sampling from this distribution such that values of variables connected by constraints are strongly correlated.

The rough idea of our first and second algorithms is to pick a random vector and examine the projections of this vector on u_i, picking a value i for u for which u_i has a large projection. (In fact, this is exactly the algorithm of Khot [20]). We have to modify this basic idea to obtain our results since the u_i's could have different



	Range of ε	Approximation Guarantee
I	$\varepsilon \geq c_2/\log k$	$\sim k^{-\frac{\varepsilon}{2-\varepsilon}}$
II	$c_1/\log n \leq \varepsilon \leq c_2/\log k$	$1 - O(\sqrt{\varepsilon \log k})$
III	$\varepsilon \leq c_1/\log n$	$1 - O(\varepsilon\sqrt{\log n \log k})$

Figure 2.1: Comparison of Approximation Guarantees for Unique Games. The figure shows the approximation guarantee of each algorithm and the range of ε where each algorithm performs better then the others.

lengths and other complications arise. Instead of picking one random vector, we pick several Gaussian random vectors. Our first algorithm (suitable for large ε) picks a small set of candidate assignments for each variable and chooses randomly amongst them (independently for every variable). It is interesting to note that such a multiple assignment is often encountered in algorithms implicit in hardness reductions involving label cover. This algorithm has a non-trivial guarantee even for very large ε. In contrast, the previous results have non-trivial guarantees only when ε is small; specifically, the algorithm of Khot [20] is applicable only when $\varepsilon = o(1/k^{10})$; the result of Gupta and Talwar is applicable only when $\varepsilon = O(1/\log n)$. As ε approaches 1 (i.e. for instances where the optimal solution satisfies only a small fraction of the constraints), the performance guarantee of our algorithm approaches that of a random assignment. Our second algorithm (suitable for $\varepsilon \in (c_1/\log n, c_2/\log k)$) carefully picks a single assignment so that almost all constraints are satisfied.

To construct our third algorithm, we introduce a new type of random partitioning scheme, which we call an *m-orthogonal separator* (where m is a parameter). Specifically, we design an algorithm that, given a set of vectors in an ℓ_2^2 space (see Definition 2.5.2), produces a random subset S such that the probability that two orthogonal vectors belong to S is equal to $1/m$ (we assume that $1/m$ is very small); and the distribution over corresponding cuts $(S, \bar{S})$ is a low distortion embedding from ℓ_2^2 into ℓ_1. In other words, for two orthogonal vectors u and v the events "$u \in S$" and "$v \in S$" are "almost" disjoint. This property is crucial for our algorithm: it essentially guarantees that we assign only one value to each vertex in a unique game. We stress that no previously known embedding satisfies this property.

In order to construct orthogonal separators we extend the methods developed for the first two algorithms and combine them with powerful metric embedding techniques developed in the works of Arora, Rao, and Vazirani [4], of Lee [25], of Chawla, Gupta, and Räcke [10], and of Arora, Lee, and Naor [3] using a new transformation of the space ℓ_2^2, which we call "normalization".

We present two constructions: one using embeddings from ℓ_2^2 into ℓ_1 and the other using embeddings from ℓ_2^2 into ℓ_2. While the second construction gives a slightly better guarantee, the first construction would be improved even if better embedding techniques from ℓ_2^2 into ℓ_1 at one scale are found (this cannot happen for embedding into ℓ_2, for which current guarantees are essentially tight).

2.2 Semidefinite Relaxation

First we reduce a unique game to an integer quadratic program. We denote the set of states by $[k] \equiv \{1, \ldots, k\}$. For each vertex u we introduce k indicator variables

$u_i \in \{0,1\}$ ($i \in [k]$) for the events $x_u = i$. For every u, the intended solution has $u_i = 1$ for exactly one i. The constraint $\pi_{uv}(x_u) = x_v$ can be restated in the following form:

$$\text{for all } i \quad u_i = v_{\pi_{uv}(i)}.$$

The unique game instance is equivalent to the following integer quadratic program:

$$\text{minimize } \frac{1}{2} \sum_{(u,v)\in E} \left(\sum_{i=1}^{k} \|u_i - v_{\pi_{uv}(i)}\|^2 \right)$$

$$\begin{aligned} \text{subject to} \quad \forall u \in V \;\; \forall i \in [k] \qquad & u_i \in \{0,1\} \\ \forall u \in V \;\; \forall i,j \in [k], i \neq j \qquad & u_i \cdot u_j = 0 \\ \forall u \in V \qquad & \sum_{i=1}^{k} u_i^2 = 1 \end{aligned}$$

Note that the objective function measures the number of *unsatisfied* constraints. The contribution of edge (u,v) to the objective function is equal to 0 if the constraint π_{uv} is satisfied, and 1 otherwise. The last two equations say that exactly one u_i is equal to 1.

We now replace each integer variable u_i with a vector variable and get a semidefinite program (SDP):

$$\text{minimize } \frac{1}{2} \sum_{(u,v)\in E} \sum_{i=1}^{k} \|u_i - v_{\pi_{uv}(i)}\|^2$$

subject to

$$\forall u \in V\ \forall i,j \in [k], i \neq j \qquad \langle u_i, u_j \rangle = 0 \tag{2.2}$$

$$\forall u \in V \qquad \sum_{i=1}^{k} \|u_i\|^2 = 1 \tag{2.3}$$

$$\forall u,v,w \in V\ \forall i,j,l \in [k] \qquad \|u_i - w_l\|^2 \leq \|u_i - v_j\|^2 + \|v_j - w_l\|^2 \tag{2.4}$$

$$\forall u,v \in V\ i,j \in [k] \qquad \langle u_i, v_j \rangle \geq 0 \tag{2.5}$$

$$\forall u,v \in V\ i,j \in [k] \qquad \langle u_i, v_j \rangle \leq \|u_i\|^2 \tag{2.6}$$

The last two constraints are triangle inequality constraints[1] for the squared Euclidean distance: inequality (2.5) is equivalent to $\|u_i - 0\|^2 + \|v_j - 0\|^2 \geq \|u_i - v_j\|^2$, and inequality (2.6) is equivalent to $\|u_i - v_j\|^2 + \|u_i - 0\|^2 \geq \|v_j - 0\|^2$. A very important constraint is that for $i \neq j$ the vectors u_i and u_j are orthogonal. This SDP was studied by Khot [20], and by Trevisan [32].

Here is an intuitive interpretation of the vector solution: Think of the elements of the set $[k]$ as states of the vertices. In the integer case, if $u_i = 1$, the vertex is in the state i. In the vector case, each vertex is in a *mixed* state, and the probability that $x_u = i$ is equal to $\|u_i\|^2$. The inner product $\langle u_i, v_j \rangle$ can be thought of as the joint probability that $x_u = i$ and $x_v = j$. The directions of vectors determine whether two states are correlated or not: If the angle between u_i and v_j is small it is likely that both events "*u is in the state i*" and "*v is in the state j*" occur simultaneously. In some sense later we will treat the lengths and the directions of vectors separately.

[1]We will use constraint 2.5 only in the second algorithm.

2.3 First Algorithm

2.3.1 Algorithm

We first describe a high level idea for the first algorithm. Pick a random Gaussian vector g (with standard normal independent components). For every vertex u add those vectors u_i whose inner product with g are above some threshold τ to the set S_u; we choose the threshold τ in such a way that the set S_u contains only one element in expectation. Then pick a random state from S_u and assign it to the vertex u (if S_u is empty do not assign any states to u). What is the probability that the algorithm satisfies a constraint between vertices u and v? Loosely speaking, this probability is equal to

$$\mathbb{E}\left[\frac{|S_u \cap \pi_{uv}(S_v)|}{|S_u| \cdot |S_v|}\right] \approx \mathbb{E}\left[|S_u \cap \pi_{uv}(S_v)|\right].$$

Assume for a moment that the SDP solution is symmetric: the lengths of all vectors u_i are the same and the squared Euclidean distance between every u_i and $v_{\pi_{uv}(i)}$ is equal to 2ε. (In fact, we can add to SDP the condition that $|u_i| = 1/\sqrt{k}$ (for all u and i) in the special case of systems of linear equations of the form $x_i - x_j = c_{ij}$ (mod p).) Since we want the expected size of S_u to be 1, we pick threshold τ such that the probability that $\langle g, u_i\rangle \geq \tau$ equals $1/k$. The random variables $\langle g, \sqrt{k}\cdot u_i\rangle$ and $\langle g, \sqrt{k}\cdot v_{\pi_{uv}(i)}\rangle$ are standard normal random variables with covariance $1-\varepsilon$ (note that we multiplied the inner products by a normalization factor of $\sqrt{k}$). For such random variables if the probability of the event $\langle g, \sqrt{k}\cdot u_i\rangle \geq t \equiv \sqrt{k}\tau$ equals $1/k$, then roughly speaking the probability of the event $\langle g, \sqrt{k}\cdot u_i\rangle \geq t \equiv \sqrt{k}\cdot\tau$ and $\langle g, \sqrt{k}\cdot v_{\pi_{uv}(i)}\rangle \geq t \equiv \sqrt{k}\cdot\tau$ equals $k^{-\varepsilon/2}\cdot 1/k$. Thus the expected size of the intersection of the sets S_u and $\pi_{uv}(S_v)$ is approximately $k^{-\varepsilon/2}$.

Unfortunately this no longer works if the lengths of vectors are different. The main problem is that if, say, u_1 is two times longer than u_2, then $\Pr(u_1 \in S_u)$ is much larger than $\Pr(u_2 \in S_u)$.

One of the possible solutions is to normalize all vectors first. In order to take into account original lengths of vectors we repeat the procedure of adding vectors to the sets S_u many times, but each vector u_i has a chance to be selected in the set S_u only in the first $s_{u,i}$ trials, where $s_{u,i}$ is some integer number proportional to the original squared Euclidean length of u_i.

We now formally present a rounding algorithm for the SDP described in the previous section.

Theorem 2.3.1. *There is a polynomial time algorithm that finds an assignment of variables which satisfies a*

$$\Omega\left(\min(1, \frac{1}{\sqrt{\varepsilon \log k}}) \cdot (1-\varepsilon)^2 \cdot \left(\frac{k}{\sqrt{\log k}}\right)^{-\varepsilon/(2-\varepsilon)}\right)$$

fraction of all constraints if the optimal solution satisfies $(1-\varepsilon)$ *fraction of all constraints.*

The algorithm is presented in Figure 2.2. We introduce some notation.

Definition 2.3.2. *Define the distance between two vertices* u *and* v *as:*

$$\varepsilon_{uv} = \frac{1}{2}\sum_{i=1}^{k} \|u_i - v_{\pi_{uv}(i)}\|^2$$

and let

$$\varepsilon_{uv}^i = \frac{1}{2}\|\tilde{u}_i - \tilde{v}_{\pi_{uv}(i)}\|^2.$$

Input: *A solution of the SDP, with the objective value* $\varepsilon \cdot |E|$.
Output: *An assignment of variables* x_u.

1. Define $\tilde{u}_i = u_i/\|u_i\|$ if $u_i \neq 0$, 0 otherwise.

   Note that vectors $\tilde{u}_1, \ldots, \tilde{u}_k$ are orthogonal unit vectors (except for those vectors that are equal to zero).

2. Pick random independent Gaussian vectors $g_1, \ldots, g_k$ with independent components distributed as $\mathcal{N}(0, 1)$.

3. For each vertex u:

   (a) Set $s_{u_i} = \lceil \|u_i\|^2 \cdot k \rceil$.

   (b) For each i project s_{u_i} vectors $g_1, \ldots, g_{s_{u_i}}$ to $\tilde{u}_i$:

   $$\xi_{u_i,s} = \langle g_s, \tilde{u}_i \rangle, \ 1 \leq s \leq s_{u_i}.$$

   Note that $\xi_{u_1,1}, \xi_{u_1,2}, \ldots, \xi_{u_1,s_{u_1}}, \ldots, \xi_{u_k,1}, \ldots, \xi_{u_k,s_{u_k}}$ are independent standard normal random variables. (Since u_i and u_j are orthogonal if $i \neq j$, their projections onto a random Gaussian vector are independent). The number of random variables corresponding to each u_i is proportional to $\|u_i\|^2$.

   (c) Fix a threshold t s.t. $\Pr(\xi \geq t) = 1/k$, where $\xi \sim \mathcal{N}(0, 1)$ (i.e. t is the $(1 - 1/k)$-quantile of the standard normal distribution; note that by Lemma 2.6.11, part 2, $t = \Theta\left(\sqrt{\log k}\right)$).

   (d) Pick those ξ_{u_i} that are larger than the threshold t:

   $$S_u = \{(i, s) : \xi_{u_i,s} \geq t\}.$$

   (e) For each vertex u, pick at random a pair (i, s) from S_u and assign $x_u = i$. If the set S_u is empty do not assign any value to the vertex: this means that all the constraints containing the vertex are not satisfied.

Figure 2.2: First Algorithm for Unique Games

If u_i *and* $v_{\pi_{uv}(i)}$ *are nonzero vectors and* α_i *is the angle between them, then* $\varepsilon^i_{uv} = 1 - \cos \alpha_i$. *For consistency, if one of the vectors is equal to zero we set* $\varepsilon^i_{uv} = 1$ *and*

$\alpha_i = \pi/2$.

We need several estimates on the joint distribution of gaussian variables in our analysis of this algorithm. We defer the proofs of these estimates to Section 2.3.2.

Lemma 2.3.3. *For every edge (u, v), state i in $[k]$ and $s \le \min(s_{u_i}, s_{v_{\pi_{uv}(i)}})$ the probability that the algorithm picks (i, s) for the vertex u and $(\pi_{uv}(i), s)$ for v at the step 3.e is*

$$\Omega\left(\min(1, \frac{1}{\sqrt{\varepsilon_{uv}^i \log k}}) \cdot \frac{1}{\sqrt{\log k}} \cdot \left(\frac{\sqrt{\log k}}{k}\right)^{2/(2-\varepsilon_{uv}^i)}\right). \tag{2.7}$$

Proof. First let us observe that $\xi_{u_i,s}$ and $\xi_{v_{\pi_{uv}(i)},s}$ are standard normal random variables with covariance $\cos\alpha_i = 1 - \varepsilon_{uv}^i$. We prove in Section 2.3.2 (Corollary 2.3.10) that the probability that $\xi_{u_i,s} \ge t$ and $\xi_{v_{\pi_{uv}(i)},s} \ge t$ is equal to (2.7).

Note that the expected number of elements in S_u is equal to $(s_{u_1} + \ldots + s_{u_k})/k$ which is at most 2. Moreover, as we will prove in Lemma 2.3.11, the conditional expected number of elements in S_u given the event $\xi_{u_i,s} \ge t$ and $\xi_{v_{\pi_{uv}(i)},s} \ge t$ is also a constant. Thus by the Markov inequality the following event happens with probability (2.7): The sets S_u and S_v contain the pairs (i, s) and $(\pi_{uv}(i), s)$ respectively and the sizes of these sets are bounded by a constant. The lemma follows. □

Definition 2.3.4. *For brevity, denote $\left(\sqrt{\log k}/k\right)^{2/(2-x)}$ by $f_k(x)$.*

Remark 2.3.5. *It is instructive to consider the case when the SDP solution is uniform in the following sense:*

1. *The lengths of all vectors u_i are the same and are equal to $1/\sqrt{k}$.*

2. All ε_{uv}^i are equal to ε.

In this case all s_{u_i} are equal to 1. And thus the probability that a constraint is satisfied is k times the probability (2.7) which is equal, up to a logarithmic factor, to $k^{-\varepsilon/(2-\varepsilon)}$. Multiplying this probability by the number of edges we get that the expected number of satisfied constraints is $k^{-\varepsilon/(2-\varepsilon)}|E|$.

In the general case, however, we need to do some extra work to average the probabilities among all states i and edges (u, v).

Recall that we interpret $\|u_i\|^2$ as the probability that the vertex u is in the state i. Suppose now that the constraint between u and v is satisfied, what is the conditional probability that u is in the state i and v is in the state $\pi_{uv}(i)$? Roughly speaking, it should be equal to $(\|u_i\|^2 + \|v_{\pi_{uv}(i)}\|^2)/2$. This motivates the following definition.

Definition 2.3.6. *Define a measure μ_{uv} on the set $[k]$:*

$$\mu_{uv}(T) = \sum_{i \in T} \frac{\|u_i\|^2 + \|v_{\pi_{uv}(i)}\|^2}{2}, \text{ where } T \subset [k].$$

Note that $\mu_{uv}([k]) = 1$. This follows from constraint (2.3).

The following lemma shows why this measure is useful.

Lemma 2.3.7. *For every edge (u, v) the following statements hold.*

1. The average value of ε_{uv}^i w.r.t. the measure μ_{uv} is less than or equal to ε_{uv}:

$$\sum_{i=1}^{k} \mu_{uv}(i)\varepsilon_{uv}^i \leq \varepsilon_{uv}.$$

2. For every i,

$$\min(s_{u_i}, s_{v_{\pi_{uv}(i)}}) \geq (1 - \varepsilon_{uv}^i)^2 \mu_{uv}(i) k.$$

Proof. 1. Indeed,

$$\begin{aligned}
\sum_{i=1}^{k} \mu_{uv}(i) \cdot \varepsilon_{uv}^i &= \sum_{i=1}^{k} \frac{\|u_i\|^2 + \|v_{\pi_{uv}(i)}\|^2 - (\|u_i\|^2 + \|v_{\pi_{uv}(i)}\|^2) \cdot \cos\alpha_i}{2} \\
&\leq \sum_{i=1}^{k} \frac{\|u_i\|^2 + \|v_{\pi_{uv}(i)}\|^2 - 2 \cdot \|u_i\| \cdot \|v_{\pi_{uv}(i)}\| \cdot \cos\alpha_i}{2} \\
&= \sum_{i=1}^{k} \frac{\|u_i - v_{\pi_{uv}(i)}\|^2}{2} = \varepsilon_{uv}
\end{aligned}$$

Note that here we used the fact that $\langle u_i, v_{\pi_{uv}(i)} \rangle \geq 0$ and, therefore, $\cos\alpha_i \geq 0$.

2. Without loss of generality assume that $\|u_i\| \leq \|v_{\pi_{uv}(i)}\|$, and hence

$$\min(s_{u_i}, s_{v_{\pi_{uv}(i)}}) = s_{u_i}.$$

Due to the triangle inequality constraint (2.6) in the SDP $\|v_{\pi_{uv}(i)}\| \cos\alpha_i \leq \|u_i\|$. Thus

$$(1 - \varepsilon_{uv}^i)^2 \mu_{uv}(i) = \cos^2\alpha_i \cdot \frac{\|u_i\|^2 + \|v_{\pi_{uv}(i)}\|^2}{2} \leq \|u_i\|^2 \leq s_{u_i}/k.$$

□

Lemma 2.3.8. *For every edge (u, v) the probability that an assignment found by the algorithm satisfies the constraint $\pi_{uv}(x_u) = x_v$ is*

$$\Omega\left(\frac{k}{\sqrt{\log k}} \cdot \min(1, \frac{1}{\sqrt{\varepsilon_{uv} \log k}}) \cdot f_k(\varepsilon_{uv})\right). \tag{2.8}$$

Proof. Denote the desired probability by P_{uv}. It is equal to the sum of the probabilities obtained in Lemma 2.3.3 over all $i \in [k]$ and $s \le \min(s_{u_i}, s_{v_{\pi_{uv}(i)}})$. In other words,

$$P_{uv} = \Omega\left(\sum_{i=1}^{k} \min(s_{u_i}, s_{v_{\pi_{uv}(i)}}) \frac{1}{\sqrt{\log k}} \min(1, \frac{1}{\sqrt{\varepsilon_{uv}^i \log k}}) f_k(\varepsilon_{uv}^i)\right).$$

Replacing $\min(s_{u_i}, s_{v_{\pi_{uv}(i)}})$ with $(1-\varepsilon_{uv}^i)^2 \mu_{uv}(i) \cdot k$ we get

$$P_{uv} = \Omega\left(\frac{k}{\sqrt{\log k}} \sum_{i=1}^{k} \mu_{uv}(i) \min(1, \frac{1}{\sqrt{\varepsilon_{uv}^i \log k}})(1-\varepsilon_{uv}^i)^2 f_k(\varepsilon_{uv}^i)\right).$$

Consider the set $M = \{i \in [k] : \varepsilon_{uv}^i \le 2\varepsilon_{uv}\}$. For i in M the term $\sqrt{\varepsilon_{uv}^i \log k}$ is bounded from above by $\sqrt{2\varepsilon_{uv} \log k}$. Thus

$$P_{uv} = \Omega\left(\frac{k}{\sqrt{\log k}} \min(1, \frac{1}{\sqrt{\varepsilon_{uv} \log k}}) \sum_{i \in M} \mu_{uv}(i)(1-\varepsilon_{uv}^i)^2 f_k(\varepsilon_{uv}^i)\right).$$

The function $(1-x)^2 f_k(x)$ is convex on $[0,1]$ (see Lemma 2.3.12). The average value of ε_{uv}^i among i in M (w.r.t. the measure μ_{uv}) is at most the average value of ε_{uv}^i among all i, which in turn is less than ε_{uv} according to Lemma 2.3.7. Finally, by the Markov inequality $\mu_{uv}(M) \ge 1/2$. Thus by Jensen's inequality

$$\begin{aligned} P_{uv} &= \Omega\left(\frac{k}{\sqrt{\log k}} \min(1, \frac{1}{\sqrt{\varepsilon_{uv} \log k}}) \mu_{uv}(M)(1-\varepsilon_{uv})^2 \cdot f_k(\varepsilon_{uv})\right) \\ &= \Omega\left(\frac{k}{\sqrt{\log k}} \min(1, \frac{1}{\sqrt{\varepsilon_{uv} \log k}})(1-\varepsilon_{uv})^2 \cdot f_k(\varepsilon_{uv})\right). \end{aligned}$$

This finishes the proof. □

We are now in position to prove the main theorem.

Theorem 2.3.1. *There is a polynomial time algorithm that finds an assignment of* *variables which satisfies a*

$$\Omega\left(\min(1, \frac{1}{\sqrt{\varepsilon \log k}}) \cdot (1-\varepsilon)^2 \cdot \left(\frac{k}{\sqrt{\log k}}\right)^{-\varepsilon/(2-\varepsilon)}\right)$$

fraction of all constraints if the optimal solution satisfies $(1-\varepsilon)$ *fraction of all constraints.*

Proof. Let us restrict our attention to a subset of edges $E' = \{(u,v) \in E : \varepsilon_{uv} \leq 2\varepsilon\}$. For (u,v) in E', since $\varepsilon_{uv} \log k \leq 2\varepsilon \log k$, we have

$$P_{uv} = \Omega\left(\frac{k}{\sqrt{\log k}} \min(1, \frac{1}{\sqrt{\varepsilon \log k}})(1-\varepsilon_{uv})^2 \cdot f_k(\varepsilon_{uv})\right).$$

Summing this probability over all edges (u,v) in E' and using convexity of the function $(1-x)^2 f_k(x)$ we get the statement of the theorem. □

2.3.2 Analysis: Technical Details

In this section we will prove some technical lemmas we used in the analysis of the first algorithm. Let us denote the probability that a standard normal random variable is greater than $t \in \mathbb{R}$ by $\tilde{\Phi}(t)$. We prove standard estimates for $\tilde{\Phi}(t)$, which we use in this section, in the Appendix.

Lemma 2.3.9. *Let* ξ *and* η *be correlated standard normal random variables,* $0 < \varepsilon < 1$, $t \geq 1$. *If* $\operatorname{cov}(\xi, \eta) \geq 1-\varepsilon$, *then*

$$\Pr(\xi \geq t \text{ and } \eta \geq t) \geq C \cdot \min(1, (\sqrt{\varepsilon} t)^{-1}) \cdot t^{-1} \cdot (t \cdot \tilde{\Phi}(t))^{\frac{2}{2-\varepsilon}}. \qquad (2.9)$$

for some positive constant C.

Proof. Let us represent ξ and η as follows:

$$\xi = \sigma X + \sqrt{1-\sigma^2}\cdot Y; \quad \eta = \sigma X - \sqrt{1-\sigma^2}\cdot Y,$$

where

$$\sigma^2 = \mathrm{Var}\left[\frac{\xi+\eta}{2}\right]; \quad X = \frac{\xi+\eta}{2\sigma}; \quad Y = \frac{\xi-\eta}{2\sqrt{1-\sigma^2}}.$$

Note that X and Y are independent standard normal random variables; and

$$\sigma^2 = \mathrm{Var}\left[\frac{\xi+\eta}{2}\right] = \frac{1}{4}\left[2 + 2\,\mathrm{cov}(\xi,\eta)\ \right] \geq 1 - \frac{\varepsilon}{2}. \tag{2.10}$$

Notice that $1/2 \leq \sigma^2 \leq 1$. We now estimate the probability (2.9) as follows

$$\begin{aligned}\Pr\big(\xi \geq t \text{ and } \eta \geq t\big) &= \Pr\left(\sigma X \geq t + \sqrt{1-\sigma^2}\cdot |Y|\right)\\ &\geq \Pr\left(X \geq \frac{t}{\sigma} + \frac{\sigma}{t}\right)\cdot \Pr\left(|Y| \leq \frac{\sigma^2}{\sqrt{1-\sigma^2}\cdot t}\right)\end{aligned}$$

By Lemma A.1.1 (part 3) from the Appendix (with $\rho = 1/\sigma$) we get

$$\begin{aligned}\Pr\big(\xi \geq t \text{ and } \eta \geq t\big) &\geq C\cdot\left(t^{-1}\cdot (t\tilde{\Phi}(t))^{1/\sigma^2}\right)\cdot \min\left(1, \frac{\sigma^2}{\sqrt{1-\sigma^2}\cdot t}\right)\\ &\geq C'\cdot \min((\sqrt{\varepsilon}\cdot t)^{-1}, 1)\cdot t^{-1}\cdot (t\cdot\tilde{\Phi}(t))^{\frac{2}{2-\varepsilon}}.\end{aligned}$$

□

Corollary 2.3.10. *Let ξ and η be standard normal random variables with covari-*

ance greater than or equal to $1-\varepsilon$*; let* $\tilde{\Phi}(t) = 1/k$*. Then*

$$\Pr\left(\xi \geq t \text{ and } \eta \geq t\right) \geq \Omega\left(\min\left(1, \frac{1}{\sqrt{\varepsilon \log k}}\right) \cdot \frac{1}{\sqrt{\log k}} \cdot \left(\frac{k}{\sqrt{\log k}}\right)^{-\frac{2}{2-\varepsilon}}\right).$$

(If $k > 6$ then $t > 1$, so this corollary follows from Lemma 2.3.9. If $k \leq 6$ then the expression in the right hand side (inside $\Omega(\cdot)$) is at most $1/k^2 \leq 1/36$; the probability in the left hand side is at most 1. Therefore, the bound holds when the constant in the omega notation is greater than 36.)

Lemma 2.3.11. *Let* ξ*,* η*,* ε*,* k *and* t *be as in Corollary 2.3.10, and let* $\xi_1, \dots, \xi_m$ *be i.i.d. standard normal random variables and* $m \leq 2k$*, then*

$$\mathbb{E}\left[\sum_{i=1}^{m} I_{\{\xi_i \geq t\}} \mid \xi \geq t \text{ and } \eta \geq t\right] = O(1),$$

where $I_{\{\xi_i \geq t\}}$ *is the indicator of the event* $\{\xi_i \geq t\}$*.*

Proof. Let X and Y be as in the proof of Lemma 2.3.9. Put $\alpha_i = \operatorname{cov}(X, \xi_i)$ and express each ξ_i as $\xi_i = \alpha_i X + \sqrt{1-\alpha_i^2} \cdot Z_i$. By Bessel's Inequality $\alpha_1^2 + \cdots + \alpha_m^2 \leq 1$ (since random variables ξ_i are orthogonal). Let $B_{t,x}$ be the event

$$B_{t,x} = \left\{\sigma x \geq t + \sqrt{1-\sigma^2}|Y|\right\}.$$

Then

$$B_{t,X} = \{\xi \geq t \text{ and } \eta \geq t\}.$$

We now estimate the value of $\mathbb{E}\left[I_{\{\xi_i \geq t\}} \mid \xi \geq t, \eta \geq t\right] = \Pr(\xi_i \geq t \mid B_{t,X})$,

$$\begin{aligned}\Pr(\xi_i \geq t \mid B_{t,X}) &= \Pr\left(\xi_i \geq t \mid B_{t,X} \text{ and } X \leq 4t\right)\Pr\left(X \leq 4t \mid B_{t,X}\right)\\ &\quad + \Pr\left(\xi_i \geq t \mid B_{t,X} \text{ and } X > 4t\right)\Pr\left(X > 4t \mid B_{t,X}\right)\\ &\leq \Pr\left(\xi_i \geq t \mid B_{t,X} \text{ and } X \leq 4t\right) + \Pr\left(X > 4t \mid B_{t,X}\right). \end{aligned} \tag{2.11}$$

Let us bound the first term.

$$\begin{aligned}\Pr\big(\xi_i \geq t \mid B_{t,X} \text{ and } X \leq 4t\big) &= \Pr\big(\alpha_i X + \sqrt{1-\alpha_i^2}\cdot Z_i \geq t \mid B_{t,X} \text{ and } X < 4t\big)\\ &\leq \max_{x\in[t/\sigma,4t]} \Pr\left(\sqrt{1-\alpha_i^2}\cdot Z_i \geq t - \alpha_i x \mid B_{t,x}\right)\\ &\stackrel{\text{A.2.2}}{\leq} \max_{x\in[t/\sigma,4t]} \Pr\left(\sqrt{1-\alpha_i^2}\cdot Z_i \geq t - \alpha_i x\right)\\ &\leq \Pr\left(Z_i \geq (1-4\alpha_i)t\right) = \tilde{\Phi}((1-4\alpha_i)t).\end{aligned}$$

Here, first we used that $X \geq t/\sigma$ when $B_{t,X}$ happens; then we used Corollary A.2.2. Now we bound the second term in (2.11).

$$\begin{aligned}\Pr\left(X > 4t \mid B_{t,X}\right) &\leq \frac{\Pr\left(X > 4t\right)}{\Pr\left(B_{t,X}\right)} \leq \frac{\tilde{\Phi}(4t)}{\Pr\left(X \geq t/\sigma + 1\right)\Pr\left(|Y| \leq 1\right)}\\ &\leq \frac{\tilde{\Phi}(4t)}{\tilde{\Phi}(\sqrt{2}t+1)\Omega(1)} = O(e^{-8t^2+(\sqrt{2}t+1)^2/2}) = O(1/m).\end{aligned}$$

We have,

$$\mathbb{E}\left[\sum_{i=1}^m I_{\{\xi_i \geq t\}} \mid \xi \geq t \text{ and } \eta \geq t\right] \leq O(1) + \sum_{i=1}^m \tilde{\Phi}((1-4\alpha_i)t).$$

Fix a sufficiently large constant c, the number of α_i that are greater than $1/c$ is at most c^2. The number of α_i such that $\log^{-1} k \leq \alpha_i \leq 1/c$ is $O(\log^2 k)$ and for them $\tilde{\Phi}((1-4\alpha_i)t) = O(k^{-1/2})$ (since c is a sufficiently large constant). Finally, if $\alpha_i < 1/\log k$, then $\tilde{\Phi}((1-4\alpha_i)t) = O(k^{-1})$. We get the bound

$$O\left(c^2 + \frac{\log^2 k}{\sqrt{k}} + \frac{m}{k}\right) = O(1).$$

This finishes the proof. □

Lemma 2.3.12. *The function* $(1-x)^2 f_k(x)$ *is convex on the interval* $[0,1]$.

Proof. Let $m = k/\sqrt{\log k}$. Compute the first and the second derivatives of f_k:

$$\begin{aligned} f_k''(x) &= \left(m^{-\frac{2}{2-x}}\right)'' = -2\log m \cdot \left(\frac{m^{-\frac{2}{2-x}}}{(2-x)^2}\right)' \\ &= 4\log m \cdot \frac{m^{-\frac{2}{2-x}}}{(2-x)^3} \cdot \left(\frac{\log m}{2-x} - 1\right). \end{aligned}$$

Now $((1-x)^2 \cdot f_k(x))'' = (1-x)^2 \cdot f_k''(x) - 4(1-x)f_k'(x) + 2f_k(x)$. Observe that $f_k(x)$ is always positive, and $f_k'(x)$ is always negative. Therefore, if $f_k''(x)$ is positive, we are done: $((1-x)^2 \cdot f_k(x))'' \geq 0$. Otherwise, we have

$$\begin{aligned} \left((1-x)^2 \cdot f_k(x)\right)'' &= (1-x)^2 \cdot f_k''(x) - 4(1-x)f_k'(x) + 2f_k(x) \\ &\geq f_k''(x) + 2f_k(x) \geq 4\log m \cdot m^{-\frac{2}{2-x}} \left(\frac{\log m}{2} - 1\right) + 2m^{-\frac{2}{2-x}} \\ &= 2m^{-\frac{2}{2-x}}(\log m - 1)^2 \geq 0. \end{aligned}$$

□

2.4 Second Algorithm

Suppose that ε is $O(1/\log k)$. In the previous section we presented an algorithm that in this case finds an assignment of variables satisfying a constant fraction of constraints. But can we do better? In this section we show how to find an assignment satisfying $1 - O(\sqrt{\varepsilon \log k})$ fraction of constraints.

2.4.1 Algorithm

We present the algorithm in Figure 2.3. The main issue we need to take care of is to guarantee that the algorithm always picks only one element in the set S_u (otherwise we loose a constant factor). This can be done by selecting the largest in absolute value $\xi_{u_i,s}$ (at step 3.c). We will also change the way we set s_{u_i}.

Denote by $[x]_r$ the function that rounds x up or down depending on whether the fractional part of x is greater or less than r. Note that if r is a random variable uniformly distributed in the interval $[0,1]$, then the expected value of $[x]_r$ is equal to x.

We first elaborate on the difference between the choice of s_{u_i} in the algorithm above and that in Algorithm 1 presented earlier. Consider a constraint $\pi_{uv}(x_u) = x_v$. Projection $\xi_{u_i,s}$ generated by u_i and $\xi_{v_{\pi_{uv}(i)},s}$ generated by $v_{\pi_{uv}(i)}$ are considered to be *matched*. On the other hand, a projection $\xi_{u_i,s}$ such that the corresponding $\xi_{v_{\pi_{uv}(i)},s}$ does not exist (or vice versa) is considered to be *unmatched*. Unmatched projections arise when $s_{u_i} \neq s_{v_{\pi_{uv}(i)}}$ and the fraction of such projections limits the probability of satisfying the constraint. Recall that in Algorithm 1, we set $s_{u_i} = \lceil \|u_i\|^2 \cdot k \rceil$. Even if u_i and $v_{\pi_{uv}(i)}$ are infinitesimally close, it may turn out that s_{u_i} and $s_{v_{\pi_{uv}(i)}}$ differ by 1, yielding an unmatched projection. As a result, some constraints that

Input: *A solution of the SDP, with the objective value* $\varepsilon \cdot |E|$.
Output: *An assignment of variables* x_u.

1. Pick a number r in the interval $[0, 1]$ uniformly at random.
2. Pick random independent Gaussian vectors $g_1, \ldots, g_{2k}$ with independent components distributed as $\mathcal{N}(0, 1)$.
3. For each vertex u:
   (a) Set $s_{u_i} = [2k \cdot \|u_i\|^2]_r$.
   (b) For each i project s_{u_i} vectors $g_1, \ldots, g_{s_{u_i}}$ to $\tilde{u}_i$:
   $$\xi_{u_i,s} = \langle g_s, \tilde{u}_i \rangle, \ 1 \le s \le s_{u_i}.$$
   (c) Select $\xi_{u_i,s}$ with the largest absolute value, where $i \in [k]$ and $s \le s_{u_i}$. Assign $x_u = i$.

Figure 2.3: Second Algorithm for Unique Games

are almost satisfied by the SDP solution (i.e ε_{uv} is close to 0) could be satisfied with low probability (by the first rounding algorithm). In Algorithm 2, we set $s_{u_i} = [2k \cdot \|u_i\|^2]_r$. This serves two purposes: Firstly, $\mathbb{E}_r \left[|s_{u_i} - s_{v_{\pi_{uv}(i)}}| \right]$ can be bounded by $2k \cdot \|u_i - v_{\pi_{uv}(i)}\|^2$, giving a small number of unmatched projections in expectation. Secondly, the number of matched projections is always at least $k/2$. These two properties are established in Lemma 2.4.3 and ensure that the expected fraction of unmatched projections is small.

Our analysis of Rounding Algorithm 2 is based on the following theorem.

Theorem 2.4.1. *Let* $\xi_1, \cdots, \xi_m$ *and* $\eta_1, \cdots, \eta_m$ *be two sequences of standard normal random variables. Suppose that the random variables in each of the sequences are independent, the covariance of every* ξ_i *and* η_j *is nonnegative, and the average* *covariance of ξ_i and η_i is at least $1-\varepsilon$:*

$$\frac{\operatorname{cov}(\xi_1,\eta_1)+\cdots+\operatorname{cov}(\xi_m,\eta_m)}{m} \geq 1-\varepsilon.$$

Then the probability that the largest r.v. in absolute value in the first sequence has the same index as the largest r.v. in absolute value in the second sequence is $1-O(\sqrt{\varepsilon\log m})$.

Now we informally sketch the proof. We will give the complete proof in Section 2.4.2. It is instructive to consider the case when $\operatorname{cov}(\xi_i,\eta_i)=1-\varepsilon$ for all i. Assume that the first variable ξ_1 is the largest in absolute value among $\xi_1,\dots,\xi_m$ and its absolute value is a positive number t. Note that the *typical* value of t is approximately $\sqrt{2\log m-\log\log m}$ (i.e t is the $(1-1/m)$-quantile of $\mathcal{N}(0,1)$). We want to show that η_1 is the largest in absolute value among $\eta_1,\dots,\eta_m$ with probability $1-O(\sqrt{\varepsilon\log m})$, specifically that the probability that any (fixed) η_i is larger than η_1 is $O(\sqrt{\varepsilon\log m}/m)$. Let us compute this probability for η_2.

Since $\operatorname{cov}(\eta_1,\xi_1)=1-\varepsilon$ and $\operatorname{cov}(\xi_2,\eta_2)=1-\varepsilon$, the random variable η_1 is equal to $(1-\varepsilon)\xi_1+\zeta_1$; and η_2 is equal to $(1-\varepsilon)\xi_2+\zeta_2$, where ζ_1 and ζ_2 are normal random variables with variance, roughly speaking, 2ε. We need to estimate the probability of the event

$$\{\eta_2\geq\eta_1\}=\{(1-\varepsilon)\xi_2+\zeta_2\geq(1-\varepsilon)\xi_1+\zeta_1\}=\{(1-\varepsilon)\xi_2+\zeta_2-\zeta_1\geq(1-\varepsilon)t\}$$

conditional on $\xi_1=t$ and $\xi_2\leq t$. For typical t this probability is almost equal to the probability of the event:

$$\{\xi_2+\zeta\geq t \text{ and } \xi_2\leq t\}=\{t-\zeta\leq\xi_2\leq t\} \tag{2.12}$$

where $\zeta = \zeta_2 - \zeta_1$.

Since the variance of the random variable ζ is $O(\varepsilon)$, we can think that $\zeta \approx O(\sqrt{\varepsilon})$. The density of ξ_2 on the interval $[t-\zeta, t]$ is approximately $\frac{e^{-t^2/2}}{\sqrt{2\pi}} \approx O(\sqrt{\log m}/m)$ (for typical t). Thus probability (2.12) is equal to $O(\sqrt{\varepsilon \log m}/m)$. This finishes our informal "proof".

Now we are ready to prove the main lemma.

Lemma 2.4.2. *The probability that the algorithm finds an assignment of variables satisfying the constraint* $\pi_{uv}(x_u) = x_v$ *is* $1 - O(\sqrt{\varepsilon_{uv} \log k})$.

Proof. If $\varepsilon_{uv} \geq 1/8$ the statement of the lemma follows trivially. So we assume that $\varepsilon_{uv} \leq 1/8$.

Let

$$M = \Big\{(i,s) : i \in [k] \text{ and } s \leq \min(s_{u_i}, s_{v_{\pi_{uv}(i)}})\Big\};$$
$$M_c = \Big\{(i,s) : i \in [k] \text{ and } \min(s_{u_i}, s_{v_{\pi_{uv}(i)}}) < s \leq \max(s_{u_i}, s_{v_{\pi_{uv}(i)}})\Big\}.$$

The set M contains those pairs (i,s) for which both $\xi_{u_i,s}$ and $\xi_{v_{\pi_{uv}(i)},s}$ are defined (i.e the *matched* projections); the set M_c contains those pairs for which only one of the variables $\xi_{u_i,s}$ and $\xi_{v_{\pi_{uv}(i)},s}$ is defined (i.e the *unmatched* projections). We will need the following lemmas.

Lemma 2.4.3. *1. The expected size of* M_c *is at most* $4\varepsilon_{uv}k$*:*

$$\mathbb{E}\left[|M_c|\right] \leq 4\varepsilon_{uv}k.$$

2. The set M *always contains at least* $k/2$ *elements:* $|M| \geq k/2$.

Proof. 1. First we find the expected value of $|s_{u_i} - s_{v_{\pi_{uv}(i)}}|$ for a fixed i. This value is equal to

$$\mathbb{E}_r \left[\left| [2k \cdot \|u_i\|^2]_r - [2k \cdot \|v_{\pi_{uv}(i)}\|^2]_r \right| \right] = 2k \cdot \left| \|u_i\|^2 - \|v_{\pi_{uv}(i)}\|^2 \right|.$$

Now by the triangle inequality constraint (2.6),

$$2k \cdot \left| \|u_i\|^2 - \|v_{\pi_{uv}(i)}\|^2 \right| \leq 2k \cdot \|u_i - v_{\pi_{uv}(i)}\|^2.$$

Summing over all i in $[k]$ we finish the proof.

2. Observe that

$$\min(s_{u_i}, s_{v_{\pi_{uv}(i)}}) \geq 2k \cdot \min(\|u_i\|^2, \|v_{\pi_{uv}(i)}\|^2) - 1$$

and

$$\min(\|u_i\|^2, \|v_{\pi_{uv}(i)}\|^2) \geq \|u_i\|^2 - |\|u_i\|^2 - \|v_{\pi_{uv}(i)}\|^2| \geq \|u_i\|^2 - \|u_i - v_{\pi_{uv}(i)}\|^2.$$

Summing over all i we get

$$\begin{aligned} |M| &= \sum_{i \in [k]} \min(s_{u_i}, s_{v_{\pi_{uv}(i)}}) \geq \sum_{i \in [k]} \left(2k \cdot \|u_i\|^2 - 2k \cdot \|u_i - v_{\pi_{uv}(i)}\|^2 - 1\right) \\ &\geq 2k - 4k\varepsilon_{uv} - k \geq k/2. \end{aligned}$$

□

Lemma 2.4.4. *The following inequality holds:*

$$\mathbb{E}_r\left[\frac{1}{|M|}\sum_{(i,s)\in M}\varepsilon_{uv}^i\right]\leq 4\varepsilon_{uv}.$$

Proof. Recall that M always contains at least $k/2$ elements. The expected value of $\min(s_{u_i}, s_{v_{\pi_{uv}(i)}})$ is equal to $2k\cdot\min(\|u_i\|^2, \|v_{\pi_{uv}(i)}\|^2)$ and is less than or equal to $2k\cdot\mu_{uv}(i)$. Thus we have

$$\begin{aligned}\mathbb{E}_r\left[\frac{1}{|M|}\sum_{(i,s)\in M}\varepsilon_{uv}^i\right] &= \mathbb{E}_r\left[\frac{1}{|M|}\sum_{i=1}^{k}\min(s_{u_i}, s_{v_{\pi_{uv}(i)}})\cdot\varepsilon_{uv}^i\right]\\ &\leq \frac{2}{k}\sum_{i=1}^{k}2k\cdot\mu_{uv}(i)\cdot\varepsilon_{uv}^i\leq 4\sum_{i=1}^{k}\mu_{uv}(i)\cdot\varepsilon_{uv}^i\leq 4\varepsilon_{uv}.\end{aligned}$$

□

Proof of Lemma 2.4.2

Applying Theorem 2.4.1 to the sequences $\xi_{u_i,s}$ $((i,s)\in M)$ and $\xi_{v_{\pi_{uv}(i)},s}$ $((i,s)\in M)$ we get that for given r the probability that the largest in absolute value random variables in the first sequence $\xi_{u_i,s}$ and the second sequence $\xi_{v_{\pi_{uv}(i)},s}$ have the same index (i,s) is

$$1-O\left(\sqrt{\log|M|\cdot\frac{1}{|M|}\sum_{(i,s)\in M}\varepsilon_{uv}^i}\right).$$

Now by Lemma 2.4.4, and by the concavity of the function $\sqrt{x}$, we have

$$\mathbb{E}_r\left[1-O\left(\sqrt{\frac{\log|M|}{|M|}\sum_{(i,s)\in M}\varepsilon_{uv}^i}\right)\right]\geq 1-O\left(\sqrt{\varepsilon_{uv}\log k}\right).$$

The probability that there is a larger $\xi_{u_i,s}$ or $\xi_{v_{\pi_{uv}(i)},s}$ in M_c is at most

$$\mathbb{E}_r\left[\frac{|M_c|}{|M|}\right] \leq \frac{4\varepsilon_{uv}k}{k/2} = 8\varepsilon_{uv}.$$

Using the union bound we get that the probability of satisfying the constraint $\pi_{uv}(x_u) = x_v$ is at least

$$1 - O(\sqrt{\varepsilon_{uv}\log k}) - 8\varepsilon_{uv} = 1 - O(\sqrt{\varepsilon_{uv}\log k}).$$

□

Theorem 2.4.5. *There is a polynomial time algorithm that finds an assignment of variables which satisfies* $1 - O(\sqrt{\varepsilon\log k})$ *fraction of all constraints if the optimal solution satisfies* $(1-\varepsilon)$ *fraction of all constraints.*

Proof. Summing the probabilities obtained in Lemma 2.4.2 over all edges (u,v) and using the concavity of the function $\sqrt{x}$ we get that the expected number of satisfied constraints is $1 - O(\sqrt{\varepsilon\log k})|E|$. □

2.4.2 Analysis: Technical Details

In this section, we present the formal proof of Theorem 2.4.1. We will follow the informal outline of the proof sketched in Section 2.4.1. We start with estimating probability (2.12).

Lemma 2.4.6. *Let* ξ *and* ζ *be two independent random normal variables with variance 1 and* σ^2 *respectively* $(0 < \sigma < 1)$. *Then for every positive* t

$$\Pr\left(\xi \leq t \text{ and } \xi + \zeta \geq t\right) = O(\sigma e^{\frac{(\sigma t+1)^2}{2}} \cdot e^{-\frac{t^2}{2}}).$$

Remark 2.4.7. *In the "typical" case $e^{(\sigma t+1)^2/2}$ is a constant.*

Proof. We have

$$\begin{aligned}
\Pr\Big(\xi \le t \text{ and } \xi+\zeta \ge t\Big) &= \int_0^\infty \Pr\left(\xi \le t \text{ and } \xi + x \ge t\right) dF_\zeta(x) \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_0^\infty \Pr\left(\xi \le t \text{ and } \xi + x \ge t\right) e^{-\frac{x^2}{2\sigma^2}}\, dx \\
&= \frac{1}{\sqrt{2\pi}}\int_0^\infty \Pr\left(\xi \le t \text{ and } \xi + \sigma\, y \ge t\right) e^{-\frac{y^2}{2}}\, dy \\
&= \frac{1}{\sqrt{2\pi}}\int_0^{t/\sigma} \Pr\left(t-\sigma\, y \le \xi \le t\right) e^{-\frac{y^2}{2}}\, dy + \frac{1}{\sqrt{2\pi}}\int_{t/\sigma}^\infty \Pr\left(t-\sigma\, y \le \xi \le t\right) e^{-\frac{y^2}{2}}\, dy.
\end{aligned}$$

Let us bound the first integral. Since the density of the random variable ξ on the interval $(t-\sigma y, t)$ is at most $\frac{1}{\sqrt{2\pi}}e^{\frac{-(t-\sigma y)^2}{2}}$ and $y \le e^y$, we have

$$\Pr\left(t-\sigma\, y \le \xi \le t\right) \le \sigma y \cdot \frac{1}{\sqrt{2\pi}}\, e^{\frac{-(t-\sigma y)^2}{2}} \le \frac{\sigma}{\sqrt{2\pi}} \cdot e^{\frac{-t^2}{2}} \cdot e^{(\sigma t+1)y}.$$

Therefore,

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}}\int_0^{t/\sigma} \Pr\left(t-\sigma\, y \le \xi \le t\right) e^{-\frac{y^2}{2}}\, dy \;&\le\; \frac{\sigma e^{\frac{-t^2}{2}}}{2\pi}\int_0^{t/\sigma} e^{(\sigma t+1)y} \cdot e^{-\frac{y^2}{2}}\, dy \\
&\le\; \frac{\sigma e^{\frac{-t^2}{2}}}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{(y-(\sigma t+1))^2}{2}} \cdot e^{\frac{(\sigma t+1)^2}{2}}\, dy \\
&=\; O\left(\sigma e^{\frac{-t^2}{2}} \cdot e^{\frac{(\sigma t+1)^2}{2}}\right).
\end{aligned}$$

We now upper bound the second integral. If $t \geq 1$, then

$$\frac{1}{\sqrt{2\pi}} \int_{t/\sigma}^{\infty} \Pr\left(t - \sigma\, y \leq \xi \leq t\right) e^{-\frac{y^2}{2}}\, dy \leq \frac{1}{\sqrt{2\pi}} \int_{t/\sigma}^{\infty} e^{-\frac{y^2}{2}}\, dy = \tilde{\Phi}(t/\sigma)$$

$$\overset{\text{by Lemma A.1.1}}{=} O\left(\frac{e^{-\frac{t^2}{2\sigma^2}}}{t/\sigma + 1}\right) = O\left(\frac{\sigma\, e^{-\frac{t^2}{2}}}{t + \sigma}\right) = O\left(\sigma\, e^{-\frac{t^2}{2}}\right).$$

If $t \leq 1$, then

$$\frac{1}{\sqrt{2\pi}} \int_{t/\sigma}^{\infty} \Pr\left(t - \sigma\, y \leq \xi \leq t\right) e^{-\frac{y^2}{2}}\, dy \leq \int_{0}^{\infty} \frac{\sigma y \cdot e^{-\frac{y^2}{2}}}{2\pi} dy = \frac{\sigma}{2\pi} \leq \sigma e^{-\frac{t^2}{2}}.$$

The desired inequality follows from the upper bounds on the first and second integrals. □

We need a slight generalization of the lemma.

Corollary 2.4.8. *Let ξ and ζ be two independent random normal variables with variance 1 and σ^2 respectively ($0 < \sigma < 1$). Then for every $t > 0$ and $0 \leq \bar{\varepsilon} < 1$*

$$\Pr\left(\xi + \zeta \geq (1 - \bar{\varepsilon})t \;\;\middle|\;\; |\xi| \leq t\right) = O\left(\frac{(\sigma + \bar{\varepsilon}t) \cdot c(\bar{\varepsilon}, \sigma, t) \cdot e^{-t^2/2}}{1 - 2\tilde{\Phi}(t)}\right),$$

where

$$c(\bar{\varepsilon}, \sigma, t) = e^{\frac{(\sigma t + 1)^2}{2} + \bar{\varepsilon} t^2}.$$

Remark 2.4.9. *As in the previous lemma, in the "typical" case $c(\bar{\varepsilon}, \sigma, t)$ is a constant.*

Proof. First note that

$$\begin{aligned}\Pr\left(\xi+\zeta \geq (1-\bar{\varepsilon})t \quad \middle| \quad |\xi| \leq t\right) &\leq \frac{\Pr\left(\xi+\zeta \geq (1-\bar{\varepsilon})t \text{ and } \xi \leq t\right)}{\Pr\left(|\xi| \leq t\right)} \\ &= \frac{\Pr\left(\xi+\zeta \geq (1-\bar{\varepsilon})t \text{ and } \xi \leq t\right)}{1-2\tilde{\Phi}(t)}.\end{aligned}$$

Now,

$$\begin{aligned}\Pr\left(\xi+\zeta \geq (1-\bar{\varepsilon})t \text{ and } \xi \leq t\right) \leq & \Pr\left(\xi+\zeta \geq t \text{ and } \xi \leq t\right) \\ & + \Pr\left((1-\bar{\varepsilon})t \leq \xi+\zeta \leq t\right).\end{aligned}$$

By Lemma 2.4.6, the first probability is bounded as follows:

$$\Pr\left(\xi+\zeta \geq t \text{ and } \xi \leq t\right) \leq O\left(\sigma e^{\frac{(\sigma t+1)^2}{2}} \cdot e^{-\frac{t^2}{2}}\right).$$

Since $\mathrm{Var}\left[\xi+\zeta\right] \leq 1+\sigma^2$, the second probability is at most

$$\Pr\left((1-\bar{\varepsilon})t \leq \xi+\zeta \leq t\right) \quad \leq \quad \bar{\varepsilon}t \cdot e^{-\frac{((1-\bar{\varepsilon})t)^2}{2(1+\sigma^2)}} \leq \bar{\varepsilon}t \cdot e^{\frac{(2\bar{\varepsilon}+\sigma^2)t^2}{2}} \cdot e^{-\frac{t^2}{2}},$$

here we used the following inequality

$$\frac{(1-\bar{\varepsilon})^2 t^2}{2(1+\sigma^2)} = \frac{(1-\bar{\varepsilon})^2(1-\sigma^2)t^2}{2(1-\sigma^4)} \geq \frac{(1-2\bar{\varepsilon}-\sigma^2)t^2}{2} \geq \frac{t^2}{2} - \frac{(2\bar{\varepsilon}+\sigma^2)t^2}{2}.$$

The corollary follows. □

In the following lemma we formally define the random variables ζ_1 and ζ_2.

Lemma 2.4.10. *Let ξ_1, ξ_2, η_1 and η_2 be standard normal random variables such that ξ_1 and ξ_2 are independent; η_1 and η_2 are independent; and*

- $\operatorname{cov}(\xi_1, \eta_1) \geq 1 - \bar{\varepsilon} \geq 0$ *and* $\operatorname{cov}(\xi_2, \eta_2) \geq 1 - \bar{\varepsilon} \geq 0$ *(for some positive* $\bar{\varepsilon}$*);*

- $\operatorname{cov}(\xi_1, \eta_2) \geq 0$ *and* $\operatorname{cov}(\xi_2, \eta_1) \geq 0$.

Then there exist normal random variables ζ_1 *and* ζ_2 *independent of* ξ_1 *and* ξ_2 *with variance at most* $2\bar{\varepsilon}$ *such that*

$$|\eta_1| - |\eta_2| \geq (1 - 4\bar{\varepsilon})|\xi_1| - (1 + 3\bar{\varepsilon})|\xi_2| - |\zeta_1| - |\zeta_2|.$$

Proof. Express η_1 as a linear combination of ξ_1, ξ_2, and a normal r.v. ζ_1 independent of ξ_1 and ξ_2:

$$\eta_1 = \alpha_1 \xi_1 + \beta_1 \xi_2 + \zeta_1,$$

similarly,

$$\eta_2 = \alpha_2 \xi_1 + \beta_2 \xi_2 + \zeta_2.$$

Note that $\alpha_1 = \operatorname{cov}(\eta_1, \xi_1) \geq 1 - \bar{\varepsilon}$ and $\beta_1 = \operatorname{cov}(\eta_1, \xi_2) \geq 0$. Thus

$$\operatorname{Var}[\zeta_1] \leq \operatorname{Var}[\eta_1] - \alpha_1^2 \leq 1 - (1 - \bar{\varepsilon})^2 \leq 2\bar{\varepsilon}.$$

Similarly, $\alpha_2 \geq 0$, $\beta_2 \geq 1 - \bar{\varepsilon}$, and $\operatorname{Var}[\zeta_2] \leq 2\bar{\varepsilon}$. Since η_1 and η_2 are independent, we have

$$\alpha_1 \alpha_2 + \beta_1 \beta_2 + \operatorname{cov}(\zeta_1, \zeta_2) = \operatorname{cov}(\eta_1, \eta_2) = 0.$$

Therefore (note that $\operatorname{cov}(\zeta_1, \zeta_2) \leq 0$; $\alpha_1 \alpha_2 \geq 0$; $\beta_1 \beta_2 \geq 0$),

$$\alpha_2 = \frac{-\beta_1 \beta_2 - \operatorname{cov}(\zeta_1, \zeta_2)}{\alpha_1} \leq \frac{\sqrt{\operatorname{Var}[\zeta_1] \operatorname{Var}[\zeta_2]}}{1 - \bar{\varepsilon}} \leq \frac{2\bar{\varepsilon}}{1 - \bar{\varepsilon}}.$$

Taking into account that $\alpha_2 \leq 1$, we get $\alpha_2 \leq \min(1, \frac{2\bar{\varepsilon}}{1-\bar{\varepsilon}}) \leq 3\bar{\varepsilon}$. Similarly, $\beta_1 \leq 3\bar{\varepsilon}$.

Finally, we have

$$
\begin{aligned}
|\eta_1| - |\eta_2| &\geq (\alpha_1 - \alpha_2)|\xi_1| - (\beta_1 + \beta_2)|\xi_2| - |\zeta_1| - |\zeta_2| \\
&\geq (1 - 4\bar{\varepsilon})|\xi_1| - (1 + 3\bar{\varepsilon})|\xi_2| - |\zeta_1| - |\zeta_2|.
\end{aligned}
$$

□

In what follows we assume that ξ_1 is the largest r.v. in absolute value among $\xi_1, \ldots, \xi_m$ and its absolute value is t. For convenience we define three events:

$$A_t = \{|\xi_i| \leq t \text{ for all } 3 \leq i \leq m\};$$

$$E_t = A_t \cap \{|\xi_1| = t \text{ and } |\xi_2| \leq t\};$$

$$E = \{|\xi_1| \geq |\xi_i| \text{ for all } i\} = \bigcup_{t \geq 0} E_t.$$

Now we are ready to combine Corollary 2.4.8 and Lemma 2.4.10.

Lemma 2.4.11. *Let $\xi_1, \cdots, \xi_m$ and $\eta_1, \cdots, \eta_m$ be two sequences of standard normal random variables. Suppose that*

1. *the random variables in each of the sequences are independent,*
2. *the covariance of every ξ_i and η_j is nonnegative,*
3. $\operatorname{cov}(\xi_1, \eta_1) \geq 1 - \bar{\varepsilon}$ *and* $\operatorname{cov}(\xi_2, \eta_2) \geq 1 - \bar{\varepsilon}$, *where* $\bar{\varepsilon} \leq 1/7$.

Then

$$\Pr\left(|\eta_1| \leq |\eta_2| \quad | \quad E_t\right) = O\left(\frac{(\sqrt{\bar{\varepsilon}} + \bar{\varepsilon}t) \cdot e^{-t^2/2} \cdot c(7\bar{\varepsilon}, \sqrt{8\bar{\varepsilon}}, t)}{1 - 2\tilde{\Phi}(t)}\right), \qquad (2.13)$$

where $c(\bar{\varepsilon}, \sigma, t)$ is from Corollary 2.4.8.

Proof. By Lemma 2.4.10, we have

$$|\eta_1| - |\eta_2| \geq (1 - 4\bar{\varepsilon})|\xi_1| - (1 + 3\bar{\varepsilon})|\xi_2| - |\zeta_1| - |\zeta_2|.$$

Therefore,

$$\begin{aligned}
\Pr\left(|\eta_1| \leq |\eta_2| \quad | \quad E_t\right) &\leq \Pr\left((1 + 3\bar{\varepsilon})|\xi_2| + |\zeta_1| + |\zeta_2| \geq (1 - 4\bar{\varepsilon})|\xi_1| \quad | \quad E_t\right) \\
&\leq \Pr\left(|\xi_2| + |\zeta_1| + |\zeta_2| \geq (1 - 7\bar{\varepsilon})t \quad | \quad E_t\right) \\
&\leq \sum_{s, s_1, s_2 \in \{\pm 1\}} \Pr\left(s\xi_2 + s_1\zeta_1 + s_2\zeta_2 \geq (1 - 7\bar{\varepsilon})t \quad | \quad E_t\right)
\end{aligned}$$

Let us fix signs $s, s_1, s_2 \in \{\pm 1\}$ and denote $\xi = s\xi_2$, $\zeta = s_1\zeta_1 + s_2\zeta_2$, then we need to show that

$$\Pr\left(\xi + \zeta \geq (1 - 7\bar{\varepsilon})t \quad | \quad E_t\right) = O\left((\sqrt{\bar{\varepsilon}} + \bar{\varepsilon}t) \cdot \frac{e^{-t^2/2} \cdot c(7\bar{\varepsilon}, \sqrt{8\bar{\varepsilon}}, t)}{1 - 2\tilde{\Phi}(t)}\right).$$

Observe that the random variables ξ, ζ and the event A_t are independent of ξ_1, thus

$$\begin{aligned}
&\Pr\left(\xi + \zeta \geq (1 - 7\bar{\varepsilon})t \quad | \quad E_t\right) \\
&\quad = \Pr\left(\xi + \zeta \geq (1 - 7\bar{\varepsilon})\, t \quad | \quad A_t \text{ and } |\xi_1| = t \text{ and } |\xi| \leq t\right) \\
&\quad = \Pr\left(\xi + \zeta \geq (1 - 7\bar{\varepsilon})\, t \quad | \quad A_t \text{ and } |\xi| \leq t\right) \\
&\quad = \Pr\left(\zeta \geq (1 - 7\bar{\varepsilon})\, t - \xi \quad | \quad A_t \text{ and } |\xi| \leq t\right).
\end{aligned}$$

Since ξ and A_t are independent, for every fixed value of ξ we can apply Corollary A.2.2. We have

$$\begin{aligned}\Pr\left(\zeta \geq (1-7\bar{\varepsilon})t - \xi \quad \middle| \quad A_t \text{ and } |\xi| \leq t\right) &\leq \Pr\left(\zeta \geq (1-7\bar{\varepsilon})t - \xi \quad \middle| \quad |\xi| \leq t\right) \\ &= \Pr\left(\xi + \zeta \geq (1-7\bar{\varepsilon})t \quad \middle| \quad |\xi| \leq t\right).\end{aligned}$$

Finally, by Corollary 2.4.8 (where $\sigma^2 = \mathrm{Var}\left[\zeta\right] \leq 8\bar{\varepsilon}$),

$$\Pr\left(\xi + \zeta \geq (1-7\bar{\varepsilon})t \quad \middle| \quad |\xi| \leq t\right) \quad = \quad O\left(\frac{(\sqrt{\bar{\varepsilon}} + \bar{\varepsilon}t) \cdot e^{-t^2/2} \cdot c(7\bar{\varepsilon}, \sqrt{8\bar{\varepsilon}}, t)}{1 - 2\tilde{\Phi}(t)}\right).$$

□

Corollary 2.4.12. *Under assumptions of Lemma 2.4.11,*

1. if $\bar{\varepsilon}t^2 \leq 1$*, then*

$$\Pr\left(|\eta_1| \leq |\eta_2| \quad \middle| \quad E_t\right) = O\left(\sqrt{\bar{\varepsilon}}\,\frac{(t+1)\cdot\tilde{\Phi}(t)}{1 - 2\tilde{\Phi}(t)}\right);$$

2. if $t > 1$*, then*

$$\Pr\left(|\eta_1| \leq |\eta_2| \quad \middle| \quad E_t\right) = O\left(\sqrt{\bar{\varepsilon}}\right).$$

Proof. 1. If $\bar{\varepsilon}t^2 \leq 1$, then $\bar{\varepsilon}t \leq \sqrt{\bar{\varepsilon}}$ and

$$c(7\bar{\varepsilon}, \sqrt{8\bar{\varepsilon}}, t) = e^{\frac{(\sqrt{8\bar{\varepsilon}}t+1)^2}{2} + 7\bar{\varepsilon}t^2} = O(1).$$

Notice that

$$\frac{(\sqrt{\bar{\varepsilon}} + \bar{\varepsilon}t)\cdot e^{-t^2/2}}{1 - 2\tilde{\Phi}(t)} = O\left(\frac{(\sqrt{\bar{\varepsilon}} + \bar{\varepsilon}t)\cdot(t+1)\cdot\tilde{\Phi}(t)}{1 - 2\tilde{\Phi}(t)}\right),$$

since (see Lemma A.1.1)

$$\tilde{\Phi}(t) = \Theta\left(\frac{e^{-t^2/2}}{t+1}\right).$$

2. If $\bar{\varepsilon} > 1/32$ the statement holds trivially. So assume that $\bar{\varepsilon} \leq 1/32$. Then

$$\frac{(\sqrt{8\bar{\varepsilon}}t + 1)^2}{2} + 7\bar{\varepsilon}t^2 \leq \frac{3t^2}{8} + O(t).$$

Thus $t \cdot e^{-\frac{t^2}{2}} \cdot c(7\bar{\varepsilon}, \sqrt{8\bar{\varepsilon}}, t)$ is upper bounded by some absolute constant. Since $t \geq 1$, the denominator $1 - 2\tilde{\Phi}(t)$ of the expression (2.13) is bounded away from 0. □

We now give a bound on the "typical" absolute value of the largest random variable.

Lemma 2.4.13. *The following inequality holds:*

$$\Pr\left(|\xi_1| \geq 2\sqrt{\log m} \quad | \quad E\right) \leq \frac{1}{m}.$$

Proof. Note that the probability of the event E is $1/m$, since all random variables $\xi_1, \ldots, \xi_m$ are equally likely to be the largest in absolute value. Thus we have (see Lemma A.1.1)

$$\Pr\left(|\xi_1| \geq 2\sqrt{\log m} \quad | \quad E\right) \leq \frac{\Pr\left(|\xi_1| \geq 2\sqrt{\log m}\right)}{\Pr\left(E\right)} \leq \frac{1}{m^2} \bigg/ \frac{1}{m} = \frac{1}{m}.$$

□

Lemma 2.4.14. *Let $\xi_1, \cdots, \xi_m$ and $\eta_1, \cdots, \eta_m$ be two sequences of standard normal random variables as in Theorem 2.4.1. Assume that* $\operatorname{cov}(\xi_1, \eta_1) \geq 1 - \bar{\varepsilon}$ *and*

$\mathrm{cov}(\xi_2, \eta_2) \geq 1 - \bar{\varepsilon}$, *where* $\bar{\varepsilon} < \min(1/(4\log m), 1/7)$. *Then*

$$\Pr\left(|\eta_1| \leq |\eta_2| \quad | \quad E\right) = O\left(\frac{\sqrt{\bar{\varepsilon}}\log m}{m}\right).$$

Proof. Write the desired probability as follows:

$$\begin{aligned}\Pr\left(|\eta_1| \leq |\eta_2| \quad | \quad E\right) = \; & \Pr\Big(|\eta_1| \leq |\eta_2| \text{ and } |\xi_1| \leq 2\sqrt{\log m} \quad | \quad E\Big) \\ & + \Pr\Big(|\eta_1| \leq |\eta_2| \text{ and } |\xi_1| \geq 2\sqrt{\log m} \quad | \quad E\Big).\end{aligned}$$

First consider the case $|\xi_1| \leq 2\sqrt{\log m}$. Denote by $dF_{|\xi_1|}$ the density of $|\xi_1|$ conditional on E. Then

$$\begin{aligned}&\Pr\Big(|\eta_1| \leq |\eta_2| \text{ and } |\xi_1| \leq 2\sqrt{\log m} \quad | \quad E\Big) \\ &\qquad = \int_0^{2\sqrt{\log m}} \Pr\left(|\eta_1| \leq |\eta_2| \quad | \quad E \text{ and } |\xi_1| = t\right) dF_{|\xi_1|}(t) \\ &\qquad = \int_0^{2\sqrt{\log m}} \Pr\left(|\eta_1| \leq |\eta_2| \quad | \quad E_t\right) dF_{|\xi_1|}(t).\end{aligned}$$

Now by Corollary 2.4.12,

$$\int_0^{2\sqrt{\log m}} \Pr\left(|\eta_1| \leq |\eta_2| \mid E_t\right) dF_{|\xi_1|}(t) = \int_0^{2\sqrt{\log m}} O\left(\frac{2\sqrt{\bar{\varepsilon}\log m}\;\tilde{\Phi}(t)}{1 - 2\tilde{\Phi}(t)}\right) dF_{|\xi_1|}(t).$$

Let us change the variable to $x = 1-2\tilde{\Phi}(t)$. What is the probability density function of $1 - 2\tilde{\Phi}(|\xi_1|)$ given E? For each i the r.v. $1 - 2\tilde{\Phi}(|\xi_i|)$ is uniformly distributed on the interval $[0, 1]$. Now $|\xi_i| > |\xi_j|$ if and only if $1 - 2\tilde{\Phi}(|\xi_i|) > 1 - 2\tilde{\Phi}(|\xi_j|)$, therefore $1 - 2\tilde{\Phi}(|\xi_1|)$ is distributed as the maximum of m independent random variables on

$[0,1]$ given E. Its density function is $(x^m)' = mx^{m-1}$ (for $x \in [0,1]$). We have

$$\begin{aligned}\int_0^{2\sqrt{\log m}} \frac{2\sqrt{\bar{\varepsilon}\log m}\ \tilde{\Phi}(t)}{1-2\tilde{\Phi}(t)} dF_{|\xi_1|}(t) &\le \int_0^{\infty} \frac{2\sqrt{\bar{\varepsilon}\log m}\ \tilde{\Phi}(t)}{1-2\tilde{\Phi}(t)} dF_{|\xi_1|}(t) \\ &= \int_0^1 \frac{2\sqrt{\bar{\varepsilon}\log m}\cdot(1-x)/2}{x}\cdot mx^{m-1}dx = m\sqrt{\bar{\varepsilon}\log m}\int_0^1 (1-x)x^{m-2}dx \\ &= m\sqrt{\bar{\varepsilon}\log m}\left(\frac{1}{m-1}-\frac{1}{m}\right) = \frac{\sqrt{\bar{\varepsilon}\log m}}{m-1}.\end{aligned}$$

Now consider the case $|\xi_1| \ge 2\sqrt{\log m}$, by Corollary 2.4.12,

$$\Pr\left(|\eta_1| \le |\eta_2| \quad | \quad E \text{ and } |\xi_1| \ge 2\sqrt{\log m}\right) = O\left(\sqrt{\bar{\varepsilon}}\right).$$

By Lemma 2.4.13,

$$\Pr\left(|\xi_1| \ge 2\sqrt{\log m} \quad | \quad E\right) \le \frac{1}{m}.$$

This concludes the proof. □

Now we will prove a lemma, which differs from Theorem 2.4.1 only by one additional condition (4).

Lemma 2.4.15. *Let $\xi_1, \cdots, \xi_m$ and $\eta_1, \cdots, \eta_m$ be two sequences of standard normal random variables. Let $\varepsilon^i = \operatorname{cov}(\xi_i, \eta_i)$. Suppose that*

1. *the random variables in each of the sequences are independent,*

2. *the covariance of every ξ_i and η_j is nonnegative,*

3. $\frac{1}{m}\sum_{i=1}^m \varepsilon^i = \varepsilon$,

4. $\varepsilon^i \le \min(1/(4\log m), 1/7)$.

Then the probability that the largest r.v. in absolute value in the first sequence has the same index as the largest r.v. in absolute value in the second sequence is $1 - O(\sqrt{\varepsilon \log m})$.

Proof. By Lemma 2.4.14,

$$\Pr\left(|\eta_1| \le |\eta_2| \quad | \quad |\xi_1| \ge \max_{j\ge 2} |\xi_j|\right) = O\left(\frac{\sqrt{\log m}}{m}\sqrt{\max(\varepsilon^1, \varepsilon^2)}\right).$$

Applying the union bound, we get

$$\begin{aligned}
\Pr(|\eta_1| \le \max_{i\ge 2}|\eta_j| \quad | \quad |\xi_1| \ge \max_{j\ge 2}|\xi_j|) &= O\left(\frac{\sqrt{\log m}}{m}\sum_{i=2}^{m}\sqrt{\max(\varepsilon^1, \varepsilon^i)}\right)\\
&= O\left(\frac{\sqrt{\log m}}{m}\cdot\left(m\sqrt{\varepsilon^1} + \sum_{i=1}^{m}\sqrt{\varepsilon^i}\right)\right)\\
&\overset{\text{by Jensen's inequality}}{\le} O\left(\sqrt{\log m}(\sqrt{\varepsilon^1} + \sqrt{\varepsilon})\right).
\end{aligned}$$

Since the probability that $|\xi_i| = \max_j |\xi_j|$ equals $1/m$ for each i, the probability that the largest r.v. in absolute value among ξ_i, and the largest r.v. in absolute value among η_j have different indexes is at most

$$O\left(\frac{1}{m}\sum_{i=1}^{m}\sqrt{\log m}\cdot(\sqrt{\varepsilon^i} + \sqrt{\varepsilon})\right) \le O\left(\sqrt{\log m}\cdot(\sqrt{\varepsilon} + \sqrt{\varepsilon})\right) = O\left(\sqrt{\varepsilon \log m}\right).$$

□

Proof of Theorem 2.4.1. Denote $\varepsilon^i = 1 - \operatorname{cov}(\xi_i, \eta_i)$. Then $(\varepsilon^1 + \cdots + \varepsilon^m) \le m\varepsilon$. We may assume that $\varepsilon < \min(1/(4\log m), 1/7)$ — otherwise, the theorem follows trivially. Consider the set $I = \{i : \varepsilon_i < \min(1/(4\log m), 1/7)\}$. Since $\varepsilon < \min(1/(4\log m), 1/7)$, the set I is not empty. Applying Lemma 2.4.15 to random variables $\{\xi_i\}_{i\in I}$ and $\{\eta_i\}_{i\in I}$, we conclude that the largest r.v. in absolute value among $\{\xi_i\}_{i\in I}$ has the same index as the largest r.v. in absolute value among $\{\xi_i\}_{i\in I}$ with probability

$$1 - O\left(\sqrt{\log|I| \cdot \frac{1}{|I|}\sum_{i\in I}\varepsilon^i}\right) = 1 - O\left(\sqrt{\varepsilon \log m}\right).$$

Since each ξ_i is the largest r.v. among $\xi_1,\ldots,\xi_m$ in absolute value with probability $1/m$, the probability that the largest r.v. among $\xi_1,\ldots,\xi_m$ does not belong to $\{\xi_i\}_{i\in I}$ is $\frac{m-|I|}{m}$. Similarly, the probability that the largest r.v. among $\eta_1,\ldots,\eta_m$ does not belong to $\{\eta_i\}_{i\in I}$ is $\frac{m-|I|}{m}$. Therefore, by the union bound, the probability that the largest r.v. in absolute value among ξ_i, and the largest r.v. in absolute value among η_j have different indexes is at most

$$1 - O(\sqrt{\varepsilon \log m}) - 2\,\frac{m-|I|}{m}. \tag{2.14}$$

We now upper bound the last term.

$$\begin{aligned} 2\,\frac{m-|I|}{m} &\overset{\text{by the Markov inequality}}{\leq} 2\,\frac{\varepsilon}{\min(1/(4\log m), 1/7)} \\ &\leq 2\,(4\log m + 7)\varepsilon = O(\varepsilon \log m) = O(\sqrt{\varepsilon \log m}). \end{aligned}$$

(Here we use that $\varepsilon \log m < 1$.)

Plugging this bound into (2.14) we get that the desired probability is $1 - O(\sqrt{\varepsilon \log m})$. This finishes the proof. □

2.5 Third Algorithm

In this section we present our third approximation algorithm for Unique Games and prove the following theorem.

Theorem 2.5.1. *There exists a polynomial time algorithm that finds an assignment of values to vertices satisfying a* $(1 - O(\varepsilon\sqrt{\log n \log k}))$ *fraction of all constraints, for any instance of Unique Games for which a* $(1 - \varepsilon)$ *fraction of all constraints is satisfiable.*

2.5.1 Overview: Orthogonal Separators

In this section we introduce a new type of embeddings from ℓ_2^2 into ℓ_1: *embeddings separating orthogonal vectors.*

Definition 2.5.2. *A set of vectors* $X \subset \mathbb{R}^d$ *is an* ℓ_2^2 *space (or* ℓ_2 *squared space)*[2] *if for every u,v, and w in X the following inequality holds*

$$\|u - v\|^2 + \|v - w\|^2 \geq \|u - w\|^2.$$

This inequality is called an ℓ_2^2 *triangle inequality. The* ℓ_2^2*-distance between two points* u *and* v *equals* $\|u - v\|^2$.

Note that the vectors in any feasible solution to the SDP for Unique Games, together with the zero vector, form an ℓ_2^2 space.

Definition 2.5.3. *Let* X *be an* ℓ_2^2 *space. We say that a distribution over subsets of* X *is an* m*-orthogonal separator of* X *with distortion* D *and probability scale* α *if the following conditions hold for* $S \subset X$ *chosen according to this distribution:*

[2] Some authors use the term "an ℓ_2^2 space" (e.g. [4]) while others prefer an alternative term, "a space of negative type" (e.g. [3]). In this dissertation, we stick with the former term.

1. *For all u in X,* $\Pr(u \in S) = \alpha \|u\|^2$.

2. *For all orthogonal vectors u and v in X,*

$$\Pr(u \in S \text{ and } v \in S) \leq \frac{\min(\Pr(u \in S), \Pr(v \in S))}{m}.$$

   Note that the right hand side is at most $\alpha \cdot \frac{\|u\|^2 + \|v\|^2}{2m}$.

3. *For all u and v in X,*

$$\Pr(I_S(u) \neq I_S(v)) \leq \alpha D \|u - v\|^2,$$

   where I_S is the indicator function of the set S.

The novelty of Definition 2.5.3 is in property 2. It says that for every *orthogonal* vectors u and v the events "$u \in S$" and "$v \in S$" are almost disjoint. Let us state the main technical result needed for the third algorithm.

Theorem 2.5.4. *There exists a randomized polynomial time algorithm that, given an ℓ_2^2 space X containing 0 and a parameter m, returns an m-orthogonal separator of X with distortion $D = O(\sqrt{\log |X| \log m})$ and probability scale $\alpha \geq 1/poly(m)$.*

In the next section we show how using this theorem we obtain an approximation algorithm for Unique Games. We shall prove Theorem 2.5.4 in Section 2.6.

2.5.2 Approximation Algorithm

Lemma 2.5.5. *The third algorithm presented in Figure 2.4 satisfies the constraint between vertices u and v with probability $1 - O(D\varepsilon_{uv})$, where ε_{uv} is the SDP contri-*

Input: *An instance of Unique Games.*
Output: *An assignment of states to vertices.*

1. Solve the SDP.
2. Mark all vertices as unprocessed.
3. while (there are unprocessed vertices)
   (a) Produce an m-orthogonal separator S with distortion D and probability scale α as in Theorem 2.5.4, where $m = 4k$ and $D = O(\sqrt{\log n \log m})$.
   (b) For all unprocessed vertices u :
   - Let $S_u = \{i : u_i \in S\}$.
   - If S_u contains exactly one element i, then assign the state i to u, and mark the vertex u as processed.
4. If the algorithm performs more than n/α iterations, assign arbitrary values to any remaining vertices (note that $\alpha \geq 1/poly(k)$).

Figure 2.4: Third Algorithm for Unique Games

bution of the term corresponding to the edge (u, v)*:*

$$\varepsilon_{uv} = \frac{1}{2} \sum_{i=1}^{k} \|u_i - v_{\pi_{uv}(i)}\|^2.$$

Proof. If $D\varepsilon_{uv} \geq 1/8$, then the statement holds trivially, so we assume that $D\varepsilon_{uv} < 1/8$. For the sake of analysis we also assume that π_{uv} is the identity permutation (this is without loss of generality, since we can just rename the states of the vertex v).

At the end of an iteration in which at least one of the vertices u or v assigned a value we mark the constraint as satisfied or not: the constraint is satisfied, if the same state i is assigned to the vertices u and v; otherwise, the constraint is not satisfied (here we conservatively count the number of satisfied constraints: a constraint marked as not satisfied in the analysis may potentially be satisfied in the future).

Consider one iteration of the algorithm. There are three possible cases:

1. Both sets S_u and S_v are equal and contain only one element, then the constraint is satisfied.

2. The sets S_u and S_v are equal, but are empty or contain more than one element, then no values are assigned at this iteration to u and v.

3. The sets S_u and S_v are not equal, then the constraint is not satisfied (a conservative assumption).

Let us estimate the probabilities of each of these events. Using the fact that for all $i \neq j$ the vectors u_i and u_j are orthogonal, and the first and second properties of orthogonal separators we get (below α is the probability scale):

$$
\begin{aligned}
\Pr(|S_u| = 1) &\geq \sum_{i\in[k]} \Pr(i \in S_u) - \sum_{\substack{i,j\in[k]\\ i\neq j}} \Pr(i \in S_u \text{ and } j \in S_u) \\
&= \sum_{i\in[k]} \Pr(u_i \in S) - \sum_{\substack{i,j\in[k]\\ i\neq j}} \Pr(u_i \in S \text{ and } u_j \in S) \\
&\geq \sum_{i\in[k]} \alpha \|u_i\|^2 - \frac{\alpha}{m} \sum_{i,j\in[k]} \frac{\|u_i\|^2 + \|u_j\|^2}{2} \\
&= \alpha - \frac{1}{4}\alpha = \frac{3}{4}\alpha.
\end{aligned}
$$

The probability that the constraint is not satisfied is at most

$$\Pr\left(S_u \neq S_v\right) \;\leq\; \sum_{i\in[k]} \Pr\left(I_S(u_i) \neq I_S(v_i)\right) \leq \alpha D \sum_{i\in[k]} \|u_i - v_i\|^2 = 2\alpha D\varepsilon_{uv}.$$

Finally the probability of satisfying the constraint is at least

$$\Pr\left(|S_u| = 1 \text{ and } S_u = S_v\right) \geq \frac{3}{4}\alpha - 2\alpha D\varepsilon_{uv} \geq \frac{1}{2}\,\alpha.$$

Since the algorithm performs n/α iterations, the probability that it does not assign any value to u or v before step 4 is exponentially small. At each iteration the probability of failure is at most $O(D\varepsilon_{uv})$ times the probability of success, thus the probability that the constraint is not satisfied is $O(D\varepsilon_{uv})$. □

We now show that the approximation algorithm satisfies $1 - O(\varepsilon\sqrt{\log n\, \log k})$ fraction of all constraints.

Proof of Theorem 2.5.1. By Lemma 2.5.5, the expected number of unsatisfied constraints is equal to

$$\sum_{(u,v)\in E} O(D \times \varepsilon_{uv}) = O(\sqrt{\log n\, \log k}) \times SDP,$$

where SDP is the SDP value. Since $SDP \leq \varepsilon|E|$, the algorithm satisfies $1 - O(\varepsilon\sqrt{\log n\, \log k})$ fraction of all constraints with high probability. □

2.6 Producing Orthogonal Separators

In this section we present two algorithms that generate m-orthogonal separators with distortions $D_1 = O(\sqrt{\log |X|} \log m)$ and $D_2 = O(\sqrt{\log |X| \log m})$. The main difference between the algorithms is that the first algorithm uses embeddings into ℓ_1 as an intermediate step, while the second one uses embeddings into ℓ_2. Thus any improvements in embeddings into ℓ_1 will result in a better distortion for the first algorithm. We also believe that the first algorithm is simpler than the second one.

The algorithms generate orthogonal separators in three steps. First we *normalize* all vectors in a special way. Namely, we transform the set X into a set of functions in $L_2[0,\infty]$, so that the image of every non-zero vector is a function with L_2 norm 1; the images of orthogonal vectors are orthogonal; the distance between the images of u and v is roughly equal to $\|u-v\|/\max(\|u\|,\|v\|)$; and the new configuration satisfies L_2^2 triangle inequalities. This normalizes vectors while ensuring that orthogonal vectors are mapped to vectors that are far apart.

Next, we embed the transformed set into the unit sphere in ℓ_1 or ℓ_2 using slightly modified previously known algorithms. After this step, the distance between the images of any two vectors u and v is at most

$$O\left(\sqrt{\log |X|}\right) \times \frac{\|u-v\|}{\max(\|u|,\|v\|)}. \tag{2.15}$$

On the other hand the distances between the images of orthogonal vectors are larger than an absolute constant. In terms of cuts this means that any two orthogonal vectors are separated with a constant probability.

Finally, we boost the probability that orthogonal vectors are separated. Then we recover the original lengths of all vectors and get rid of the $1/\max(\|u|,\|v\|)$

term in the distortion (2.15).

2.6.1 Normalization: Embedding into $L_2[0,\infty]$

The space $L_2[0,\infty]$ is the space of square integrable functions $f : [0,\infty) \to \mathbb{R}^d$ equipped with the following inner product:

$$\langle f_1, f_2 \rangle = \int_0^{+\infty} \langle f_1(t), f_2(t) \rangle \, dt;$$

and norm:

$$\|f\|_2 = \sqrt{\langle f, f \rangle} \equiv \sqrt{\int_0^{+\infty} \|f(t)\|^2 \, dt}.$$

We construct a mapping φ from $\mathbb{R}^d$ into $L_2[0,\infty]$ as follows

$$\varphi(u)(t) = \begin{cases} u, & \text{if } t \le 1/\|u\|^2; \\ 0, & \text{otherwise.} \end{cases}$$

We map the zero vector to 0. Let us see what properties the embedding φ has.

Lemma 2.6.1. *Let $X \subset \mathbb{R}^d$ be an ℓ_2^2 metric space containing the zero vector. Then*

1. *The image $\varphi(X)$ satisfies triangle inequalities in L_2^2:*

$$\forall u, v, w \in X \;\; \|\varphi(u) - \varphi(v)\|_2^2 + \|\varphi(v) - \varphi(w)\|_2^2 \ge \|\varphi(u) - \varphi(w)\|_2^2.$$

2. *For all vectors u and v in X,*

$$\langle \varphi(u), \varphi(v) \rangle = \frac{\langle u, v \rangle}{\max(\|u\|^2, \|v\|^2)}.$$

3. For all non-zero vectors u in X, $\|\varphi(u)\|_2^2 = 1$.

4. For all orthogonal u and v in X, the images $\varphi(u)$ and $\varphi(v)$ are also orthogonal.

5. For all non-zero vectors u and v in X,

$$\|\varphi(v) - \varphi(u)\|_2^2 \leq \frac{2\,\|v-u\|^2}{\max(\|u\|^2, \|v\|^2)}.$$

Proof. 1. The triangle inequality for the functions $\varphi(u)$, $\varphi(v)$ and $\varphi(w)$ is equivalent to the following inequality:

$$\int_0^\infty \|\varphi(u)(t) - \varphi(v)(t)\|^2 + \|\varphi(v)(t) - \varphi(w)(t)\|^2 - \|\varphi(u)(t) - \varphi(w)(t)\|^2 \, dt \geq 0.$$

This inequality holds for every t, since the vectors $\varphi(u)(t)$, $\varphi(v)(t)$ and $\varphi(w)(t)$ lie in the set $\{0, u, v, w\} \subset X$ and vectors in X satisfy ℓ_2^2 triangle inequalities.

2. Without loss of generality assume that $\|u\| \leq \|v\|$, then

$$\langle \varphi(u), \varphi(v) \rangle = \int_0^\infty \langle \varphi(u)(t), \varphi(v)(t) \rangle \, dt = \int_0^{1/\|v\|^2} \langle u, v \rangle \, dt = \frac{\langle u, v \rangle}{\|v\|^2}.$$

Parts 3 and 4 follow from part 2.

5. Assume without loss of generality that $\|u\| \leq \|v\|$, then

$$\begin{aligned}
\|\varphi(v) - \varphi(u)\|_2^2 &= 2 - \langle \varphi(u), \varphi(v) \rangle \\
&= \frac{1}{\|v\|^2} \cdot \left(\|v-u\|^2 + \|v\|^2 - \|u\|^2\right) \\
&\leq \frac{2}{\|v\|^2} \cdot \left(\|v-u\|^2\right).
\end{aligned}$$

Here we used the triangle inequality $\|v - 0\|^2 \leq \|v-u\|^2 + \|u-0\|^2$. □

Remark 2.6.2. *The embedding into $L_2^2[0,\infty]$ can be represented efficiently. Note that $L_2[0,\infty]$ is a Hilbert space (and thus isometric to ℓ_2), so the metric on every finite subset of $L_2^2[0,\infty]$ is uniquely determined by its Gram matrix*[3]*. Hence we just need to compute the Gram matrix for the vectors/functions from $\varphi(X)$. This can be done using the formula from Lemma 2.6.1 (item 2). Hence, we have the following corollary.*

Corollary 2.6.3. *There exists a polynomial time algorithm that, given an ℓ_2^2 space X, computes the Gram matrix of the set of vectors $\varphi(X)$.*

2.6.2 Embedding into ℓ_1 and ℓ_2

We use the following theorem of Arora, Lee, and Naor [3], which is based on the results of Arora, Rao and Vazirani [4], Lee [25], and Chawla, Gupta and Räcke [10].

Theorem 2.6.4 ([3], Theorem 3.1)**.** *There exist constants $C \geq 1$ and $0 < p < 1/2$ such that for every finite ℓ_2^2 space X with distance $d(u,v) = \|u-v\|^2$ and every $\Delta > 0$, the following holds. There exists a distribution μ over subsets $U \subset X$ such that for every $u, v \in X$ with $d(u,v) \geq \Delta$,*

$$\mu\left\{U : u \in U \text{ and } d(v,U) \geq \frac{\Delta}{C\sqrt{\log|X|}}\right\} \geq p.$$

Note that we can efficiently sample from the distribution μ. We need the following easy corollaries.

Corollary 2.6.5. *There exists an efficient algorithm that, given an ℓ_2^2 space X, generates random subsets Y such that the following conditions hold.*

[3] The Gram matrix of a set of vectors is the matrix, where (ij)-th element is equal to the inner product of i-th and j-th vectors.

1. For every u and v in X,

$$\Pr\left(I_Y(u) \neq I_Y(v)\right) \leq D\,\|u-v\|^2.$$

2. For every u and v s.t. $\|u-v\| \geq 1$,

$$\Pr\left(I_Y(u) \neq I_Y(v)\right) \geq \beta,$$

where β is a universal constant, $D = O(\sqrt{\log|X|})$.

Proof. We apply Theorem 2.6.4 to the space X with $d(u,v) = \|u-v\|^2$ and $\Delta = 1$. Let r be a random variable uniformly distributed in $[0, \frac{1}{C\sqrt{\log|X|}}]$, where C is the constant from Theorem 2.6.4. Let Y be the r-neighborhood of U. Then

$$\begin{aligned}\Pr\left(I_Y(u) \neq I_Y(v)\right) &= \Pr\left(d(u,U) < r \leq d(v,U) \text{ or } d(v,U) < r \leq d(u,U)\right)\\ &\leq C\sqrt{\log|X|}\cdot \mathbb{E}\left[|d(u,U)-d(v,U)|\right] \leq C\sqrt{\log|X|}\cdot\|u-v\|^2.\end{aligned}$$

We verified condition 1 for $D = C\sqrt{\log|X|}$. Now if $\|u-v\|^2 \geq 1$ by Theorem 2.6.4 we have

$$\Pr\left(u \in Y, v \notin Y\right) \geq \Pr\left(u \in U \text{ and } d(v,U) \geq \frac{1}{C\sqrt{\log|X|}}\right) \geq p.$$

Therefore, $\Pr\left(I_Y(u) \neq I_Y(v)\right) = \Pr\left(u \in Y, v \notin Y\right) + \Pr\left(u \notin Y, v \in Y\right) \geq 2p$. Thus condition 2 holds for $\beta = 2p$. □

Corollary 2.6.6 (*cf.* [3], Lemma 3.5)**.** *There exists an efficient algorithm, that constructs an embedding h of an ℓ_2^2 space X into L_2 such that the following conditions* *hold.*

1. *For all u and v in X,*

$$\|h(u) - h(v)\| \leq D\,\|u - v\|^2.$$

2. *For every u and v s.t. $\|u - v\| \geq 1$,*

$$\|h(u) - h(v)\| \geq 2\gamma.$$

3. *The set $h(X)$ lies in the unit ball:*

$$\forall u \in X\ \|h(u)\| \leq 1.$$

where γ is a universal constant; $D = O(\sqrt{\log|X|})$.

Proof. We apply Theorem 2.6.4 to the space X with $d(u,v) = \|u - v\|^2$ and $\Delta = 1$. Let $L_2(\mu)$ be the space of functions from $\{U : U \subset X\}$ to $\mathbb{R}$ equipped with the inner product $\langle f, g \rangle = \mathbb{E}\left[f(U)g(U)\right]$, where U is distributed according to the distribution μ from Theorem 2.6.4. Define an embedding h of X into $L_2(\mu)$ as follows:

$$h(u) = \min(C\sqrt{\log|X|} \cdot d(u,U), 1).$$

We verify that all conditions 1–3 are satisfied.

1. We prove that the expansion of h is at most $D \equiv C\sqrt{\log |X|}$.

$$\begin{aligned}
\|h(u) - h(v)\|^2_{L_2(\mu)} &= \mathbb{E}\left[|h(u) - h(v)|^2\right] \\
&= E|\min(D \cdot d(u,U), 1) - \min(D \cdot d(v,U), 1)|^2 \\
&\leq \mathbb{E}\left[\left(D \cdot \|u - v\|^2\right)^2\right] = (D \cdot \|u - v\|^2)^2.
\end{aligned}$$

2. Now if $\|u - v\|^2 \geq 1$ by Theorem 2.6.4 we have

$$\Pr\left(u \in U \text{ and } d(v,U) \geq \frac{1}{C\sqrt{\log |X|}}\right) \geq p.$$

Therefore, $\Pr\left(h(u) = 0, h(v) = 1\right) \geq p$. Hence

$$\|h(u) - h(v)\|^2_{L_2(\mu)} \geq \Pr\left(h(u) = 0, h(v) = 1\right) + \Pr\left(h(u) = 1, h(v) = 0\right) \geq 2p.$$

We verified condition 2 for $\gamma = \sqrt{p/2}$.

3. We have

$$\|h(u)\|^2 = \mathbb{E}\left[\min(C\sqrt{\log |X|} \cdot d(u,U), 1)^2\right] \leq 1.$$

$\square$

Corollary 2.6.7. *There exists an efficient algorithm, that constructs an embedding ψ of an ℓ_2^2 space X into ℓ_2 such that the following conditions hold.*

1. *For all u and v in X,*

$$\|\psi(u) - \psi(v)\| \leq D\,\|u - v\|^2.$$

2. For every u and v s.t. $\|u - v\| \geq 1$,

$$\|\psi(u) - \psi(v)\| \geq 2\gamma.$$

3. The set $\psi(X)$ lies on the unit sphere: $\forall u \in X$

$$\|\psi(u)\| = 1,$$

where γ is a universal constant; $D = O(\sqrt{\log |X|})$.

Proof. Construct an embedding $h(u)$ from Corollary 2.6.6. We can assume that $h(u)$ is an embedding into ℓ_2 since it is isometric to L_2. Define a new embedding as follows:

$$\psi(u) = h(u)/2 + \sqrt{1 - \|h(u)\|^2/4} \cdot e,$$

where e is a unit vector orthogonal to all vectors in $h(X)$. It is easy to see that the embedding ψ satisfies conditions 2 and 3. Let us check condition 1:

$$\begin{aligned} \|\psi(u) - \psi(v)\| &\leq \|h(u) - h(v)\|/2 + |\sqrt{1 - \|h(u)\|^2/4} - \sqrt{1 - \|h(v)\|^2/4}| \\ &\leq C_1 \|h(u) - h(v)\| \leq C_1 \, D \|u - v\|^2, \end{aligned}$$

since $0 \leq \|h(u)\|/2 \leq 1/2$, and the function $\sqrt{1 - x^2}$ is a Lipschitz function on the interval $[0, 1/2]$. □

2.6.3 Generating Orthogonal Separators via ℓ_1

In this section we present an algorithm to generate orthogonal separators with distortion $O(\sqrt{\log |X|} \log m)$. This result is not as strong as the one given in the next section, but is arguably simpler, and demonstrates a number of the same ideas. Using this algorithm, in conjunction with Lemma 2.5.5, implies the following result:

Theorem 2.6.8. *There exists a polynomial time algorithm that finds an assignment of values to vertices satisfying a* $(1 - O(\varepsilon\sqrt{\log n}\log k))$ *fraction of all constraints, for any instance of Unique Games for which a* $(1-\varepsilon)$ *fraction of all constraints is satisfiable.*

The algorithm presented in Figure 2.5 generates orthogonal separators as specified above.

Lemma 2.6.9. *The algorithm generates an* m*-orthogonal separator of* X *with distortion* $O(\sqrt{\log|X|}\log m)$ *and probability scale* $\alpha = 1/|X|$.

Proof. Let us verify that all the conditions of Definition 2.5.3 hold.

1. Fix an arbitrary u. Conditional on the event $r \le \|u\|^2$ the probability of picking u in S is equal to $1/|X|$. Thus

$$\Pr(u \in S) = \frac{1}{|X|} \cdot \Pr\left(r \le \|u\|^2\right) = \frac{1}{|X|} \cdot \|u\|^2.$$

2. Fix orthogonal vectors u and v from X. By Lemma 2.6.1 (parts 3 and 4), $\|\varphi(u) - \varphi(v)\|_2^2 = 2$, hence by Corollary 2.6.5,

$$\Pr(I_{Y_i}(u) = I_{Y_i}(v)) \le 1 - \beta.$$

Thus the probability that $W(u) = W(v)$ is at most $(1-\beta)^l \le \frac{1}{m}$. The probability

Input: *An ℓ_2^2 set of vectors X (containing 0), a parameter m.*
Output: *A random set S.*

1. Set $l = \lceil \ln m/\beta \rceil$ (where β is as in Corollary 2.6.5).
2. Obtain $\varphi(X)$, a normalization of X, as described in Section 2.6.1.
3. Apply the algorithm from Corollary 2.6.5 to the set $\varphi(X)$, to generate l random independent subsets $Y_1, \ldots, Y_l \subset \varphi(X)$.
4. For every vector $u \in X$, construct a word $W(u)$ of length l corresponding to inclusion or exclusion of $\varphi(u)$ from the sets Y_i:
$$W(u) = I_{Y_1}(\varphi(u)) \ldots I_{Y_l}(\varphi(u)).$$
5. Pick a random word W in $\{0,1\}^l$ s.t. the probability that $W = W(u)$ (for each u) equals $1/|X|$. This is feasible since the number of distinct words constructed in step 4 is at most $|X|$ (possibly we may pick a word not corresponding to any $W(u)$).
6. Pick a random uniform value r in the interval $(0,1)$.
7. Find all vectors u of ℓ_2^2-length at least r such that $W(u) = W$:
$$S = \left\{ u \in X : \|u\|^2 \geq r \text{ and } W(u) = W \right\}.$$
8. Return S.

Figure 2.5: Generating Orthogonal Separators via ℓ_1

that u and v are in S is as follows:

$$\begin{aligned} \Pr(u, v \in S) &= \Pr\big(W(u) = W(v) \text{ and } W = W(u) \text{ and } r \leq \min(\|u\|^2, \|v\|^2)\big) \\ &= \Pr(W(u) = W(v)) \cdot \Pr(W = W(u)) \cdot \Pr\big(r \leq \min(\|u\|^2, \|v\|^2)\big) \\ &\leq \frac{1}{|X|} \cdot \frac{\min(\|u\|^2, \|v\|^2)}{m} = \frac{\min(\Pr(u \in S), \Pr(v \in S))}{m}. \end{aligned}$$

3. Fix u and v from X and assume $\|u\| \leq \|v\|$. Similarly to part 2, we have

$$\begin{aligned}\Pr\left(I_S(u) \neq I_S(v)\right) &= \Pr(W(u) \neq W(v) \text{ and } (W = W(u) \text{ or } W = W(v)) \\ &\quad \text{and } r \leq \|u\|^2) + \Pr\left(W = W(v) \text{ and } \|u\|^2 \leq r \leq \|v\|^2\right) \\ &\leq \frac{2\|u\|^2}{|X|} \cdot \Pr\left(W(u) \neq W(v)\right) + \frac{1}{|X|}\left(\|v\|^2 - \|u\|^2\right).\end{aligned}$$

Now, by Corollary 2.6.5 (part 1) and Lemma 2.6.1 (part 5),

$$\begin{aligned}\Pr\left(W(u) \neq W(v)\right) &\leq \sum_{i=1}^{l} \Pr\left(I_{Y_i}(\varphi(u)) \neq I_{Y_i}(\varphi(v))\right) \\ &\leq l\sqrt{\log|X|} \cdot \|\varphi(u) - \varphi(v)\|_2^2 \leq \frac{2l\sqrt{\log|X|} \cdot \|u - v\|^2}{\|v\|^2}.\end{aligned}$$

Using the ℓ_2^2 triangle inequality $\|v\|^2 - \|u\|^2 \leq \|u - v\|^2$ we get

$$\begin{aligned}\Pr\left(I_S(u) \neq I_S(v)\right) &\leq \frac{1}{|X|}\left(\frac{4\|u\|^2}{\|v\|^2} \cdot l\sqrt{\log|X|} + 1\right)\|u - v\|^2 \\ &= \frac{1}{|X|}\|u - v\|^2 \cdot O(\sqrt{\log|X|}\,\log m).\end{aligned}$$

□

2.6.4 Generating Orthogonal Separators via ℓ_2

In this section we prove Theorem 2.5.4, which in turn implies Theorem 2.5.1 (using Lemma 2.5.5). We present an algorithm to generate orthogonal separators using embeddings into ℓ_2.

Theorem 2.6.10. *The algorithm generates an m-orthogonal separator of X with distortion $O(\sqrt{\log|X|\log m})$ and probability scale $\alpha = 1/m'$.*

Input: *An ℓ_2^2 set of vectors X (containing 0), a parameter m.*
Output: *A random set S.*

1. Fix $m' = p(m)$, where $p(x)$ is a polynomial we specify later.
2. Obtain $\varphi(X)$, a normalization of X, as described in Section 2.6.1.
3. Embed $\varphi(X)$ into the unit sphere in ℓ_2 (see Corollary 2.6.7). Denote the image of the vector $\varphi(u)$ by $\psi(u)$.
4. Generate a random Gaussian vector g with independent components distributed as $\mathcal{N}(0,1)$.
5. Fix a threshold t s.t. $\Pr(\xi \geq t) = 1/m'$, where $\xi \sim \mathcal{N}(0,1)$ (i.e. t is $(1-1/m')$-quantile of the standard normal distribution).
6. Pick a random uniform value r in the interval $(0,1)$.
7. Find all vectors u of ℓ_2^2-length at least r such that $\langle \psi(u), g\rangle \geq t$:
$$S = \left\{u \in X : \|u\|^2 \geq r \text{ and } \langle \psi(u), g\rangle \geq t\right\}.$$
8. Return S.

Figure 2.6: Generating Orthogonal Separators via ℓ_2

Proof. Let us verify that all the conditions of Definition 2.5.3 hold.

1. Fix an arbitrary u. Conditional on the event $r \leq \|u\|^2$ the probability of picking u in S is equal to $1/m'$. Thus

$$\Pr(u \in S) = \frac{1}{m'} \cdot \Pr\left(r \leq \|u\|^2\right) = \frac{1}{m'} \cdot \|u\|^2.$$

2. Fix orthogonal vectors u and v from X. Similarly to Lemma 2.6.9 (part 2), we

have

$$
\begin{aligned}
&\Pr(u \in S \text{ and } v \in S) \\
&\qquad = \Pr\big(\langle \psi(u), g\rangle \geq t \text{ and } \langle \psi(v), g\rangle \geq t \text{ and } r \leq \min(\|u\|^2, \|v\|^2)\big) \\
&\qquad = \Pr\left(\langle \psi(u), g\rangle \geq t \text{ and } \langle \psi(v), g\rangle \geq t\right) \cdot \min(\|u\|^2, \|v\|^2) \\
&\qquad \leq \Pr\left(\langle (\psi(u)+\psi(v))/2, g\rangle \geq t\right) \cdot \min(\|u\|^2, \|v\|^2) \\
&\qquad = m' \Pr\left(\langle (\psi(u)+\psi(v))/2, g\rangle \geq t\right) \cdot \min(\Pr\left(u \in S\right), \Pr\left(v \in S\right)).
\end{aligned}
$$

We need to show that

$$\Pr\left(\langle (\psi(u)+\psi(v))/2, g\rangle \geq t\right) \leq 1/(m \cdot m').$$

By Lemma 2.6.1 (parts 3 and 4), $\|\varphi(u) - \varphi(v)\|_2^2 = 2$. Thus by Corollary 2.6.7, $\|\psi(u) - \psi(v)\| \geq 2\gamma$, where γ is a positive constant. Hence

$$\mathrm{Var}\left[\langle (\psi(u)+\psi(v))/2, g\rangle\right] = \left\| \frac{\psi(u)+\psi(v)}{2} \right\|^2 \leq 1 - \gamma^2.$$

Now by Lemma A.1.1 (part 3) from the Appendix (we assume that the polynomial $p(x)$ is chosen so that $m' > 6$ and thus $t > 1$)

$$
\begin{aligned}
\Pr\left(\left\langle \frac{\psi(u)+\psi(v)}{2}, g \right\rangle \geq t\right) &\leq \tilde{\Phi}\left(\frac{t}{\sqrt{1-\gamma^2}}\right) \leq \frac{1}{t}\left(C \cdot \frac{t}{m'}\right)^{\frac{1}{1-\gamma^2}} \\
&= \frac{1}{m'} \cdot O\left(\left(\frac{\log m'}{m'}\right)^{\frac{1}{1-\gamma^2}-1}\right).
\end{aligned}
$$

(here $\tilde{\Phi}(x)$ denotes the probability that a standard normal random variable is greater than x). Recall that we fixed m' to be $p(m)$, where $p(x)$ is a polynomial. It is easy to see that for an appropriate $p(x)$ (that depends only on the constant γ) the expression $O\left(\left(\frac{\log m'}{m'}\right)^{1/(1-\gamma^2)-1}\right)$ is less than $1/m$, therefore the value of the right hand side is less than $1/(m \cdot m')$.

3. Before we proceed with the proof, let us first prove the following lemma about normal variables.

Lemma 2.6.11. *Let X and Y be two standard normal random variables with covariance $1-2\varepsilon^2$; and let $\tilde{\Phi}(t) = 1/m$, $t > 1$. Then*

$$\Pr\left(X \geq t \text{ and } Y \leq t\right) = O(\varepsilon\sqrt{\log m}/m).$$

Proof. If $\varepsilon t \geq 1$ or $\varepsilon \geq 1/2$, then we are done, since $\varepsilon\sqrt{\log m} = \Omega(\varepsilon t) = \Omega(1)$ and

$$\Pr\left(X \geq t \text{ and } Y \leq t\right) \leq \Pr\left(X \geq t\right) = \frac{1}{m}.$$

So assume that $\varepsilon t \leq 1$ and $\varepsilon < 1/2$. We apply Lemma 2.4.6 to random variables Y and $X/(1-2\varepsilon^2) - Y)$. Note that the covariance of the random variables is 0, hence they are independent. Note that $\sigma^2 \equiv \mathrm{Var}[X/(1-2\varepsilon^2) - Y] = O(\varepsilon^2)$. We have

$$\begin{aligned}\Pr\left(X \geq t \text{ and } Y \leq t\right) &= \Pr\big(Y \leq t \text{ and } Y + (X/(1-2\varepsilon^2) - Y) \geq t/(1-2\varepsilon^2)\big) \\ &\leq \Pr\big(Y \leq t \text{ and } Y + (X/(1-2\varepsilon^2) - Y) \geq t\big) \\ &= O\left(\sigma e^{\frac{(\sigma t+1)^2}{2}} e^{\frac{-t^2}{2}}\right) = O(\varepsilon\tilde{\Phi}(t)t) = O(\varepsilon\sqrt{\log m}/m).\end{aligned}$$

□

Now we verify the third condition. For all u and v from X,

$$
\begin{aligned}
&\Pr(u \in S \text{ and } v \notin S) \\
&\qquad = \Pr\big(\langle \psi(u), g\rangle \geq t \text{ and } \langle \psi(v), g\rangle \leq t \text{ and } r \leq \min(\|u\|^2, \|v\|^2)\big) \\
&\qquad\quad + \Pr\big(\langle \psi(u), g\rangle \geq t \text{ and } \|v\|^2 \leq r \leq \|u\|^2\big) \\
&\qquad \leq \Pr\left(\langle \psi(u), g\rangle \geq t \text{ and } \langle \psi(v), g\rangle \leq t\right) \cdot \min(\|u\|^2, \|v\|^2) \\
&\qquad\quad + 1/m' \cdot |\|u\|^2 - \|v\|^2|.
\end{aligned}
$$

By Lemma 2.6.11,

$$
\begin{aligned}
\Pr\left(() \, \langle \psi(u), g\rangle \geq t \text{ and } \langle \psi(v), g\rangle \leq t\right) &\leq O(\|\psi(v) - \psi(u)\| \sqrt{\log m'}/m') \\
&\leq O\left(\|\varphi(v) - \varphi(u)\|_2^2 \cdot \sqrt{\log |X|} \cdot \sqrt{\log m'}/m'\right) \\
&\leq O\left(\frac{\|v-u\|^2}{\max(\|u\|^2, \|v\|^2)} \cdot \sqrt{\log |X|} \cdot \sqrt{\log m'}/m'\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\Pr(I_S(u) \neq I_S(v)) &= \Pr\left(u \in S \text{ and } v \notin S\right) + \Pr\left(u \notin S \text{ and } v \in S\right) \\
&\leq O\left(\frac{\|v-u\|^2}{\max(\|u\|^2, \|v\|^2)} \cdot \sqrt{\log |X|} \cdot \sqrt{\log m'}/m'\right) \\
&\qquad \times \min(\|u\|^2, \|v\|^2) + 2/m' \cdot |\|u\|^2 - \|v\|^2| \\
&\leq O\left(\|v-u\|^2 \sqrt{\log |X|} \cdot \sqrt{\log m'}/m'\right).
\end{aligned}
$$

This completes the proof. □

2.7 d to 1 Games

In this section we extend our results to d-to-1 games.

Definition 2.7.1. *We say that* $\Pi \subset [k] \times [k]$ *is a d-to-1 predicate, if for every* i *there are at most* d *different values* j *such that* $(i, j) \in \Pi$*, and for every* j *there is at most one* i *such that* $(i, j) \in \Pi$*.*

Definition 2.7.2 (d to 1 Games)**.** *We are given a directed constraint graph* $G = (V, E)$*, a set of variables* x_u *(for all vertices* u*) and d-to-1 predicates* $\Pi_{uv} \subset [k] \times [k]$ *for all edges* (u, v)*. Our goal is to assign a value from the set* $[k]$ *to each variable* x_u*, so that the maximum number of the constraints* $(x_u, x_v) \in \Pi_{uv}$ *is satisfied.*

Note that even if all constraints of a d-to-1 game are satisfiable it is hard to find an assignment of variables satisfying all constraints. We will show how to satisfy

$$\Omega\left(\frac{1}{\sqrt{\log k}} \cdot (1-\varepsilon)^4 \cdot \left(\frac{k}{\sqrt{\log k}}\right)^{-\frac{\sqrt{d}-1+\varepsilon}{\sqrt{d}+1-\varepsilon}}\right)$$

fraction of all constraints (the multiplicative constant in the Ω notation depends on d). Notice that this value can be obtained by replacing ε in formula (1.1) with $\varepsilon' = 1 - (1-\varepsilon)/\sqrt{d}$ (and changing $(1-\varepsilon)^2$ to $(1-\varepsilon)^4$).

Even though we do not require that for a constraint Π_{uv} each i in $[k]$ belongs to some pair $(i, j) \in \Pi_{uv}$, let us assume that for each i there exists j s.t. $(i, j) \in \Pi_{uv}$; and for each j there exists i s.t. $(i, j) \in \Pi_{uv}$. As we see later this assumption is not important.

In order to write a relaxation for d-to-1 games, we introduce the following notation:

$$w_{uv}^i = \sum_{j:(i,j)\in\Pi_{uv}} v_j.$$

The SDP is as follows:

$$\text{minimize } \frac{1}{2}\sum_{(u,v)\in E}\left(\sum_{i=1}^{k} \| \; u_i - w_{uv}^i \; \|^2\right)$$

subject to

$$\forall u \in V \; \forall i,j \in [k], i \neq j \qquad \langle u_i, u_j\rangle = 0 \tag{2.16}$$

$$\forall u \in V \qquad \sum_{i=1}^{k} \|u_i\|^2 = 1 \tag{2.17}$$

$$\forall (u,v) \in V \; i,j \in [k] \qquad \langle u_i, v_j\rangle \geq 0 \tag{2.18}$$

$$\forall (u,v) \in V \; i \in [k] \; 0 \leq \langle u_i, w_{uv}^i\rangle \leq \min(\|u_i\|^2, \|w_{uv}^i\|^2) \tag{2.19}$$

An important observation is that $\|w_{uv}^1\|^2 + \ldots + \|w_{uv}^k\|^2 = 1$, here we use the fact that for a fixed edge (u,v) each v_j is a summand in one and only one w_{uv}^i.

We use the first algorithm for Unique Games (described in Section 2.3, see Figure 2.2) for rounding a vector solution. For analysis we will need to change some notation:

$$\tilde{w}_{uv}^i = \begin{cases} w_{uv}^i/\|w_{uv}^i\|, \text{ if } w_{uv}^i \neq 0; \\ 0, \text{ otherwise} \end{cases}$$

$$\varepsilon_{uv}^i = \frac{\|\tilde{u}_i - \tilde{w}_{uv}^i\|^2}{2}$$

$$\varepsilon_{uv}^i{}' = 1 - \frac{1-\varepsilon_{uv}^i}{\sqrt{d}}$$

$$\mu_{uv}(i) = \frac{\|u_i\|^2 + \|w_{uv}^i\|^2}{2}$$

The following lemma explains why we get the new dependency on ε.

Lemma 2.7.3. *For every edge* (u, v) *and state* i *there exists* j' *s.t.* $(i, j') \in \Pi_{uv}$ *and* $\|\tilde{u}_i - \tilde{v}_{j'}\|^2/2 \leq {\varepsilon_{uv}^i}'$.

Proof. Let u_i' be the projection of the vector $\tilde{u}_i$ to the linear span of the vectors v_j (where $(i, j) \in \Pi_{uv}$). Let α_i be the angle between $\tilde{u}_i$ and w_{uv}^i; and let β_i be the angle between $\tilde{u}_i$ and u_i'. Clearly, $\|u_i'\| = \cos \beta_i \geq \cos \alpha_i = 1 - \varepsilon_{uv}^i$. Since all $\tilde{v}_j$ ($(i, j) \in \Pi_{uv}$) are orthogonal unit vectors, there exists $\tilde{v}_{j'}$ s.t. $\langle \tilde{v}_{j'}, u_i' \rangle \geq \|u_i'\|/\sqrt{d}$. Hence, $\langle \tilde{v}_{j'}, \tilde{u}_i \rangle = \langle \tilde{v}_{j'}, u_i' \rangle \geq (1 - \varepsilon_{uv}^i)/\sqrt{d}$. □

For every edge (u, v) and state i, find j' as in the previous lemma and define a function[4] $\pi_{uv}(i) = j'$. Then replace every constraint $(x_u, x_v) \in \Pi_{uv}$ with a stronger constraint $\pi_{uv}(x_u) = x_v$. Now we can apply the original analysis of Algorithm 1 to the new problem. In the proof we need to substitute ${\varepsilon_{uv}^i}'$ for ε_{uv}^i, $1 - (1 - \varepsilon_{uv})/\sqrt{d}$ for ε_{uv}, and $1 - (1 - \varepsilon)/\sqrt{d}$ for ε. The only missing step is the following lemma.

Lemma 2.3.7$'$. *For every edge* (u, v) *the following statements hold.*

1. *The average value of* ε_{uv}^i *w.r.t. the measure* μ_{uv} *is less than or equal to* ε_{uv}.

2. *The average value of* ${\varepsilon_{uv}^i}'$ *w.r.t. the measure* μ_{uv} *is less than or equal to* $1 - \frac{1-\varepsilon_{uv}}{\sqrt{d}}$.

3. $\min(s_{u_i}, s_{v_{\pi_{uv}(i)}}) \geq d(1 - {\varepsilon_{uv}^i}')^4 \mu_{uv}(i) k$.

Proof. Let α_i be the angle between u_i and w_{uv}^i and let α_i' be the angle between u_i and $v_{\pi_{uv}(i)}$.

[4]The function π_{uv} is not necessarily a permutation.

1. Indeed,

$$\begin{aligned}\sum_{i=1}^{k} \mu_{uv}(i) \cdot \varepsilon_{uv}^{i} &= \sum_{i=1}^{k} \frac{\|u_i\|^2 + \|w_{uv}^i\|^2 - (\|u_i\|^2 + \|w_{uv}^i\|^2) \cdot \cos \alpha_i}{2} \\ &\leq \sum_{i=1}^{k} \frac{\|u_i\|^2 + \|w_{uv}^i\|^2 - 2 \cdot \|u_i\| \cdot \|w_{uv}^i\| \cdot \cos \alpha_i}{2} \\ &= \sum_{i=1}^{k} \frac{\|u_i - w_{uv}^i\|^2}{2} = \varepsilon_{uv}.\end{aligned}$$

2. This follows from part 1 and the definition of ${\varepsilon_{uv}^{i}}'$.

3. Due to the triangle inequality constraint, $\|w_{uv}^i\| \cos \alpha_i \leq \|u_i\|$. Thus

$$(1 - \varepsilon_{uv}^{i})^2 \mu_{uv}(i) = \cos^2 \alpha_i \cdot \frac{\|u_i\|^2 + \|w_{uv}^i\|^2}{2} \leq \|u_i\|^2.$$

Similarly $\|v_{\pi_{uv}(i)}\| \cos \alpha_i' \leq \|u_i\|$ and

$$(1 - {\varepsilon_{uv}^{i}}')^2 \|u_i\|^2 \leq \cos^2 \alpha_i' \cdot \|u_i\|^2 \leq \|v_{\pi_{uv}(i)}\|^2.$$

Combining these two inequalities and noting that $(1 - {\varepsilon_{uv}^{i}}') = (1 - \varepsilon_{uv}^{i})/\sqrt{d}$, we get

$$d(1 - {\varepsilon_{uv}^{i}}')^4 \mu_{uv}(i) \leq (1 - {\varepsilon_{uv}^{i}}')^2 \|u_i\|^2 \leq \|v_{\pi_{uv}(i)}\|^2.$$

The lemma follows. □

We now address the issue that for some edges (u, v) and states j there may not necessarily exist i s.t. $(i, j) \in \Pi_{uv}$. We call such j a state of degree 0. The key observation is that in our algorithms we may enforce additional constraints like $x_u = i$ or $x_u \neq i$ by setting $u_i = 1$ or $u_i = 0$ respectfully. Thus we can add extra states and enforce that the vertices are not in these states. Then we add pairs (i, j) where i is a new state, and j is a state of degree 0 (or vice-versa). Alternatively we can rewrite the objective function by adding an extra term:

$$\text{minimize } \frac{1}{2} \sum_{(u,v)\in E} \left(\sum_{i=1}^{k} \left\| \, u_i - w_{uv}^i \, \right\|^2 + \|w_{uv}^0\|^2 \right),$$

where w_{uv}^0 is the sum of v_j over j of degree 0.

Chapter 3

MAX 2 CSP

3.1 Semidefinite Relaxation

In this section, we describe the vector program (SDP) for MAX 2CSP/MAX 2SAT. For convenience we replace each negation $\bar{x}_i$ with a new variable x_{-i} that is equal by the definition to $\bar{x}_i$. First, we transform our problem to a MAX 2SAT formula: we replace

- each constraint of the form $x_i \wedge x_j$ with two clauses x_i and x_j;
- each constraint of the form $x_i \oplus x_j$ with two clauses $x_i \vee x_j$ and $x_{-i} \vee x_{-j}$;
- finally, each constraint x_i with $x_i \vee x_i$.

It is easy to see that the fraction of *unsatisfied* constraints in the formula is equal, up to a factor of 2, to the number of unsatisfied constraints in the original MAX 2CSP instance. Therefore, if we satisfy $1 - O(\min(\sqrt{\varepsilon}, \varepsilon\sqrt{\log n}))$ fraction of all constraints in the 2SAT formula, we will also satisfy $1 - O(\min(\sqrt{\varepsilon}, \varepsilon\sqrt{\log n}))$ fraction of all constraints in MAX 2CSP. In what follows, we will consider only 2SAT formulas.

To avoid confusion between 2SAT and SDP constraints we will refer to them as clauses and constraints respectively.

We now rewrite all clauses in the form $x_i \to x_j$, where $i, j \in \{\pm 1, \pm 2, \ldots, \pm n\}$. For each x_i, we introduce a vector variable v_i in the SDP. We also define a special unit vector v_0 that corresponds to the value 1: in the intended (integral) solution $v_i = v_0$, if $x_i = 1$; and $v_i = -v_0$, if $x_i = 0$. The SDP contains the constraints that all vectors are unit vectors; v_i and v_{-i} are opposite; and ℓ_2^2-triangle inequalities.

For each clause $x_i \to x_j$ we add the term

$$\frac{1}{8}\left(\|v_j - v_i\|^2 - 2\langle v_j - v_i, v_0\rangle\right)$$

to the objective function. In the intended solution this expression equals to 1, if the clause is not satisfied; and 0, if it is satisfied. Therefore, our SDP is a relaxation of MAX 2SAT (the objective function measures how many clauses are not satisfied).

We get an SDP relaxation for MAX 2SAT:

$$\text{minimize } \frac{1}{8} \sum_{\text{clauses } x_i \to x_j} \|v_j - v_i\|^2 - 2\langle v_j - v_i, v_0\rangle$$

subject to

$$\begin{aligned}
\|v_i - v_j\|^2 + \|v_j - v_k\|^2 \geq \|v_i - v_k\|^2 \qquad & \text{for all } i, j, k \in \{0, \pm 1, \ldots, \pm n\} \\
\|v_i\|^2 = 1 \qquad & \text{for all } i \in \{0, \pm 1, \ldots, \pm n\} \\
v_i = -v_{-i} \qquad & \text{for all } i \in \{\pm 1, \ldots, \pm n\}
\end{aligned}$$

In a slightly different form, this semidefinite program was introduced by Feige and Goemans [14]. Later, Zwick [34] used this SDP in his algorithm.

Note that since all vectors v_i are unit vectors, triangle inequalities can be written as

$$\|v_i - v_j\|^2 + \langle v_i - v_j, v_k \rangle \geq 0.$$

In particular, triangle inequalities imply that the contribution of each clause $x_i \to x_j$ to the SDP objective function is nonnegative:

$$\|v_j - v_i\|^2 - 2\langle v_j - v_i, v_0 \rangle \geq 0 \qquad \text{for each clause } x_i \to x_j. \tag{3.1}$$

In fact, our first algorithm (which satisfies $1 - O(\sqrt{\varepsilon})$ fraction of all constraints) requires only triangle inequalities of type (3.1).

3.2 First Algorithm

In this section, we present an approximation algorithm that finds a solution satisfying $1 - O(\sqrt{\varepsilon})$ fraction of all constraints given a $1 - \varepsilon$ satisfiable instance. The approximation algorithm is shown in Figure 3.1. We interpret the inner product $\langle v_i, v_0 \rangle$ as the bias towards rounding v_i to 1. The algorithm rounds vectors orthogonal to v_0 ("unbiased" vectors) using the random hyperplane technique. If, however, the inner product $\langle v_i, v_0 \rangle$ is positive, the algorithm shifts the random hyperplane; and it is more likely to round v_i to 1 than to 0.

It is easy to see that the algorithm always obtains a valid assignment to variables: if $x_i = 1$, then $x_{-i} = 0$ and vice versa.

A clause $x_i \to x_j$ is not satisfied by the algorithm if $\xi_i \geq t_i$ and $\xi_j \leq t_j$ (i.e. x_i

Input: *An instance of MAX 2SAT.*
Output: *An assignment of variables* x_i.

1. Solve the SDP for MAX 2SAT. Denote by SDP the objective value of the solution and by ε the fraction of the constraints "unsatisfied" by the vector solution, that is,
$$\varepsilon = \frac{SDP}{\#\text{constraints}}.$$
2. Pick a random Gaussian vector g with independent components distributed as $\mathcal{N}(0,1)$.
3. For every i,
   (a) Project the vector g to v_i:
   $$\xi_i = \langle g, v_i \rangle.$$
   Note, that ξ_i is a standard normal random variable, since v_i is a unit vector.
   (b) Pick a threshold t_i as follows:
   $$t_i = -\langle v_i, v_0 \rangle \big/ \sqrt{\varepsilon}\ .$$
   (c) If $\xi_i \geq t_i$, set $x_i = 1$, otherwise set $x_i = 0$.

Figure 3.1: First Algorithm for MAX 2SAT

is set to 1; and x_j is set to 0). The following lemma bounds the probability of this event.

Lemma 3.2.1. *Let ξ_i and ξ_j be two standard normal random variables with covariance $1 - 2\Delta^2$ (where $\Delta \geq 0$). For all real numbers t_i, t_j and $\delta = (t_j - t_i)/2$ we have (for some absolute constant C)*

1. If $t_j \le t_i$,

$$\Pr(\xi_i \ge t_i \textit{ and } \xi_j \le t_j) \le C\min(\Delta^2/|\delta|, \Delta).$$

2. If $t_j \ge t_i$,

$$\Pr(\xi_i \ge t_i \textit{ and } \xi_j \le t_j) \le C(\Delta + 2\delta).$$

Proof. 1. First note that if $\Delta = 0$, then the above inequality holds, since $\xi_i = \xi_j$ almost surely. If $\Delta \ge 1/2$, then the right hand side of the inequality becomes $\Omega(1) \times \min(1/|\delta|, 1)$. Since $\max(t_i, -t_j) \ge |\delta|/2$, the inequality follows from the bound $\tilde{\Phi}(|\delta|/2) \le O(1/|\delta|)$. So we assume $0 < \Delta < 1/2$.

Let $\xi = (\xi_i + \xi_j)/2$ and $\eta = (\xi_i - \xi_j)/2$. Notice that $\mathrm{Var}[\xi] = 1 - \Delta^2$, $\mathrm{Var}[\eta] = \Delta^2$; and random variables ξ and η are independent. We estimate the desired probability as follows:

$$\begin{aligned}\Pr(\xi_j \le t_j \text{ and } \xi_i \ge t_i) &= \Pr\left(\eta \ge \left|\xi - \frac{t_i + t_j}{2}\right| + \frac{t_i - t_j}{2}\right) \\ &= \int_{-\infty}^{+\infty} \Pr\left(\eta \ge \left|t - \frac{t_i + t_j}{2}\right| + \frac{t_i - t_j}{2}\right) dF_\xi(t).\end{aligned}$$

Note that the density of the normal distribution with variance $1 - \Delta^2$ is always less than $1/\sqrt{2\pi(1 - \Delta^2)} < 1$, thus we can replace $dF_\xi(t)$ with dt.

$$
\begin{aligned}
\Pr(\xi_j \le t_j \text{ and } \xi_i \ge t_i) &\le \int_{-\infty}^{+\infty} \tilde{\Phi}\left(\frac{\left|t - \frac{t_i+t_j}{2}\right| + \frac{t_i-t_j}{2}}{\Delta}\right) dt \\
&= \int_{-\infty}^{+\infty} \tilde{\Phi}\left(\frac{|t| + |\delta|}{\Delta}\right) dt \\
&= \Delta \int_{-\infty}^{+\infty} \tilde{\Phi}\left(|s| + |\delta|/\Delta\right) ds \quad \text{(by Lemma A.1.1)} \\
&\le C'\Delta \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(|s|+|\delta|/\Delta)^2}{2}} ds \\
&= 2C'\Delta \cdot \tilde{\Phi}(|\delta|/\Delta) \quad \text{(by Lemma A.1.1)} \\
&\le 2C' \min(\Delta^2/|\delta|, \Delta).
\end{aligned}
$$

2. We have

$$
\begin{aligned}
\Pr(\xi_j \le t_j \text{ and } \xi_i \ge t_i) &\le \Pr(\xi_j \le t_j \text{ and } \xi_i \ge t_j) + \Pr(t_i \le \xi_i \le t_j) \\
&\le C(\Delta + 2\delta).
\end{aligned}
$$

For estimating the probability $\Pr(\xi_j \le t_j \text{ and } \xi_i \ge t_j)$ we used part 1 with $t_i = t_j$. □

Theorem 3.2.2. *The approximation algorithm finds an assignment satisfying a $1 - O(\sqrt{\varepsilon})$ fraction of all constraints, if a $1 - \varepsilon$ fraction of all constraints is satisfied in the optimal solution.*

Proof. We shall estimate the probability of satisfying a clause $x_i \to x_j$. Set $\Delta_{ij} = \|v_j - v_i\|/2$ (so that $\operatorname{cov}(\xi_i, \xi_j) = \langle v_i, v_j\rangle = 1 - 2\Delta_{ij}^2$) and $\delta_{ij} = (t_j - t_i)/2 \equiv -\langle v_j - v_i, v_0\rangle/(2\sqrt{\varepsilon})$. The contribution of the term to the SDP is equal to $c_{ij} = (\Delta_{ij}^2 + \delta_{ij}\sqrt{\varepsilon})/2$.

Consider the following cases (we use Lemma 3.2.1 in all of them):

1. If $\delta_{ij} \geq 0$, then the probability that the clause is not satisfied (i.e. $\xi_i \geq t_i$ and $x_j \leq t_j$) is at most

$$C(\Delta_{ij} + 2\delta_{ij}) \leq C(\sqrt{2c_{ij}} + 4c_{ij}/\sqrt{\varepsilon}).$$

2. If $\delta_{ij} < 0$ and $\Delta_{ij}^2 \leq 4c_{ij}$, then the probability that the clause is not satisfied is at most

$$C\Delta_{ij} \leq 2C\sqrt{c_{ij}}.$$

3. If $\delta_{ij} < 0$ and $\Delta_{ij}^2 > 4c_{ij}$, then the probability that the constraint is not satisfied is at most

$$\frac{C\Delta_{ij}^2}{|\delta_{ij}|} = \frac{C\Delta_{ij}^2}{(\Delta_{ij}^2 - 2c_{ij})/\sqrt{\varepsilon}} \leq \frac{C\sqrt{\varepsilon}\,\Delta_{ij}^2}{\Delta_{ij}^2 - \Delta_{ij}^2/2} = 2C\sqrt{\varepsilon}.$$

Combining these cases we get that the probability that the clause is not satisfied is at most

$$4C(\sqrt{c_{ij}} + c_{ij}/\sqrt{\varepsilon} + \sqrt{\varepsilon}).$$

The expected fraction of unsatisfied clauses is equal to the average of such probabilities over all clauses. Recall, that ε is equal, by the definition, to the average value of c_{ij}. Therefore, the expected number of unsatisfied constraints is $O(\sqrt{\varepsilon} + \varepsilon/\sqrt{\varepsilon} + \sqrt{\varepsilon})$ (here we used Jensen's inequality for the function $\sqrt{\cdot}$). $\square$

3.3 Second Algorithm

3.3.1 Preliminaries

For every instance of MAX 2SAT we consider a corresponding directed graph $G = (V, E)$: the set of vertices equals $V = \{\pm 1, \pm 2, \ldots, \pm n\}$; two vertices i and j are connected by a directed edge (i, j) if there is a clause $x_i \rightarrow x_j$ or a clause $x_{-j} \rightarrow x_{-i}$ (note that these two clauses are equivalent) in the MAX 2SAT instance. The graph G is symmetric in the following sense: if (i, j) is an edge then $(-j, -i)$ is an edge as well.

Definition 3.3.1. *Given a set of vertices S, denote the set of edges outgoing from S by $\delta^{out}(S)$; the set of edges incoming to S by $\delta^{in}(S)$. Denote by $\delta^{out}_M(S)$ [$\delta^{in}_M(S)$] the set of edges outgoing from [incoming to] S in the subgraph $G[M]$ of G induced by a vertex set M.*

Consider a feasible solution $\{v_i\}_{i \in V}$ to the semidefinite relaxation for MAX 2SAT. We introduce a directed distance function $d : V \times V \rightarrow \mathbb{R}^+ \cup \{0\}$ as follows

$$d(i, j) = \|v_i - v_j\|^2 - \langle v_0, v_j - v_i \rangle \equiv \|v_i - v_j\|^2 - \|v_0 - v_i\|^2 + \|v_0 - v_j\|^2.$$

Lemma 3.3.2. *The function $d(\cdot,\cdot)$ satisfies the following four properties:*

$$\begin{aligned} &\forall i \in X && d(i,i) = 0 \\ &\forall i,j \in X && d(i,j) \geq 0 \\ &\forall i,j,k \in X && d(i,j) + d(j,k) \geq d(i,k) \\ &\forall i,j \in X && d(i,j) = d(-j,-i). \end{aligned}$$

Remark 3.3.3. *We will call a distance function that satisfies the first three properties* a directed semimetric[1]*; if additionally a function satisfies the fourth property we will call it* symmetric directed semimetric.

Proof. Clearly, $d(i,i) = 0$. The fact that d is nonnegative trivially follows from the triangle inequality constraint of the semidefinite relaxation. Now we verify the third condition (the triangle inequality for d):

$$\begin{aligned} d(i,j) + d(j,k) &= \|v_i - v_j\|^2 - \|v_0 - v_i\|^2 + \|v_0 - v_j\|^2 \\ &\quad + \|v_j - v_k\|^2 - \|v_0 - v_j\|^2 + \|v_0 - v_k\|^2 \\ &= \|v_i - v_j\|^2 + \|v_j - v_k\|^2 - \|v_0 - v_i\|^2 + \|v_0 - v_k\|^2 \\ &\geq \|v_i - v_j\|^2 - \|v_0 - v_i\|^2 + \|v_0 - v_k\|^2 = d(i,k) \end{aligned}$$

□

We define the distance between sets and points in the natural way:

- $d(S,T) = \min_{i \in S; j \in T} d(i,j)$;
- $d(i,S) = d(\{i\},S)$;

[1] Directed semimetrics are sometimes called quasi-semimetrics.

- $d(S,i) = d(S,\{i\})$;

Definition 3.3.4. *Given a set of vertices $M \subset V$, we define its volume as follows:*

$$\text{vol}(M) = \sum_{\substack{(i,j)\in E\\ i,j\in M}} d(i,j).$$

Definition 3.3.5. *We say that two sets of vertices, S and T, are Δ-separated with respect to d if $d(S,T) \geq \Delta$. Similarly, we say that two sets of vectors in Euclidean space, S and T, are Δ-separated with respect to ℓ_2^2 if for every $u \in S$ and $v \in T$*

$$\|u - v\|^2 \geq \Delta.$$

3.3.2 Separation Theorem

In this section, we remind the reader the separation theorem of Arora, Rao and Vazirani [4]. Then we prove a variant of this theorem for the directed semimetric d.

We will use the algorithm of Arora, Rao, Vazirani [4] with the separation guarantee of Lee [25].

Theorem 3.3.6 (Arora, Rao, Vazirani [4], Lee [25])**.** *Consider a set X of points in Euclidean space. Assume that it satisfies ℓ_2^2 triangle inequalities:*

$$\forall u, v, w \in X \; \|u - v\|^2 + \|v - w\|^2 \geq \|u - w\|^2,$$

and that it is spread:

$$\sum_{u,v\in X} \|u - v\|^2 \geq c|X|^2,$$

where c is a constant that does not depend on X. Then there exist two sets $S, T \subset X$

that are $\Omega(1/\sqrt{\log n})$-separated with respect to ℓ_2^2 and such that each of them contains a constant fraction of all points in X.

Furthermore, there is a randomized polynomial-time algorithm for finding these subsets S, T.

Note that a feasible SDP solution $\{v_i\}$ for MAX 2SAT is spread. Indeed,

$$
\begin{aligned}
&\sum_{i,j\in V} \|v_i - v_j\|^2 \\
&\qquad = \sum_{\substack{i\in V,\ j\in V\\ i>0,\ j>0}} \left(\|v_i - v_j\|^2 + \|v_i - v_{-j}\|^2 + \|v_{-i} - v_j\|^2 + \|v_{-i} - v_{-j}\|^2\right) \\
&\qquad = \sum_{\substack{i\in V,\ j\in V\\ i>0,\ j>0}} 8 = 8n^2 = 2|V|^2.
\end{aligned}
$$

Therefore, we can apply Theorem 3.3.6 to it. Moreover, we may assume that the algorithm returns sets S and T that are symmetric: $S = -T$ (Indeed, the first step of the ARV algorithm partitions V into symmetric sets S' and T': $S' = -T'$. At the deletion step we have some freedom in choosing matchings. We should always choose symmetric matchings, that is if (i, j) belongs to the matching, $(-i, -j)$ should also belong to the matching. See also [25, Corollary 4.10].).

Corollary 3.3.7. *Let $\{v_i\}_{i\in V}$ be a feasible solution of the semidefinite relaxation for MAX 2SAT. Then there exists a set $S \subset V$ such that $|S| = \Omega(n)$ and sets $\{v_i : i \in S\}$ and $\{-v_i : i \in S\}$ are $\Omega(1/\sqrt{\log n})$-separated with respect to ℓ_2^2.*

Furthermore, there is a randomized polynomial-time algorithm for finding S.

Now we prove a variant of this corollary for weighted graphs. Our proof is based on the approach of Chawla, Gupta and Räcke [10]: we introduce many duplicate vertices for every vertex of the weighted graph; the number of duplicates is (roughly) proportional to the weight of the vertex.

Assume that every vertex i has a weight w_i and that $w_i = w_{-i}$. Denote the weight of all vertices in a set M by $W(M)$.

Corollary 3.3.8. *There exists a set $S \subset V$ such that $W(S) = \Omega(W(V))$ and sets $\{v_i : i \in S\}$ and $\{-v_i : i \in S\}$ are $\Omega(1/\sqrt{\log n})$-separated with respect to ℓ_2^2. Furthermore, there is a randomized polynomial-time algorithm for finding S.*

Proof. Denote $W = W(V)$. For every vertex i, we introduce

$$m_i = \left\lfloor w_i \middle/ \frac{W}{n^2} \right\rfloor.$$

duplicate vertices $(i, 1), \ldots, (i, m_i)$ and let $v_{(i,j)} = v_i$. Clearly, the set $\{v_{(i,j)}\}$ satisfies the SDP constraints since $v_{(i,j)} = v_{(-i,j)}$ and all vectors $v_{(i,j)}$ lie in the set $\{v_i\}$. Since $m_i \le n^2$ there are at most $2n^3$ new vertices.

We apply Corollary 3.3.7 to vectors $\{v_{(i,j)}\}$. We obtain a set S_{dup} such that sets $\{v_{i,j} \equiv v_i : (i,j) \in S_{dup}\}$ and $\{-v_{i,j} \equiv -v_i : (i,j) \in S_{dup}\}$ are Δ-separated with respect to ℓ_2^2 distance, and S_{dup} contains a constant fraction of duplicate vertices. Let S be the set of vertices i such that at least one duplicate of i belongs to S_{dup}. First, sets $\{v_i : i \in S\}$ and $\{-v_i : i \in S\}$ are Δ-separated. Then

$$W(S) = \sum_{i \in S} w_i \ge \sum_{i \in S} \frac{m_i W}{n^2} = \frac{W}{n^2} \sum_{i \in S} m_i \ge \frac{W}{n^2} |S_{dup}|.$$

On the other hand,

$$|S_{dup}| = \Omega\left(\sum_{i \in V} m_i\right) \ge \Omega\left(\sum_{i \in V} \left(\frac{w_i n^2}{W} - 1\right)\right) = \Omega\left(n^2 - |V|\right) = \Omega(n^2).$$

Therefore, $W(S) = \Omega(W)$. $\square$

We are ready to prove a separation theorem for the directed semimetric $d(\cdot, \cdot)$.

Theorem 3.3.9. *There exists a polynomial-time randomized algorithm that given a graph $G = (V, E)$, a feasible solution v_i, and a set of non-negative weights $\{w_i\}_{i \in V}$ ($w_i = w_{-i}$) finds a set $S \subset V$ such that*

- *sets S and $-S \equiv \{-i : i \in S\}$ are Δ-separated with respect to d;*
- *$W(S) = \Omega(W(V))$.*

Proof. We apply the algorithm from Corollary 3.3.8. The algorithm returns a set of vertices $\hat{S}$ such that the sets $\left\{v_i : i \in \hat{S}\right\}$ and $\left\{-v_i : i \in \hat{S}\right\}$ are $\Omega(1/\sqrt{\log n})$-separated with respect to ℓ_2^2. Let

$$S^+ = \{i \in \hat{S} : \langle v_0, v_i \rangle \ \geq 0\}$$
$$S^- = \{i \in \hat{S} : \langle v_0, v_i \rangle \ \leq 0\}.$$

Note that sets S^+ and $-S^- \equiv \{-i : i \in S^-\}$ are $\Omega(1/\sqrt{\log n})$-separated with respect to d: If $i \in S^+$, $j \in -S^-$, then

$$d(i, j) = \|v_i - v_j\|^2 + 2\langle v_0, v_i - v_j \rangle \geq \|v_i - v_j\|^2 \geq \Omega(1/\sqrt{\log n}).$$

Similarly sets $-S^-$ and S^+ are $\Omega(1/\sqrt{\log n})$-separated. Since $\hat{S} = S^+ \cup S^-$, one of the sets S^+ or S^- contains a constant fraction of the total weight. Therefore, either $S = S^+$ or $S = -S^-$ satisfies the condition of the theorem. $\square$

Input: *A directed graph $G = (V, E)$; a feasible solution $\{v_i\}_{i \in V}$.*
Output: *A partitioning of V into three sets S, R, and $-S$.*

1. Let
$$w_i = \sum_{j:(i,j)\in E} d(i,j) + \sum_{j:(j,i)\in E} d(j,i).$$
(Note that $w_i = w_{-i}$.)

2. Run the separation algorithm from Theorem 3.3.9 with weights w_i. Denote its output by S^*. Denote the distance between sets S^* and $-S^*$ by Δ.

3. Define a level cut E_t ($t \in (0, \Delta)$):
$$E_t = \{(i,j) \in E : d(S^*, i) \le t \text{ and } d(S^*, j) \ge t\} \cup \{(i,j) \in E : d(j, -S^*) \le t \text{ and } d(i, -S^*) \ge t\}.$$

4. Find $t_0 \in (0, \Delta/2)$ which minimizes the size of E_t.

5. Let S be the set of vertices that are reachable from S^* in the graph $G - E_{t_0}$.

6. Let $R = M \setminus (S \cup -S)$.

7. Return S, R and $-S$.

Figure 3.2: Finding $(S, R, -S)$ Partitioning

3.3.3 Algorithm and Analysis

Theorem 3.3.10. *Given a directed graph $G = (V, E)$ such that $V = -V$ and a feasible solution $\{v_i\}_{i \in V}$, the algorithm presented in Figure 3.2 finds a partitioning of V into three sets S, R, and $-S$ such that*

1.
$$|\delta^{out}(S)| + |\delta^{in}(-S)| = O\left(\sqrt{\log |V|}\,\mathrm{vol}(M)\right).$$

2. The volume of R is at most a constant fraction of the volume of the graph:

$$\mathrm{vol}(R) \le c\ \mathrm{vol}(M), \ \textit{where } c < 1 \textit{ is a constant.}$$

Proof. First, we prove that sets S and $-S$ are disjoint. Note that for every $i \in S$, $d(S^*, i) < \Delta/2$; similarly, for every $j \in -S$, $d(j, -S^*) < \Delta/2$. Therefore, for every $i \in S$, $j \in -S$, we have

$$d(i,j) \ge d(S^*, -S^*) - d(S^*, i) - d(j, -S^*) > 0.$$

Hence $i \ne j$.

By the definition of S, $\delta^{out}(S) \subset E_{t_0}$; $\delta^{in}(-S) \subset E_{t_0}$. Using standard arguments we get

$$\begin{aligned} \mathrm{vol}(V) &= \sum_{(i,j)\in E} d(i,j) \ge \sum_{\substack{(i,j)\in E \\ d(S^*,j)\ge d(S^*,i)}} \Big(d(S^*, j) - d(S^*, i) \Big) \\ &= \sum_{\substack{(i,j)\in E \\ d(S^*,j)\ge d(S^*,i)}} \int_{d(S^*,i)}^{d(S^*,j)} dt \ge \int_0^{\Delta/2} |E_t| dt. \end{aligned}$$

hence $|E_{t_0}| \le \frac{2}{\Delta} \mathrm{vol}(V)$ and

$$|\delta^{out}(S)| + |\delta^{in}(-S)| \le \frac{2}{\Delta} \mathrm{vol}(M) = O\left(\sqrt{\log |V|}\, \mathrm{vol}(V) \right).$$

Now let us estimate the volume of R. Since $W(S) = \Omega(W(V))$, $W(R) \le cW(G)$

for some constant $c < 1$. On the other hand,

$$\text{vol}(R) = \sum_{\substack{(i,j)\in E\\ i,j\in R}} d(i,j) \le \frac{1}{2}\sum_{i\in R} w_i = \frac{W(R)}{2};$$
$$\text{vol}(V) = \frac{1}{2}\sum_{i\in V} w_i = \frac{W(V)}{2}.$$

Therefore, $\text{vol}(R) \le cW(G)$. □

Applying this algorithm recursively, we get an algorithm for the MAX 2SAT problem.

Input: *An instance of MAX 2SAT.*
Output: *An assignment of variables x_i.*

1. Let $G = (V, E)$ be the graph corresponding to the MAX 2SAT instance.
2. Solve the semidefinite relaxation for MAX 2SAT. Obtain a solution $\{v_i\}_{i\in V}$.
3. Let $k = 0$, $R_0 = V$.
4. **while** R_k is not empty
   (a) Find $(S, R, -S)$ partitioning of $G[R_k]$ (the graph induced on G by R_i).
   (b) Let $S_{k+1} = S$, $R_{k+1} = R$, $k = k + 1$.
5. Let $S = S_1 \cup \cdots \cup S_k$.
6. For every $i \in V$, let
$$x_i = \begin{cases} true, & \text{if } i \in S;\\ false, & \text{otherwise.}\end{cases}$$

Figure 3.3: Second Algorithm for MAX 2SAT

Theorem 3.3.11. *Given an instance of MAX 2SAT where $1 - \varepsilon$ fraction of all* *constraints are satisfiable, the algorithm presented in Figure 3.3 finds a solution that satisfies a* $1 - O(\varepsilon\sqrt{\log n})$ *fraction of all constraints. The algorithm runs in randomized polynomial time.*

Proof. Note that at each iteration k, V is a disjoint union of sets $S_1 \cup \ldots S_k$, R_k, and $-(S_1 \cup \ldots S_k)$. Since when the algorithm stops $R_k = \varnothing$, sets S and $-S$ partition V. Therefore, our assignment to x_i is valid: if x_i is true then x_{-i} is false and vice versa.

Note that a clause $x_i \rightarrow x_j$ is not satisfied if and only if $i \in S$ and $j \in -S$, that is, when the edge (i, j) goes from S to $-S$. We use the following estimate on the number of such edges:

$$\begin{aligned}&|\delta^{out}(S_1)| + |\delta^{in}(-S_1)| + |\delta^{out}_{R_1}(S_2)| + |\delta^{in}_{R_1}(-S_2)| + \ldots \\ &\qquad = O\left(\sqrt{\log n}\right) \cdot (\mathrm{vol}(V) + \mathrm{vol}(R_1) + \ldots).\end{aligned}$$

The key observation is that the volume of R_i decreases geometrically, so the number of unsatisfied clauses is $O(\sqrt{\log n}) \cdot \mathrm{vol}(V)$. □

Chapter 4

Conclusion and Future Work

4.1 Conclusions

In this dissertation, we presented approximation algorithms for two Constraint Satisfaction Problems, Unique Games and MAX 2CSP. Our algorithms have better approximation guarantees than previously known algorithms in all ranges of parameters. Moreover, three our algorithms are nearly optimal (assuming the Unique Games Conjecture).

Our results for Unique Games have interesting complexity implications. They show that when the fraction of unsatisfiable constraints ε is less than $c/\log k$ almost all constraints can be satisfied in polynomial time, which exponentially improves the previous bound of $\sim 1/k^{10}$. In particular, our bound disproves some stronger versions of the Unique Games Conjecture considered in the literature before.

We developed new techniques for rounding SDP solutions. In particular, in Section 2.3, we showed how to deal with SDP solutions in which different vectors have different lengths; in Sections 2.3.2 and 2.4.2, we proved several useful bounds for the joint distribution of normal variables; in Section 2.6, we introduced and constructed a new type of metric embeddings, m-orthogonal separators; finally, in Section 3.3.2, we proved a separation theorem for directed metric spaces. These techniques are of interest on their own. We hope that they will prove useful for solving other combinatorial problems.

4.2 Future Work

Many known approximation results and several results presented in this dissertation are optimal or near optimal assuming the Unique Games Conjecture. It is therefore a very interesting question whether the Unique Conjecture is true, and these results are indeed optimal; or whether it is false, and we just need more powerful methods to obtain better approximation guarantees, and thus disprove the Unique Games Conjecture.

One promising avenue for future research is to study stronger SDP relaxations (e.g. relaxations from the Lasserre hierarchy) for combinatorial optimization problems such as MAX CUT, Vertex Cover, and Unique Games. Of course, it would be great if this research led to improved approximation guarantees to these problems. However, even negative results, that is, integrality gap examples for Lasserre hierarchy, would be interesting, and they would be a strong evidence in favor of the Unique Games Conjecture.

Bibliography

[1] A. Agarwal, M. Charikar, K. Makarychev, and Y. Makarychev. $O(\sqrt{\log n})$ *approximation algorithms for Min UnCut, Min 2CNF Deletion, and directed cut problems.* In Proceedings of the 37th ACM Symposium on Theory of Computing, pp. 573–581, 2005.

[2] G. Andersson, L. Engebretsen, and J. Håstad. *A new way to use semidefinite programming with applications to linear equations mod p.* Journal of Algorithms, Vol. 39, pp. 162–204, 2001.

[3] S. Arora, J. R. Lee, and A. Naor. *Euclidean distortion and the sparsest cut.* In Proceedings of the 37th ACM Symposium on Theory of Computing, pp. 553–562, 2005.

[4] S. Arora, S. Rao, and U. Vazirani. *Expander flows, geometric embeddings and graph partitioning.* In Proceedings of the 36th ACM Symposium on Theory of Computing, pp. 222–231, 2004.

[5] P. Austrin *Balanced Max 2-Sat might not be the hardest.* In Proceedings of the 39th ACM Symposium on Theory of Computing, pp. 189–197, 2007.

[6] P. Austrin *Towards Sharp Inapproximability For Any 2-CSP.* In Proceedings of the 48th IEEE Symposium on Foundations of Computer Science, pp. 307–317,2007.

[7] M. Charikar, K. Makarychev, and Y. Makarychev. *Near-Optimal Algorithms for Maximum Constraint Satisfaction Problems.* In Proceedings of the 18th ACM-SIAM Symposium on Discrete Algorithms, pp. 62–68, 2007.

[8] M. Charikar, K. Makarychev, and Y. Makarychev. *Near-Optimal Algorithms for Unique Games.* In Proceedings of the 38th ACM Symposium on Theory of Computing, pp. 205–214, 2006.

[9] E. Chlamtac, K. Makarychev, and Y. Makarychev. *How to Play Unique Games Using Embeddings.* In Proceedings of the 47th IEEE Symposium on Foundations of Computer Science, pp. 687–696, 2006.

[10] S. Chawla, A. Gupta, and H. Räcke. *Approximations for generalized sparsest cut and embeddings of ℓ_2 into ℓ_1.* In Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithms, pp. 102–111, 2005.

[11] S. Chawla, R. Krauthgamer, R. Kumar,Y. Rabani, and D. Sivakumar. *On the hardness of approximating multicut and sparsest-cut.* In Proceedings of the 20th IEEE Conference on Computational Complexity, pp. 144–153, 2005.

[12] I. Dinur, E. Mossel, and O. Regev. *Conditional hardness for approximate coloring.* ECCC Technical Report TR05-039, 2005.

[13] U. Feige and D. Reichman. *On systems of linear equations with two variables per equation.* In Proceedings of the 7th International Workshop on Approxi-

mation Algorithms for Combinatorial Optimization, vol. 3122 of Lecture Notes in Computer Science, pp. 117–127, 2004.

[14] U. Feige and M. Goemans. *Approximating the value of two prover proof systems, with applications to MAX 2SAT and MAX DICUT.* In Proceedings of the 3rd IEEE Israel Symposium on the Theory of Computing and Systems, pp. 182–189, 1995.

[15] U. Feige and L. Lovász. *Two-prover one round proof systems: Their power and their problems.* In Proceedings of the 24th ACM Symposium on Theory of Computing, pp. 733–741, 1992.

[16] N. Garg, V. V. Vazirani, and M. Yannakakis. *Approximate max-flow min-(multi)cut theorems and their applications.* In Proceedings of the 25th ACM Symposium on Theory of Computing, pp. 698–707, 1993.

[17] M. Goemans and D. Williamson. *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming.* Journal of the ACM, vol. 42, no. 6, pp. 1115–1145, Nov. 1995.

[18] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization.* Springer Verlag, 1988.

[19] A. Gupta and K. Talwar. *Approximating Unique Games.* In Proceedings of the 17th ACM-SIAM Symposium on Discrete Algorithms, pp. 99–106, 2006.

[20] S. Khot. *On the power of unique 2-prover 1-round games.* In Proceedings of the 34th ACM Symposium on Theory of Computing, pp. 767–775, 2002.

[21] S. Khot and O. Regev. *Vertex cover might be hard to approximate to within* $2-\varepsilon$. In Proceedings of the 18th IEEE Conference on Computational Complexity, pp. 379–386, 2003.

[22] S. Khot and N. 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, pp. 53–62, 2005.

[23] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. *Optimal inapproximability results for MAX-CUT and other 2-variable CSPs?* ECCC Report TR05-101, 2005.

[24] P. Klein, S. Plotkin, S. Rao, and E. Tardos. *Approximation algorithms for Steiner and directed multicuts.* Journal of Algorithms, vol. 22 n. 2, pp. 241–269, 1997.

[25] J. R. Lee. *On distance scales, embeddings, and efficient relaxations of the cut cone.* In Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithms, pp. 92–101, 2005.

[26] M. Lewin, D. Livnat, and U. Zwick. *Improved Rounding Techniques for the MAX 2-SAT and MAX DI-CUT Problems.* In Proceedings of the 9th International IPCO Conference on Integer Programming and Combinatorial Optimization, pp. 67–82, 2002.

[27] S. Matuura and T. Matsui. *New approximation algorithms for MAX 2SAT and MAX DICUT.* Journal of Operations Research Society of Japan, vol. 46, pp. 178–188, 2003.

[28] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. *Noise stability of functions with low influences: invariance and optimality.* In Proceedings of the 46th IEEE Symposium on Foundations of Computer Science, pp. 21–40, 2005.

[29] A. Samorodnitsky and L. Trevisan. *Gowers Uniformity, Influence of Variables, and PCPs.* In Proceedings of the 38th ACM symposium on Theory of computing, pp. 11–20, 2006.

[30] Z. Šidák. *Rectangular Confidence Regions for the Means of Multivariate Normal Distributions.* Journal of the American Statistical Association, vol. 62, no. 318, pp. 626–633, Jun. 1967.

[31] L. Stockmeyer. *Planar 3-colorability is polynomial complete.* ACM SIGACT News, Vol. 5, Issue 3, pp. 19–25, July 1973.

[32] L. Trevisan. *Approximation Algorithms for Unique Games.* In Proceedings of the 46th IEEE Symposium on Foundations of Computer Science, pp. 197–205, 2005.

[33] U. Zwick. *Analyzing the MAX 2-SAT and MAX DI-CUT approximation algorithms of Feige and Goemans.* Manuscript.
Available at `www.cs.tau.ac.il/~zwick/my-online-papers.html`.

[34] U. Zwick. *Finding almost satisfying assignments.* In Proceedings of the 30th ACM Symposium on Theory of Computing, pp. 551–560, 1998.

Appendix A

Properties of Normal Distribution

A.1 Bounds on Normal Distribution Function

For completeness we prove some standard results about the normal distribution in this section. Denote the probability that a standard normal random variable is greater than $t \in \mathbb{R}$ by $\tilde{\Phi}(t)$, in other words

$$\tilde{\Phi}(t) \equiv 1 - \Phi_{0,1}(t) = \Phi_{0,1}(-t),$$

where $\Phi_{0,1}$ is the normal distribution function.

Lemma A.1.1. *1. For every $t > 0$,*

$$\frac{t}{\sqrt{2\pi}(t^2+1)} e^{-\frac{t^2}{2}} < \tilde{\Phi}(t) < \frac{1}{\sqrt{2\pi}t} e^{-\frac{t^2}{2}}.$$

2. There exist constants c_1, C_1, c_2, C_2 such that for every $t > 0$ and $0 < p < 1/3$

the following bounds hold

$$\frac{c_1}{\sqrt{2\pi}(t+1)}e^{-\frac{t^2}{2}} \le \tilde{\Phi}(t) \le \frac{C_1}{\sqrt{2\pi}(t+1)}e^{-\frac{t^2}{2}};$$

$$c_2\sqrt{\log(1/p)} \le \tilde{\Phi}^{-1}(p) \le C_2\sqrt{\log(1/p)};$$

3. *There exist constants c_3 and C_3 such that for every $t > 1$, $\rho \ge 1$ the following inequality holds*

$$(c_3 t\tilde{\Phi}(t))^{\rho^2} \le t\tilde{\Phi}(\rho t + \frac{1}{\rho t}) \le t\tilde{\Phi}(\rho t) \le (C_3 t\tilde{\Phi}(t))^{\rho^2}$$

Proof. 1. Observe, that in the limit $t \to \infty$ all three expressions are equal to 0. Hence the lemma follows from the following inequalities on the derivatives:

$$\left(\frac{t}{\sqrt{2\pi}(t^2+1)}e^{-\frac{t^2}{2}}\right)' = -\frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} + \frac{2e^{-\frac{t^2}{2}}}{\sqrt{2\pi}(t^2+1)^2} > -\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}} = \tilde{\Phi}(t)',$$
$$\left(\frac{1}{\sqrt{2\pi}t}e^{-\frac{t^2}{2}}\right)' = -\frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} - \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}t^2} < -\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}} = \tilde{\Phi}(t)'$$

2. This trivially follows from (1).

3. Using(2) we get

$$t\tilde{\Phi}(\rho t) \le \frac{C_1 t e^{-\frac{(\rho t)^2}{2}}}{\rho t + 1} = C_1\left(\frac{t+1}{(\rho t+1)^{1/\rho^2}t^{1-1/\rho^2}} \cdot \frac{te^{-t^2/2}}{t+1}\right)^{\rho^2}$$
$$\le C_1\left(\left(\frac{t+1}{t}\right)^{1-1/\rho^2}\left(t\cdot\tilde{\Phi}(t)\right)\right)^{\rho^2} \le \left(2C_1\, t\tilde{\Phi}(t)\right)^{\rho^2}.$$

Similarly,

$$t\tilde{\Phi}(\rho t + \frac{1}{\rho t}) \geq \frac{c_1 t e^{-\frac{(\rho t + 1/(\rho t))^2}{2}}}{\rho t + 1} = c_1 \left(\frac{(t+1)e^{-1/\rho^2 - 1/(2t^2)}}{(\rho t + 1)^{1/\rho^2} t^{1-1/\rho^2}} \cdot \frac{t e^{-t^2/2}}{t+1} \right)^{\rho^2}$$

$$\geq c_1 \left(\frac{1}{e^{3/2}\rho^{1/\rho^2}} \left(\frac{t+1}{t} \right)^{1-1/\rho^2} \left(t \cdot \tilde{\Phi}(t) \right) \right)^{\rho^2} \geq \left(\frac{c_1}{e^2} t\tilde{\Phi}(t) \right)^{\rho^2}.$$

Here we used that $\rho^{1/\rho^2} \leq e^{1/(2e)} < \sqrt{e}$. □

A.2 Šidák Theorem

In Section 2.3.2 and Section 2.4.2, we use a corollary of the following result of Šidák [30]:

Theorem A.2.1 (Šidák)**.** *Let $\xi_1, \ldots, \xi_k$ be normal random variables with mean zero, then for any positive $t_1, \ldots, t_k$,*

$$\Pr\left(|\xi_1| \leq t_1, |\xi_2| \leq t_2, \ldots, |\xi_k| \leq t_k\right) \geq \Pr\left(|\xi_1| \leq t_1\right) \Pr\left(|\xi_2| \leq t_2, \ldots, |\xi_k| \leq t_k\right).$$

Note that these random variable do not have to be independent.

Corollary A.2.2. *Let $\xi_1, \ldots, \xi_k$ be normal random variables with mean zero, then for any positive $t_1, \ldots, t_k$,*

$$\Pr\left(\xi_1 \geq t_1 \;\middle|\; |\xi_2| \leq t_2, \ldots, |\xi_k| \leq t_k\right) \leq \Pr\left(\xi_1 \geq t_1\right).$$

Proof. By Theorem A.2.1,

$$\Pr\left(|\xi_1| \leq t_1 \;\middle|\; |\xi_2| \leq t_2, \ldots, |\xi_k| \leq t_k\right) \geq \Pr\left(|\xi_1| \leq t_1\right).$$

Thus

$$\begin{aligned}
&\Pr(\xi_1 \geq t_1 \mid |\xi_2| \leq t_2, \ldots, |\xi_k| \leq t_k) \\
&\qquad = \frac{1}{2} - \frac{1}{2} \Pr\left(|\xi_1| \leq t_1 \mid |\xi_2| \leq t_2, \ldots, |\xi_k| \leq t_k\right) \\
&\qquad \leq \frac{1}{2} - \frac{1}{2} \Pr\left(|\xi_1| \leq t_1\right) = \Pr\left(\xi_1 \geq t_1\right).
\end{aligned}$$

□