Title
Merging Techniques for Combinatorial Optimization: Spectral Graph Theory and Semidefinite Programming

Author
Kolla, Alexandra

Publication Date
2009

Peer reviewed|Thesis/dissertation

Merging Techniques for Combinatorial Optimization: Spectral Graph Theory and Semidefinite Programming

by

Alexandra Kolla

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

in

Computer Science

in the

GRADUATE DIVISION
of the
UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor Umesh Vazirani, Chair
Professor Luca Trevisan
Professor Alberto Grünbaum

Fall 2009

Merging Techniques for Combinatorial Optimization: Spectral Graph Theory and Semidefinite Programming

Abstract

Merging Techniques for Combinatorial Optimization: Spectral Graph Theory and Semidefinite Programming

by

Alexandra Kolla
Doctor of Philosophy in Computer Science

University of California at Berkeley

Professor Umesh Vazirani, Chair

In this thesis, we study three problems related to expanders, whose analysis involves understanding the intimate connection between expanders, spectra and semidefinite programming. Our first result is related to Khot's *Unique Games* conjecture (UGC) [**?**], whose validity is one of the most central open problems in computational complexity theory. We show that UGC is *false* on expander graphs. This result, in particular, rules out a natural way of proving hardness of approximation for SPARSEST CUT. Our second result is in the area of *graph sparsification*. We say that a graph H is a *sparsifier* for a graph G if the respective graph Laplacians of the two graphs satisfy $x^T\mathcal{L}_H x \approx x^T\mathcal{L}_G x$ for all vectors x. Given a union of two graphs $G+W$, we show how to choose a sparse subgraph $W' \subseteq W$ so that $G+W'$ is a good sparsifier for $G+W$. We apply the result to optimizing the algebraic connectivity of a graph by adding very few edges. We also show how to use this result in order to create optimal *ultrasparsifiers* for every graph, which can be used as good graph-theoretic *preconditioners* for symmetric, positive semidefinite, diagonally dominant linear systems. Lastly, we study the integrality gap of the well known Sparsest Cut semidefinite program. We present a simple construction and analysis of an $\Omega(\log\log N)$ integrality gap for Sparsest Cut while our gap instance, vector solution, and analysis are somewhat simpler and more intuitive than those which appear in the literature.

Our techniques use tools both from the area of semidefinite programming and spectral graph theory. This is not surprising since, when delving into the beautiful theory underlying spectra and SDPs, one finds that they are deeply connected in more ways than one. Semidefinite programs are nothing but linear programs with variables representing entries of a matrix, together with eigenvalue bounds for that matrix and could, therefore capture the spectral expansion of a graph. Similarly, most of eigenvalue optimization problems can be cast as SDPs, which leads to developing semidefinite programming based algorithms for a plethora of other important graph problems. In this thesis we further explore the connections between expansion, spectra and SDPs by applying them to solving these three problems described above.

To my grandmother

Contents

Acknowledgements

Being a graduate student at Berkeley has been for me an exciting journey through routes that sometimes have been extremely rewarding and other times have been long and difficult but definitely "different" that what I expected when I first came in. Looking back at those last five years I feel that I would never have been exposed to all the wonderful things I learned and experienced without a certain number of people being by my side.

I am deeply thankful to my thesis advisor, Umesh Vazirani. The advice he gave me throughout the years, helped me shape into who I have become not only as an academic but as a person as well. From him, I learned how to follow my dreams and not be afraid: "You just need to find your eigenvector" he said when I was confused and uncertain about my future as a graduate student. I learned how to meet drop-dead deadlines: "If you only have one minute till the deadline, then everything gets done in the last minute". Sure enough, he also made a great contribution to my writing skills. "Your paper looks great, fantastic job. Now you need to re-write it!" or, some times, if I was lucky enough he would say: "Your thesis looks great, it is so good that you might even need to re-write it". Most importantly, the greatest lesson I got from him is that the secret to success is not so much trying to fix my weaknesses and change who I am in the process but rather embrace them and work more on my strengths so that I can follow my own path and develop myself as my own independent researcher.

I am extremely grateful to my undergraduate thesis advisor, Spiros Argyros. Apart from being an exceptional mathematician and inspiring teacher he is also a very selfless, modest person. To him I owe a very large part of my love for mathematics. His most precious piece of advise to me was never to worry about how successful I am but just do what I love and love what I do and success will come. By following his advise throughout my academic life I feel I have become a much more happier, relaxed and successful researcher. Many thanks to Luca, Satish and Alberto Grünbaum for serving in my thesis and quals committee and giving me useful advice throughout my years in Berkeley. Also thanks to Christos for convincing me to choose to come to Berkeley when I was trying to decide which graduate school to go to. For the technical side of the thesis, thanks to my coauthors Sanjeev Arora, Subash Khot, Madhur Tulsiani, David Steurer, Nisheeth Vishnoi, Yury Makarychev, Shanghua Teng, Amin Saberi, James Lee, for working with me on the problems I was interested on and for their valuable contribution. I would like to especially thank Madhur, Yury and Shanghua for being such brilliant, very supportive and enthusiastic coworkers and friends. Many thanks to my office-mate and friend Henry Lin for helping me out with every small technical detail that I couldn't fix, and for posting very helpful notes on what needs to be done to file a thesis. Thanks also to several people that invited me to work with them at Microsoft and gave me the opportunity to experience a great academic environment and meet brilliant scientists: Yuval Peres, Jennifer Chayes and Christian Borgs.

My experience in Berkeley would have been much shorter if it wasn't for the great waves that California coast has to offer. It would also have been much less fun and relaxing without the surfing. Thanks to all my surfing friends and to all the people in the Theory group that trusted me, let me take them out in the water and ended up becoming surfers themselves. Thanks to my non-academic friend Yvonne for supporting me and distracting me throughout the years. Thanks to the RSF for being there to accommodate my passion

for exercise.

I would particularly like to thank Dan for being there for me, for the good and the bad moments. Thanks for all the fun we had and for accommodating me when I needed a place to stay, for staying with me when I was sick, for being so loving and supporting. Finally, I would like to place a special thanks to my parents and my late grandmother. I cannot imagine being luckier than having those three people by my side, to be proud of me for all my successes and look out for me and always make sure I am safe and happy.

Chapter 1

Introduction

Expander graphs are a surprising powerful tool that has found a wide range of applications in computer science including solving problems in communication and construction of error correcting codes as well as proving results in number theory and computational complexity. Informally, expanders are graphs in which every subset of vertices expands quickly, in the sense that it is connected to many vertices in the set of complementary vertices. In this thesis, we study three problems related to expanders, whose analysis involves understanding the intimate connection between expanders, spectra and semidefinite programming.

Before introducing these connections, it is useful to consider the closely related notion of a sparse cut in a graph. Formally, if we let $(S, \overline{S})$ denote a cut, i.e. a partition of the graph in two disjoint pieces, then the *sparsest cut* $h(G)$ of a graph G is defined as follows:

$$h(G) = \min_{S:|S|\leq V/2} \frac{E(S, \overline{S})}{|S|} \tag{1.1}$$

Such sparse graph *partitions* are central objects of study in the theory of Markov chains, geometric embeddings and are a natural algorithmic primitive in numerous settings, including clustering, divide and conquer approaches, PRAM emulation, VLSI layout, and packet routing in distributed networks.

Since the problem of finding the *sparsest cut*, i.e. the cut that minimizes the edge expansion, is NP-hard, much of recent work has focused on approximating $h(G)$. One way to approximate the sparsest cut is using the closely related notion that of *algebraic connectivity* of a graph, that is, the second smallest eigenvalue of the graph Laplacian. The intrinsic relation between the *algebraic connectivity* of a graph and the *edge expansion* can be summarized in the Alon-Millman-Cheeger inequality [**?**], [**?**], [**?**]:

Theorem 1 *([?], [?], [?]) Let d_{max} be the maximum degree of a node in G. Then*

$$\lambda_2 \leq h(G) \leq \sqrt{2 d_{max} \lambda_2} \tag{1.2}$$

Revisiting the notion of expanders we can now define an expander graph to be a graph with no sparse cut, or equivalently, with large algebraic connectivity.

Notably, the best approximation algorithms for sparsest cut use Semidefinite Programming (SDP). This is not surprising since semidefinite programs are nothing but linear programs with variables representing entries of a matrix, together with eigenvalue bounds for that matrix and could, therefore capture the spectral expansion of a graph. Similarly, most of eigenvalue optimization problems can be cast as SDPs, which leads to developing semidefinite programming based algorithms for a plethora of other important graph problems. In this thesis we further explore the connections between expansion, spectra and SDPs and apply them to several more sophisticated problems related to expander graphs.

The first problem (**chapter 3**), is related to investigating the validity of Khot's *Unique Games* conjecture (UGC) [**?**], which asserts that for a certain constraint satisfaction problem, it is NP-hard to decide whether there is a labeling that satisfies almost all the constraints or, for every labeling, the fraction of the constraints satisfied is very small. Since its origin, the UGC has been applied with remarkable success to prove tight hardness of approximation results for several important NP-hard problems such as Vertex Cover [**?**], Maximum Cut [**?**].

An interesting class of Unique Games are games where the underlying graph is an expander. Some researchers conjectured that these were good candidates for the hardest instances of UGC. There is also another important reason for studying this special class of Unique Games: Despite the remarkable progress in proving hardness of approximation results assuming the conjecture, the state of the art for innaproximability of the (uniform) Sparsest Cut (Sparsest Cut) problem is very unsatisfying . In fact, it seems unlikely that there is a reduction from Unique Games to Sparsest Cut, unless one assumes that the starting Unique Games instance has some expansion property.

We show that UGC for expanders is *false*. We present two distinct algorithms that find good assignments for instances of Unique Games when the underlying graph has some significant expansion. The first algorithm is based on a spectral partitioning approach and performs well for arbitrary Γ-max-lin constraints on the underlying Unique Games graph. The second algorithm is based on semidefinite programming and covers the case where the constraints are arbitrary permutations. We note that even though there is a reduction from the case where the Unique Games instance has arbitrary constraints to the case where the constraints are Γ-max-lin [**?**], it does not preserve the expansion of the underlying graphs. Therefore the results of the second algorithm are not implied nor can be obtained by the first.

Chapter 4 of the thesis is motivated by the problem of deciding which k edges to add to a given graph G on order to maximize its expansion. In is not hard to use SDP to find a *fractional* graph W of total *weight* k such that $G + W$ maximizes the expansion. In order to find the optimum k edges we need, however, to *round* the SDP solution. To do so, we present a novel method of *sparsification*. Given a union of two graphs $G+W$, we investigate the question of choosing a sparse subgraph $W' \subseteq W$ so that $G + W'$ *approximates* $G + W$. The notion of approximation we use is that the two graphs are spectrally *close* and has been used in the literature by several authors [**?**], [**?**], [**?**]. We prove that for every two graphs G, W there exists a subgraph $W' \subseteq W$ such that the number of edges in W' as well as the

quality of approximation of the spectrum of $G + W$ depend on the *relative* spectrum of W with respect to G. We apply the result to optimizing the algebraic connectivity of a graph by adding very few edges. We also show how to use this result in order to create optimal *ultrasparsifiers* for every graph, which can be used as good graph-theoretic *preconditioners* for symmetric, positive semidefinite, diagonally dominant linear systems.

In **chapter 5,** we study the integrality gap of the well known Sparsest Cut semidefinite program. The best approximation algorithm known for that problem is an SDP based algorithm and achieves an approximation guarantee of $O(\sqrt{\log n})$ [**?**] which is believed to be tight. And yet, the best lowerbound known for that SDP is $\Omega(\log \log n)$ and was proved by Devanur, Khot, Saket, and Vishnoi [STOC 2006], following an integrality gap for non-uniform demands due to Khot and Vishnoi [FOCS 2005]. Their constructions involve a complicated SDP solution and analysis. In this chapter, we present a simple construction and analysis of an $\Omega(\log \log N)$ integrality gap for Sparsest Cut while our gap instance, vector solution, and analysis are somewhat simpler and more intuitive. Furthermore, our approach is rather general, and provides a variety of different gap examples derived from quotients of the hypercube. It also illustrates why the lower bound is stuck at $\Omega(\log \log N)$, and why new ideas are needed in order to derive stronger examples.

In the rest of the introduction we give more detailed overview of the three main chapters in this thesis.

1.1 Algorithms for Expanding Instances of Unique Games

Unique Games is a constraint satisfaction problem where one is given a constraint graph $G = (V, E)$, a label set $[k]$ and for each edge $e = (u, v)$, a bijective mapping $\pi_{uv} : [k] \mapsto [k]$. The goal is to assign to each vertex in G a label from $[k]$ so as to maximize the fraction of the constraints that are "satisfied," where an edge $e = (u, v)$ is said to be *satisfied* by an assignment if u is assigned a label i and v is assigned a label j such that $\pi_{uv}(i) = j$. The value of a labeling $\Lambda \colon V \to [k]$ is the fraction of the constraints satisfied by it and is denoted by $\mathsf{val}(\Lambda)$. For a Unique Games instance $\mathcal{U}$, we denote by $\mathsf{opt}(\mathcal{U})$ the maximum value of $\mathsf{val}(\Lambda)$ over all labelings. This optimization problem was first considered by Cai, Condon, and Lipton [**?**]. The Unique Games Conjecture (UGC) of Khot [**?**] asserts that for such a constraint satisfaction problem, for arbitrarily small constants $\varepsilon, \delta > 0$, it is NP-hard to decide whether there is a labeling that satisfies $1 - \varepsilon$ fraction of the constraints or, for every labeling, the fraction of the constraints satisfied is at most δ as long as the size of the label set, k, is allowed to grow as a function of ε and δ.

Since its origin, the UGC has been successfully used to prove (often optimal) hardness of approximation results for several important NP-hard problems such as MIN-2SAT-DELETION [**?**], Vertex Cover [**?**], Maximum Cut [**?**], Graph Coloring [**?**], and non-uniform Sparsest Cut [**?**, **?**]. In addition, in recent years, Unique Games Conjecture(UGC) has also proved to be intimately connected to the limitations of Semidefinite Programming. Making this connection precise, [**?**] shows that if UGC is true, then for every constraint satisfaction problem(CSP) the best approximation ratio is given by a certain

simple SDP. However, one fundamental problem that has resisted attempts to prove inapproximability results, even assuming UGC, is the (uniform) SPARSEST CUT problem. This problem has a $O(\sqrt{\log n})$ approximation algorithm by Arora, Rao, and Vazirani [**?**], but no hardness result beyond NP-hardness is known (recently, in [**?**], a PTAS is ruled out under a complexity assumption stronger than $\mathsf{P} \neq \mathsf{NP}$). In fact, it seems unlikely that there is a reduction from UNIQUE GAMES to SPARSEST CUT, unless one assumes that the starting UNIQUE GAMES instance has some expansion property. This is because if the UNIQUE GAMES instance itself has a sparse cut, then the instance of SPARSEST CUT produced by such a reduction also has a sparse cut (this is certainly the case for known reductions, i.e. [**?**, **?**]), irrespective of whether the UNIQUE GAMES instance is a YES or a NO instance. This motivates the following question: is UNIQUE GAMES problem hard even with the promise that the constraint graph is an expander? A priori, this could be true even with a very strong notion of expansion (as some of the authors of this paper speculated), leading to a superconstant hardness result for SPARSEST CUT and related problems like MINIMUM LINEAR ARRANGEMENT.

1.1.1 Our Results

The main result of this chapter is that the UNIQUE GAMES problem is actually easy when the constraint graph is even a relatively weak expander. The main notion of expansion that we consider is when the algebraic connectivity of a graph G, denoted by $\lambda := \lambda_2(G)$, is bounded away from 0. As we noted earlier, the size of balanced cuts (relative to the total number of edges) in a graph is also a useful notion of expansion and the results in this paper can be extended to work in that setting. We present two approximation algorithms to find good assignments when they exist, for expanding instances of UNIQUE GAMES.

In the first part of the chapter, we study the case of random unique games generated by picking a random regular graph of degree d (or a random $G_{n,p}$ graph of average degree d) and picking a random permutation for each edge. We show that with high probability over the choice of instances, the value of the SDP from [**?**] and [**?**] is at most δ for $d = \Omega(1/\delta^4 + 1/\varepsilon^4)$.

Using techniques from the above analysis, we also study the problem of recovering planted solutions for random unique games and finding good solutions when the given Unique game is a Γ-max-lin expanding instance. Specifically, we start with studying the model where a random instance *consistent with a given solution* is chosen to start with, and an adversary then perturbs ε fraction of the constraints. Thus, the given instance has one planted solution with value $1-\varepsilon$. We give an algorithm which recovers w.h.p. a solution of value at least $1-O(\varepsilon)$ *even when the perturbations are adversarial.* The result for Γ-max-lin expanding constraint graphs follows easily from this analysis.

To obtain both the above results, we analyze the dual of the SDP. We reduce the problem of estimating the value of the SDP to estimating the eigenvalues for an associated matrix M. Since most known eigenvalue analyses are for matrices with independent entries, which does not happen to be the case with M, we adapt the analyses from [**?**] and [**?**] to our purposes. The planted and expanding Γ-max-lin cases are dealt with by analyzing the eigenvectors of this matrix.

In the second part of the paper we show the following theorem.

Theorem 2 *There is a polynomial time algorithm for* Unique Games *that, given $\varepsilon > 0$, distinguishes between the following two cases:*

- YES *case: There is a labeling which satisfies at least $1-\varepsilon$ fraction of the constraints.*
- NO *case: Every labeling satisfies less than $1 - O(\frac{\varepsilon}{\lambda}\log(\frac{\lambda}{\varepsilon}))$ fraction of the constraints.*

A consequence of the result is that when the Unique Games instance is $(1-\varepsilon)$-satisfiable and $\lambda \gg \varepsilon$, the algorithm finds a labeling to the Unique Games instance that satisfies 99% of the constraints. An important feature of the algorithm is that its performance does not depend on the number of labels k.

1.1.2 Comparison to previous work

Most of the algorithms for Unique Games (which can be viewed as attempts to disprove the UGC) are based on the SDP relaxation proposed by Feige and Lovász [**?**]. Their paper showed that if the Unique Games instance is unsatisfiable, then the value of the SDP relaxation is bounded away from 1, though they did not give quantitative bounds. Khot [**?**] gave a SDP-rounding algorithm to find a labeling that satisfies $1 - O(k^2\varepsilon^{1/5}\log(1/\varepsilon))$ fraction of the constraints when there exists a labeling that satisfies $1-\varepsilon$ fraction of the constraints. The SDP's analysis was then revisited by many papers. On an $(1-\varepsilon)$-satisfiable instance, these papers obtain a labeling that satisfies at least $1 - f(\varepsilon, n, k)$ fraction of the constraints where $f(\varepsilon, n, k)$ is $\sqrt[3]{\varepsilon \log n}$ in Trevisan [**?**], $\sqrt{\varepsilon \log k}$ in Charikar, Makarychev, and Makarychev [**?**], $\varepsilon\sqrt{\log n \log k}$ in Chlamtac, Makarychev, and Makarychev [**?**], and $\varepsilon \log n$ via an LP based approach in Gupta and Talwar [**?**]. Trevisan [**?**] also gave a combinatorial algorithm that works well on expanders. On an $(1-\varepsilon)$-satisfiable instance, he showed how to obtain a labeling satisfying $1 - \varepsilon \log n \log \frac{1}{\lambda}$ fraction of the constraints. All these results require ε to go to 0 as either n or k go to infinity in order to maintain their applicability[1]. Our main result is the first of its kind where under an additional promise of a natural graph property, namely expansion, the performance of the algorithm is independent of k and n. Furthermore, our analysis steps away from the edge-by-edge analysis of previous papers in favor of a more global analysis of correlations, which may be useful for other problems. We also provide an integrality gap for this SDP to show that, quantitatively, our main result is tight up to log factors.

We note that if we impose a certain structure on our constraints, namely if they are of the form ΓMAX2LIN, the latest results continue to hold when λ is replaced by stronger relaxations for the expansion of G, similar in spirit to the relaxations obtained by SDP hierarchies [**?**, **?**, **?**]. In particular, we show that λ can be replaced by the value of such a relaxation for expansion of G after a constant number of rounds.

Application to parallel repetition Since our main result shows an upper bound on the integrality gap for the standard SDP, the analysis of Feige and Lovász [**?**] allows us to

[1] On the other hand, the UGC allows k to grow arbitrarily as a function of ε, and therefore, all known algorithms fall short of disproving UGC.

prove (see Section 3.5) a *parallel repetition theorem* for unique games with expansion. We show that the r-round parallel repetition value of a UNIQUE GAMES instance with value at most $1-\varepsilon$ is at most $(1-\Omega(\varepsilon\cdot\lambda/\log\frac{1}{\varepsilon}))^r$. In addition to providing an alternate proof, when $\lambda \gg \varepsilon^2\log(1/\varepsilon)$, this is better than the general bound for nonunique games, where the best bound is $(1-\Omega(\varepsilon^3/\log k))^r$ by Holenstein [**?**], improving upon Raz's Theorem [**?**]. We note that recently, Safra and Schwartz [**?**] also showed a parallel repetition theorem for games with expansion, and their result works even for general games. Also, Rao [**?**] has proved a better parallel repetition theorem for, so called, projection games, which are more general than unique games. His result does not assume any expansion of the game graph.

1.2 Subgraph Sparsifiers and Applications

Graph Sparsifiers and Ultrasparsifiers. A *sparsifier* of a graph $G=(V,E,w)$ is a d-sparse graph H that is similar to G in some useful way. (We say that a graph is d-sparse if it has at most dn edges). The purpose that H serves is that it can be used as a proxy for G in computations without introducing too much error and at the same time achieve significantly faster running time. Many notions of "similarity" have been considered. We will mainly be interested in the spectral notion of similarity introduced by Spielman and Teng [**?**], [**?**]: we say that H is a κ-approximation of G if for all $x\in\mathbf{R}^V$,

$$x^T\mathcal{L}_G x \le x^T\mathcal{L}_H x \le \kappa x^T\mathcal{L}_G x \tag{1.3}$$

where $\mathcal{L}_G$ and $\mathcal{L}_H$ are the Laplacian matrices of G and H.

Equivalently, for such $H\subseteq G$ we will use the notation $H\preceq G\preceq\kappa H$ to imply that equation 4.1 holds.

In the case where G is the complete graph, excellent spectral sparsifiers are supplied by Ramanujan Graphs [**?**], [**?**]. These are d-regular graphs H all of whose non-zero Laplacian eigenvalues lie between $d-2\sqrt{d-1}$ and $d+2\sqrt{d-1}$. Thus, if we take a Ramanujan graph on n vertices and multiply the weight of every edge by $n/(d-2\sqrt{d-1})$, we obtain a graph that $\frac{d+2\sqrt{d-1}}{d-2\sqrt{d-1}}$-approximates the complete graphs

In [**?**] the authors showed that every graph can be approximated at least this well by a graph with only twice as many edges. Namely, they showed that

Theorem 3 *For every $d>1$, every undirected graph $G=(V,E,w)$ on n vertices contains a weighted subgraph $H=(V,F,\tilde{w})$ with $\lceil d(n-1)\rceil$ edges (i.e. average degree at most $2d$) that satisfies:*

$$x^T\mathcal{L}_G x \le x^T\mathcal{L}_H x \le \frac{d+2\sqrt{d-1}}{d-2\sqrt{d-1}}\cdot x^T\mathcal{L}_G x$$

A graph is *k-ultra-sparse* if it has at most n-1+k edges. We note that a spanning tree is 0-ultra-sparse. An *ultra-sparsifier* of a graph $G=(V,E,w)$ is a d-sparse graph $U\subseteq G$ that approximates G in some useful way. We use the notion of ultrasparsifiers as it appears in [**?**]. Namely, U is a (κ,k)- ultrasparsifier of G if it has the following properties

- $U\preceq G\preceq\kappa\cdot U$
- U has less than $n-1+k$ edges.

1.2.1 Our Results

In this chapter we consider a variation of the spectral sparsification problem where we are required to keep a subgraph of the original graph. Formally, given a union of two weighted graphs G and W and an integer k, we are asked to find a k-edge weighted graph W_k such that $G+W_k$ is a good spectral sparsifer of $G+W$. We will refer to this problem as the subgraph (spectral) sparsification. We present a nontrivial condition on G and W such that a good sparsifier exists and give a polynomial time algorithm to find the sparsifer.

As a significant application of our technique, we show that for each positive integer k, every n-vertex weighted graph has an $(n-1+k)$-edge spectral sparsifier with relative condition number at most $\frac{n}{k}\log n\tilde{O}(\log\log n)$ where $\tilde{O}()$ hides lower order terms. Our bound is within a factor of $\tilde{O}(\log\log n)$ from optimal. This nearly settles a question left open by Spielman and Teng about ultrasparsifiers, which is a key component in their nearly linear-time algorithms for solving diagonally dominant symmetric linear systems.

We also present another application of our technique to spectral optimization in which the goal is to maximize the algebraic connectivity of a graph (e.g. turn it into an expander) with a limited number of edges.

Ultrasparsifiers and Their Use in Solving Systems of Linear Equations. In recent years, *ultrasparsifiers* have been frequently used as provably good graph theoretic preconditioners that enable the fast solution of *linear systems*. Given an $n\times n$ symmetric, positive semidefinite, diagonally dominant matrix A with m non-zero entries, and an $n-$dimensional vector $\mathbf{b}$, a linear system solver is required to produce a vector $\tilde{\mathbf{x}}$ within relative distance ε of the solution to $A\mathbf{x}=\mathbf{b}$. Iterative methods for solving such systems consist of successively computing better approximations of $\mathbf{x}$. These methods are improved by *preconditioning*, which consists of solving $B^{-1}A\mathbf{x}=B^{-1}\mathbf{b}$ for a *preconditioner* B that is carefully chosen.

Vaidya first discovered [**?**] that for symmetric, positive semidefinite, diagonally dominant matrices one could use combinatorial techniques to construct provably good matrices (ultrasparsifiers) B for preconditioning. Since the running time of the linear system solver directly depends on the parameters k and N above, there has been a line of work by several authors [**?**] [**?**], [**?**], [**?**], [**?**] that aims in constructing ultrasparsifiers with the fewest number of edges possible. In [**?**], the authors showed the existence of $((n/k)\log^{O(1)}n, k2^{O(\sqrt{\log n\log\log n})})$ ultrasparsifiers for every graph G, thus exhibiting a linear system solver that runs in time

$$m\log^{O}(1)m+O(m\log(1/\varepsilon))+n2^{O(\sqrt{\log n\log\log n})}\log(1/\varepsilon)$$

In [**?**] the authors also conjecture the existence of ultrasparsifiers with parameters $(\frac{n}{k}\log n,k)$.

In this chapter we also show how to use our result to obtain $(\frac{n}{k}\log n\tilde{O}(\log\log n),k)$-ultrasparsifiers for any graph G. Here we use the notation $\tilde{O}$ to hide lower-order terms. We note that up to those lower order terms, these ultrasparsifiers are optimal in the sense that for any constant degree expander graph G, we cannot construct ultrasparsifiers with better parameters.

Optimizing Algebraic Connectivity by Adding few Edges . In this chapter, we present an approximation algorithm for the following problem: given a graph $G = (V, E_{base})$, a set of candidate edges E_{cand}, and a parameter k, add at most k candidate edges to G so as to maximize its algebraic connectivity, that is, find a subset $E \subset E_{cand}$ that maximizes $\lambda_2(\mathcal{L}_{G+E})$. The problem was introduced by Ghosh and Boyd [**?**], who presented a heuristic for it. It is known that the problem is NP-hard [**?**]. But prior to this work, no approximation algorithm was known for it.

We use two upper bounds for the cost of the combinatorial solution in order to prove an approximation guarantee: one upper bound is the SDP value, λ_{SDP}, and the other is $\lambda_{k+2}(\mathcal{L}_G)$ (see Lemma 72). Note that neither of these two bounds are good approximations for the value of the optimum solution by themselves (for instance, if G consists of n isolated vertices, (V, E_{cand}) is an expander, $k < n$, then the value of the combinatorial solution is 0 but $\lambda_{SDP} \sim k/n$), but their combinations lead to a good upper bound for the optimum solution λ_{OPT}.

1.3 Integrality Gaps for Sparsest Cut

As pointed out earlier, the notion of graph expansion plays a central role in the modern theory of computation. Moreover, given an input graph $G = (V, E)$, the computational problem of computing the least expanding set in G, or the extent to which G is an expander, is a fundamental one in algorithm design. If we let $E(S, \overline{S})$ denote the set of edges between $S \subseteq V$ and its complement and define, similarly to the edge expansion $h(G)$,

$$\Phi(G) = \min \left\{ \frac{|E(S, \overline{S})|}{|S||\overline{S}|} : S \subseteq V \right\},$$

then calculating $\Phi(G)$ (and the set which achieves the minimum) if the well-known *uniform* Sparsest Cut problem. Since the problem is NP-hard, much recent work has focused on approximating $\Phi(G)$.

The first such algorithm, due to Leighton and Rao [**?**], achieved an $O(\log N)$-approximation, where $N = |V|$, and was based on a linear programming relaxation that computes an all-pairs multi-commodity flow in G. Later, Linial, London, and Rabinovich [**?**], and Aumann and Rabani [**?**], found a connection between rounding this linear programming (and its generalizations) and the problem of embedding finite metric spaces into L_1.

Around this time, a natural semi-deifnite programming (SDP) relaxation was proposed. This relaxation can be written succinctly as

$$\mathsf{SDP}(G) = \min \left\{ \frac{\sum_{uv \in E} \|x_u - x_v\|^2}{\sum_{u,v \in V} \|x_u - x_v\|^2} : \|x_u - x_v\|^2 \leq \|x_u - x_w\|^2 + \|x_w - x_v\|^2 \ \forall u, v, w \in V \right\},$$

where the minimum ranges over all vectors $\{x_u\}_{u \in V} \subseteq \mathbb{R}^{N-1}$. The latter constraints are referred to alternatively as the "negative-type inequalities," the "ℓ_2^2 inequalities," or the "squared triangle inequalities," and the geometric constraints they place on the solution are still poorly understood.

In fact, Goemans and Linial [?, ?] conjectured that the integrality gap of this relaxation is only $O(1)$ (in fact, they conjectured that a more general "non-uniform" version of the problem satisfied this bound). In a seminal work of Arora, Rao, and Vazirani [?], it was shown that the integrality gap is at most $O(\sqrt{\log N})$, but the question of lower bounds on the integrality gap remained open, largely because of the difficulty of producing interesting systems of vectors that satisfied the ℓ_2^2 inequalities.

Finally, in a remarkable paper, Khot and Vishnoi [?] disproved the non-uniform Goemans-Linial conjecture using a connection with the Unique Games conjecture [?]. A year later, Devanur, Khot, Saket, and Vishnoi [?] showed how one can obtain a gap for the *uniform* version defined above. Their quantitative lower bound is $\Omega(\log\log N)$, and the exponential gap between this and the $O(\sqrt{\log N})$ upper bound still remains.

Problematically, both the constructions of [?] and [?] are shrouded in mystery. The construction and analysis have often been referred to as "difficult," "impenetrable," "extremely technical," and "magic" (the last description coming from the authors themselves). The goal of the present work is to present a simple, self-contained construction and analysis of an $\Omega(\log\log N)$ integrality gap. Our inputs instances, vector solutions, and analysis are all simpler and more intuitive than their counterparts in [?] and [?].

It is difficult to overestimate the importance of the Sparsest Cut problem, the preceding SDP, and its place in the larger theory of approximation algorithms. We mention, first of all, that the algorithm and analysis of [?] drove a huge wave of new results in approximation algorithms. Furthermore, the Sparsest Cut problem and the analysis of this SDP were some of the primary driving forces in the field of metric embeddings, and led to a number of beautiful results and connections. The SDP combines the flow-based constraints of the Leighton-Rao LP, together with the second (Laplacian) eigenvalue bound used in spectral partitioning (see [?]), and in this sense represents a new frontier in algorithm design.

Finally, we mention that the uniform Sparsest Cut problem is still very poorly understood from the standpoint of approximation algorithms. It is known to be hard to approximate within $1 + \varepsilon_0$, for some small constant $\varepsilon_0 > 0$, unless NP has subexponential-time algorithms [?], but no better lower bound is known, even assuming the unique games conjecture. On the other hand, as we previously mentioned, the best upper bound is $O(\sqrt{\log N})$.

1.3.1 Our Results

We present simple construction and analysis of an $\Omega(\log\log N)$ integrality gap for Sparsest Cut SDP. Our gap instances are simply quotients of the standard hypercube—which we will represent by $Q_n = \{\frac{-1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}^n$—under some action by permutations of the coordinates. The sparsity of cuts in these graphs was studied by Khot and Naor [?], and those authors also suggested them as a possible source for integrality gaps.

For instance, consider the cyclic shift operator $\sigma(x_1, x_2, \ldots, x_n) = (x_2, \ldots, x_n, x_1)$, and define the quotient metric

$$d(u, v) = \min\left\{\|u - \sigma^i v\|_1 : i = 0, 1, \ldots, n-1\right\},$$

which is clearly σ-invariant, i.e. $d(u,v) = d(\sigma u, v) = d(u, \sigma v)$, and hence actually a metric on the orbits of Q_n under the action of σ. It is straightforward to verify that d satisfies the triangle inequality.

Our approach is simply to define vectors $\{x_u\}_{u \in Q_n}$ such that $\|x_u - x_v\|^2 \approx d(u,v)$ holds for all $u, v \in \mathcal{P}$, where $\mathcal{P}$ is a certain "pseudorandom" subset of Q_n, and $|Q_n \setminus \mathcal{P}| = o(|Q_n|)$. We use this connection (and the fact that d is a metric) to prove the triangle inequalities for $\{x_u\}_{u \in \mathcal{P}}$. We then map all the points of $Q_n \setminus \mathcal{P}$ to some fixed x_{u_0} for $u_0 \in \mathcal{P}$. Being such a small fraction of points, their contribution to the SDP is inconsequential.

For cyclic shifts, our vector solution is essentially the following,

$$x_u = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} (\sigma^i u)^{\otimes t}, \tag{1.4}$$

for some small $t = O(1)$. In general, we simply average over the action of a group, and take small tensor powers.

Now, our $\mathcal{P}$ is essentially the set of points whose orbits are not too self-correlated, e.g. points $u \in Q_n$ with $\langle u, \sigma^i u \rangle \leq n^{-1/3}$, say, for every $i \in \{1, 2, \ldots, n-1\}$. To show that $d(u,v) \approx \|x_u - x_v\|^2$ for $u, v \in \mathcal{P}$, we will assume that $\|x_u\| = 1$ for every $u \in \mathcal{P}$ (this is almost true, by virtue of the definition of $\mathcal{P}$). In this case, it suffices to prove that $1 - \langle x_u, x_v \rangle \approx 1 - \lambda(u,v)$, where

$$\lambda(u,v) = \max \left\{ \langle u, \sigma^i v \rangle : i = 0, 1, \ldots, n-1 \right\}$$

is the associated "quotient inner product."

To see that this holds, we write

$$\langle x_u, x_v \rangle = \sum_{i=0}^{n-1} \langle u, \sigma^i v \rangle^t. \tag{1.5}$$

Now, if $\lambda(u,v) \geq 1 - \delta$, then $\langle x_u, x_v \rangle \geq (1-\delta)^t \geq 1 - \delta t$. On the other hand, if $\langle x_u, x_v \rangle \geq 1 - \delta$, we need to find a single $i \in [n]$ for which $\langle u, \sigma^i v \rangle \approx \langle x_u, x_v \rangle$. Since we are taking t^{th} powers in (1.5), any small inner products $\langle u, \sigma^j v \rangle$ are dampened out. But if there were two distinct indices i, j for which $\langle u, \sigma^i v \rangle$ and $\langle u, \sigma^j v \rangle$ were both moderately large, then $\langle u, \sigma^{i-j} u \rangle$ would also be large, which doesn't happen because $u \in \mathcal{P}$. Hence $\langle x_u, x_v \rangle$ can only be close to 1 if the contribution comes almost entirely from one shift. This matching property is precisely what yields the triangle inequalities.

Bibliography

Chapter 3 is based on work that was presented at the ACM Symposium on the Theory of Computing, 2007 [**?**]. Chapter 4 is based on joint work with Yury Makarychev, Shang-Hua Teng and Amin Saberi [**?**] and chapter 5 is based on joint work with James R. Lee [**?**].

Chapter 2

Spectral Graph Theory and SDP Basics

In this chapter we give the essential definitions and properties of graph spectra and convex optimization, with special emphasis to semidefinite programming. We start with a brief overview of some basic results from Linear Algebra and Matrix Theory. In section 2.2 we present some basic facts of spectral graph theory. We first define the adjacency matrix of a graph and examine its eigenvalues and eigenvectors. Then, we focus on the graph *Laplacian* and the spectrum of the Laplacian matrix with special emphasis on the second smallest Laplacian eigenvalue and its relation to several graph invariants, including connectivity and expanding properties. As a main source of reference regarding the Laplacian spectrum, the reader is referred to [**?**].

Next, in section 2.3, we start with some preliminary definitions and facts on convex functions and convex optimization. We proceed in 2.3.1, to introduce a special case of convex optimization, namely *Semidefinite programming* , which will be the focus of the rest of the section. We present several of the main properties of semidefinite programs [**?**], [**?**] and show that such programs arise in a variety of ways: as certain geometric extremal problems, as relaxations (stronger than linear relaxations) of combinatorial optimization problems, in optimizing eigenvalue bounds in graph theory to mention some of them. We conclude by giving examples of how semidefinite programs are used in the design of approximation algorithms for NP-hard optimization problems.

2.1 Linear Algebra Principles

Spectra of Matrices. We start with a review of eigenvalues and eigenvectors of matrices, with special focus on symmetric matrices.
Let A be an $n \times n$ real matrix. An *eigenvector* of A is a vector u such that Au is parallel to u. In other words, $Au = \lambda u$ for some real or complex number λ. the number λ is called the *eigenvalue* of A belonging to eigenvector u. Note that λ is an eigenvalue iff the matrix

$A - \lambda I$ is singular, equivalently, iff

$$\det(A - \lambda I) = 0 \tag{2.1}$$

This is an algebraic equation of degree n for λ and has therefore n roots with multiplicity.

Definition 4 *The trace of a square matrix A is defined as*

$$tr(A) = \sum_{i=1}^{n} A_{ii}$$

It is a well-known fact that the trace of A is the sum of the eigenvalues of A, each taken with the same multiplicity as it accurs among the roots of equation 2.1.
If the matrix A is symmetric, then its eigenvalues are real. Also, there is an orthonormal basis $u_1, \cdots, u_n$ of the space consisting of eigenvectors of A, so that the corresponding eigenvalues $\lambda_1, \cdots, \lambda_n$ are precisely the roots of 2.1. Then A can be written as

$$A = \sum_{i=1}^{n} \lambda_i u_i {u_i}^T$$

Equivalently, every symmetric matrix can be written as $U^T AU$ where U is an orthogonal matrix and D is a diagonal matrix with diagonal entries the eigenvalues of A.

We also state the following well-known result.

Theorem 5 *(Min–max characterization of eigenvalues of a symmetric matrix A) : Let A be an $n \times n$ hermitian matrix. Let $\lambda_1 \leq \cdots \leq \lambda_n$ be its eigenvalues listed in increasing order. Then*

$$\lambda_j(A) = \min_{S_j} \max_{x \in S_j, \|x\|=1} x^T Ax$$

Where S_j is a j dimensional subspace.
Alternatively, if $\lambda_1 \geq \cdots \geq \lambda_n$ are listed in decreasing order, then

$$\lambda_j(A) = \max_{S_j} \min_{x \in S_j, \|x\|=1} x^T Ax$$

Where S_j is a j dimensional subspace.

Positive Semidefinite Matrices. We next turn our attention to a certain class of symmetric matrices that are called *positive semidefinite* matrices.

Definition 6 *A symmetric $n \times n$ matrix is called positive semidefinite if all its eigenvalues are non-negative. The matrix is positive definite if all its eigenvalues are positive.*

Below we summarize equivalent ways of characterizing positive semidefinite matrices.

Proposition 7 *For a real symmetric $n \times n$ matrix A, the following are equivalent:*

- *(i) A is positive semidefinite (PSD for short);*
- *(ii) the quadratic form $x^T Ax$ is nonnegative for every $x \in \mathbb{R}^n$;*
- *(iii) A can be written as the Gram matrix of n vectors $u_1, \cdots, u_n \in \mathbb{R}^m$ for some m; this means that $a_{ij} = \langle u_i, u_j \rangle$. Equivalently $A = U^T U$ for some matrix U;*
- *(iv) A is a nonnegative linear combination of matrices of the type $x^T x$;*
- *(v) The determinant of every symmetric minor of A is nonnegative.*

Some well known definitions and facts follow.

Definition 8 *(Frobenius inner product of matrices) For two $n \times n$ matrices A and B we define the following inner product:*

$$A \bullet B = \sum_{i,j} a_{ij} b_{ij} = tr(A^T B)$$

Definition 9 *A convex cone in $\mathbb{R}^n$ is a set of vectors V, with the following properties:*

1. *$v \in V$ implies $\alpha v \in V$ for any positive scalar α.*
2. *$v \in V$ and $u \in V$ implies that $v + u \in V$.*

Proposition 10

- *The sum of two PSD matrices is PSD.*
- *If A and B are PSD matrices then $tr(AB) \geq 0$ and equality holds iff $AB = 0$. However, the product of two PSD matrices need not be PSD.*
- *A matrix A is PSD iff $A \bullet B \geq 0$ for every PSD matrix B.*

Given the above, we conclude with the geometric statement that the set of all positive semidefinite matrices forms a convex closed cone in $\mathbb{R}^{n \times n}$ with vertex 0.

2.2 Spectra of Graphs

2.2.1 The Adjacency Matrix.

We remind the reader that for a graph G, the adjacency matrix $A = A_G$ is defined as :

$$A_G = \begin{cases} 1 & \text{if } (u,v) \in E \\ 0 & \text{if } (u,v) \notin E \end{cases}$$

If the graph has n vertices, A_G has n real eigenvalues $\mu_1 \geq \mu_2 \geq \cdots \mu_n$. The eigenvectors that correspond to these eigenvalues form an orthonormal basis of $\mathbb{R}^n$. We note

that if the graph is d-regular then the largest eigenvalue is equal to d and the corresponding eigenvector is the all-one's vector.

We can use the Courant-Fisher Theorem to characterize the spectrum of A. The largest eigenvalue satisfies

$$\mu_1 = \max_{x \in \mathbb{R}^n} \frac{x^T A x}{x^T x}$$

If we denote the first eigenvector by x_1 then

$$\mu_2 = \max_{x \in \mathbb{R}^n, x \perp x_1} \frac{x^T A x}{x^T x}$$

Similar definitions hold for the eigenvalues μ_i, $i \geq 3$.

We also allow weighted graphs which are viewed as a graph which has for each pair (u, v) of vertices, assigned a certain weight w_{uv} . The weights are usually real numbers and they must satisfy the following conditions:

- (i) $a_{uv} = w_{vu}$ for all $v, u \in V(G)$, and
- (ii) $w_{vu} \neq 0$, if and only if v and u are adjacent in G .

Unweighted graphs can be viewed as a special case of weighted graphs, by specifying, for each $u, v \in V(G)$, the weight w_{uv} to be equal to the number of edges between u and v. In the weighted case, we can similarly define the adjacency matrix $A = A(G) = [w_{uv}]$.

There are many useful connections between the eigenvalues of a graph and its combinatorial properties.

As an example, we state the following ([**?**], [**?**]). The proof follows easily from interlacing eigenvalues.

Proposition 11 *The maximum size $\omega(G)$ of a clique in G is at most $\mu_1 + 1$. This bound remains valid even if we replace the non-diagonal 0's in the adjacency matrix by arbitrary real numbers.*

The following bound on the chromatic number is due to Hoffman [**?**].

Proposition 12 *The chromatic number $\chi(G)$ of G is at least $1 - (\frac{\mu_1}{\mu_n})$. This bound remains valid even if we replace the 1's in the adjacency matrix by arbitrary real numbers.*

2.2.2 The Graph Laplacian and its Spectrum.

Basic Properties. Let $d(v)$ denote the degree of $v \in V(G)$, $d(v) = \sum_u w_{uv}$, and let $D = D(G)$ be the diagonal matrix indexed by $V(G)$ and with $d_{vv} = d(v)$. The matrix $\mathcal{L} = \mathcal{L}_G = D(G) - A(G)$ is called the Laplacian matrix of G . The matrix $\mathcal{L}_G$ is sometimes called the Kirchhoff matrix of G due to its role in the well-known Matrix-Tree Theorem which is usually attributted to Kirchhoff.

The following properties were established by several authors [**?**], [**?**], [**?**] for the case of unweighted graphs. The proofs carry over to the weighted case if all the weights are non-negative.

Theorem 13 *Let $G = (V, E, w)$ be a weighted undirected graph with n vertices and m edges and weights $w_e > 0$. Then:*

- *(a) $\mathcal{L}_G$ has only real eigenvalues,*
- *(b) $\mathcal{L}_G$ is positive semidefinite,*
- *(c) its smallest eigenvalue is $\lambda_1 = 0$ and a corresponding eigenvector is $\mathbf{1}$. The multiplicity of 0 as an eigenvalue of $\mathcal{L}_G$ is equal to the number of connected components of G.*

We note the next useful expression for the inner product (quadratic form) $\langle \mathcal{L}_G x, x\rangle$ which holds also in the weighted case:

$$\langle \mathcal{L}_G x, x\rangle = \sum_{uv} w_{uv}(x_u - x_v)^2 \tag{2.2}$$

Let $\lambda_1, \cdots, \lambda_n$ be the eigenvalues of $\mathcal{L}_G$ in increasing order and repeated according to their multiplicity. So, item (c) above is equivalent to the statement that $\lambda_1 = 0$, and $\lambda_2 > 0$ if and only if G is connected.

The Incidence Matrix and the Cut Space of a Graph. We next derive an alternative expression for the graph Laplacian and use it to show theorem (13). Namely, we observe that if we orient the edges of G arbitrarily, we can write its Laplacian as $\mathcal{L} = B^T W B$ where $B_{m\times n}$ is the *signed edge-vertex incidence matrix*, given by

$$B(e,u) = \begin{cases} 1 & \text{if u is e's head,} \\ -1 & \text{if u is e's tail,} \\ 0 & \text{otherwise} \end{cases}$$

and $W_{m\times m}$ is the diagonal matrix with $W(e,e) = w_e$. Denote the row vectors of B by $\{b_e\}_{e\in E}$ and note that ${b_{(u,v)}}^T = \chi_v - \chi_u$, where χ denotes the characteristic vector of a node. It is well known that $\mathrm{im}(B) \subseteq \mathbf{R}^m$ is the *cut space* of G, [**?**].

We will assume from here on that the graph G is connected. For item (b) of theorem (13), it is immediate that $\mathcal{L}$ is positive semidefinite since:

$$x^T \mathcal{L} x = x^T B^T W B x = {\|W^{1/2} B x\|_2}^2 \geq 0$$

For (c), we also have $\ker(\mathcal{L}) = \ker(W^{1/2}B) = \mathrm{span}(\mathbf{1})$, since

$$\begin{aligned} & x^T \mathcal{L} x = 0 \Leftrightarrow {\|W^{1/2} B x\|_2}^2 = 0 \\ \Leftrightarrow \quad & \sum_{uv} w_{uv}(x_u - x_v)^2 = 0 \\ \Leftrightarrow \quad & (x_u - x_v) = 0 \text{ for all edges } (u,v) \\ \Leftrightarrow \quad & x \text{ is constant, since G is connected.} \end{aligned}$$

The Pseudoinverse of the Laplacian of a connected graph. Since $\mathcal{L}$ is symmetric, we can diagonalize it and write

$$\mathcal{L} = \sum_{i=2}^{n} \lambda_i u_i {u_i}^T$$

Where u_i are the orthonormal eigenvectors of $\mathcal{L}$. Since the graph is connected, all eigenvalues but λ_1 are strictly positive. The *Moore-Penrose Pseudoinverse* of $\mathcal{L}$ is defined as

$$\mathcal{L}^\dagger = \sum_{i=2}^{n} \frac{1}{\lambda_i} u_i {u_i}^T \tag{2.3}$$

Note that $\ker(\mathcal{L}) = \ker(\mathcal{L}^\dagger)$ and that

$$\mathcal{L}\mathcal{L}^\dagger = \mathcal{L}^\dagger\mathcal{L} = \sum_{i=2}^{n} u_i {u_i}^T$$

Which is simply the identity on $\mathrm{im}(\mathcal{L}) = \ker(\mathcal{L})^\perp = \mathbf{R}^n \setminus \mathrm{span}(\mathbf{1})$.

The Effective Resistance. The effective resistance R_{ij} between two nodes i, j of a weighted graph is the electrical resistance between the nodes of the corresponding resistor network with branch conductances given by the edge weights. In other words, R_{ij} is the potential difference that appears across nodes i and j when a unit current source if applied between them. Formally, we can define the effective resistance as follows. For more on effective resistance the reader is referred to [**?**], [**?**].
Let v be the solution to the equation

$$\mathcal{L}_G v = e_i - e_j$$

where e_i denotes the ith unit vector, with 1 in the ith position and zero everywhere else. Note that $e_i - e_j$ is in the range of $\mathcal{L}_G$ so the above equation has a solution. We define R_{ij} as

$$R_{ij} = u_i - u_j$$

We define the effective resistance matrix $R \in \mathbb{R}^{n \times n}$ as $R = [R_{ij}]$. We note that R is symmetric with 0 in the diagonal.

There are several alternative formulas to express the effective resistance between two nodes [**?**], [**?**]. We are primarily interested in the expressions that involve the pseudoinverse of $\mathcal{L}_G$.

Claim 14 *The following formula gives R_{ij} in terms of ${\mathcal{L}_G}^\dagger$.*

$$R_{ij} = (e_i - e_j)^T {\mathcal{L}_G}^\dagger (e_i - e_j) = ({\mathcal{L}_G}^\dagger)_{ii} + ({\mathcal{L}_G}^\dagger)_{jj} - 2({\mathcal{L}_G}^\dagger)_{ij}$$

Proof:

We note that $\mathcal{L}_G {\mathcal{L}_G}^\dagger = I - \mathbf{1}\mathbf{1}^\dagger/n$ and multiply $\mathcal{L}_G v$ on the left hand side by ${\mathcal{L}_G}^\dagger$

to get $(I - \mathbf{1}\mathbf{1}^\dagger/n)v = {\mathcal{L}_G}^\dagger(e_i - e_j)$ so

$$(e_i - e_j)^T {\mathcal{L}_G}^\dagger(e_i - e_j) = (e_i - e_j)v = u_i - v_j = R_{ij}$$

□

We also note that from the above claim it follows that

$$R_{ij} = (e_i - e_j)^T \mathcal{L}_G + \mathbf{1}\mathbf{1}^\dagger/n^{-1}(e_i - e_j) \tag{2.4}$$

λ_2 - The Algebraic Connectivity of Graphs. The second smallest Laplacian eigenvalue λ_2 of graphs is probably the most important information contained in the spectrum of a graph. This eigenvalue is related to several important graph invariants, and it has been extensively investigated. Most of the results are consequences of the well-known Courant-Fischer principle which, as we noted before, states that

$$\lambda_2 = \max_{x \in \mathbb{R}^n, x \perp \mathbf{1}, x \neq \mathbf{0}} \frac{x^T \mathcal{L}_G x}{x^T x} \tag{2.5}$$

Fiedler [**?**] obtained another expression for λ_2. He also called the number $\lambda_2(\mathcal{L}_G)$ *algebraic connectivity* influenced by its relation to the classical connectivity parameters of the graph- the vertex connectivity and the edge connectivity.

Proposition 15 *Let G be a weighted graph with non-negative weights a_{uv} . Then*

$$\lambda_2 = 2n \min_{x \in \Phi} \frac{\sum_{uv \in E(G)} a_{uv}(x_u - x_v)^2}{\sum_{u \in V(G)} \sum_{v \in V(G)} (x_u - x_v)^2} \tag{2.6}$$

where Φ is the set of all non-constant vectors.

The intrinsic relation between the *algebraic connectivity* of a graph and the *edge expansion* can be summarized in the Alon-Millman-Cheeger inequality [**?**], [**?**], [**?**]. We first define the *edge expansion* (sparsest cut) of a graph G.

Definition 16 *Let $(S, \overline{S})$ denote a cut, i.e. a partition of the graph in two disjoint pieces. The edge expansion $h(G)$ of a graph G is defined as follows:*

$$h(G) = \min_{S : |S| \leq V/2} \frac{E(S, \overline{S})}{|S|} \tag{2.7}$$

Equivalently, if we let $x_S \in \{0,1\}^n$ be the characteristic vector of the cut $(S, \overline{S})$, with $x_S(v) = 1$ if $v \in S$ and $x_S(v) = 0$ if $v \in \overline{S}$, and $|S| \leq \frac{|V|}{2}$, then

$$h(G) = \min_{x_S \in \{0,1\}^n, x \neq 0} \frac{x_S^T \mathcal{L}_G x_S}{x_S^T x_S} \tag{2.8}$$

Theorem 17 *([?], [?], [?]) Let d_{max} be the maximum degree of a node in G. Then*

$$\lambda_2 \leq h(G) \leq \sqrt{2 d_{max} \lambda_2} \tag{2.9}$$

Remark 1 *There is a very strong relationship between the eigenvalues of the adjacency matrix of G and the eigenvalues of the Laplacian. In particular, for d–regular graphs, $\lambda_i = d - \mu_i$.*

2.3 Convex Functions and Convex Optimization

In this section, we give some principles regarding convex functions and introduce convex optimization.

Convex Functions. We first define convex sets.

Definition 18 *A set C is convex if the line segment between any two points in C lies in C, i.e., if for any $x, y \in C$ and any $0 \leq \theta \leq 1$, we have*

$$\theta x + (1 - \theta)y \in C$$

Definition 19 *(Convex Function) A function $f : \mathbf{R}^n \to \mathbf{R}$ is convex if $dom(f)$ is a convex set and if for all x, $y \in dom(f)$ and all $0 \leq \theta \leq 1$.*

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

The Epigraph. The graph of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$\{(x, f(x))|x \in \mathbf{dom}f\}$$

which is a subset of $\mathbf{R}^{n+1}$. The *epigraph* of f is defined as

$$\mathbf{epi}f = \{(x, t)|x \in \mathbf{dom}f, f(x) \leq t\}$$

Note that a function is convex if and only if its epigraph is a convex set.

Convex Optimization. A convex optimization problem is one of the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \qquad ii = 1, 2, \cdots, m \end{array}$$

where the functions $f_0, f_i : \mathbf{R}^n \to \mathbf{R}$ are *convex*.

There is in general no analytical formula for the solution of convex optimization problems, but (as with linear programming problems) there are very effective methods for solving them. Interior-point methods work very well in practice, and in some cases can be proved to solve the problem to a specified accuracy with a number of operations that does not exceed a polynomial of the problem dimensions.

We next note a well-known fact about convex programs. For a proof see, for example [**?**], chapter 11.

Fact 20 *Consider a convex program as above. If there is a separation oracle that runs in polynomial time for deciding whether a point is feasible, or, alternatively, if there is a membership oracle for the convex set, the set is centered (i.e., we are given a small ball inside the set), and there is separation oracle for level sets of the objective function, then the convex program can be solved in polynomial time by some interior point method.*

2.3.1 Semidefinite Programming

2.3.2 Properties of Semidefinite Programs

A special case of convex program is a *semidefinite program* (SDP for short). Namely, it is a convex optimization problem of a linear objective function over the intersection of the cone of positive semi-definite matrices with an affine space. It has the following form :

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & x_1 A_1 + x_2 A_2 \dots x_n A_n - B \succeq 0 \end{aligned}$$

where $A_1, A_2, \dots A_n, B$ are symmetric square matrices and $c \in \mathbb{R}^n$ is a given vector. We can think of $X = x_1 A_1 + x_2 A_2 \dots x_n A_n - B$ as a matrix whose entries are linear functions of the variables.

The dual of the above SDP is

$$\begin{aligned} \text{maximize} \quad & B \bullet Y \\ \text{subject to} \quad & A_1 \bullet Y = c_1 \\ & A_2 \bullet Y = c_2 \\ & \vdots \\ & A_n \bullet Y = c_n \\ & Y \succeq 0 \end{aligned}$$

From (weak) duality, we have the following inequality between the optimal values of the primal and dual SDP: $\text{val}_{\text{primal}} \leq \text{val}_{\text{dual}}$.

Strong duality obtains if the primal semidefinite program above is strictly feasible, i.e., there exists an x with

$$x_1 A_1 + x_2 A_2 \dots x_n A_n - B \succ 0$$

When strong duality holds, we have the equality : $\text{val}_{\text{primal}} = \text{val}_{\text{dual}}$.

The special case when $A_1, \cdots, A_n, B$ are diagonal matrices is just a generic linear program, and it is very fruitful to think of semidefinite programs as generalizations of linear programs.

2.3.3 Semidefinite Programs for Eigenvalue Optimization

In many cases semidefinite programs arise in the form of minimizing or maximizing an appropriate linear combination of eigenvalues of a symmetric matrix subject to linear constraints on the matrix. In this section we give some examples of eigenvalue optimization problems that can be cast as an SDP.

We will focus on propositions 12 and 11 that appeared in section 2.2. We make the following observation :

In order to determine the best choice of the "free" entries of the matrix as promised in the statement of propositions 12 and 11, we can write a semidefinite program. For the bound on the maximum size of a clique as in 11, we fix the diagonal entries at 0, the entries corresponding to edges at 1, but are free to choose the entries corresponding to nonadjacent pairs of vertices. We want to minimize the largest eigenvalue. This can be written as a semidefinite program:

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & tI - X \succeq 0 \\ & X_{ii} = 0 \qquad \forall i \in V \\ & X_{ij} = 1 \qquad \forall ij \in E \end{aligned}$$

In addition, it turns out that the semidefinite program for 12 is the dual of the above, and their common optimum value is a parameter that was called θ function by Lovasz [**?**].

2.3.4 Semidefinite Programs in Approximation Algorithms

Semidefinite programs are often used as relaxations of 0/1 quadratic programs. In obtaining the relaxations, we often replace 0/1 variables $x_1, \ldots, x_n$ by vectors $\mathbf{v_1}, \ldots, \mathbf{v_n}$. Alternatively, we may think of solving for an $n \times n$ positive semidefinite matrix Y as above, such that $Y_{ij} = \mathbf{v_i} \cdot \mathbf{v_j}$. In what follows we will give an example of an approximation algorithm for a 0/1 quadratic optimization problem that is based on semidefinite programming.

Definition 21 *(MAXCUT) Given a graph $G = (V, E)$, the MAXCUT problem asks to output a partition of the vertices V so as to maximize the number of edges crossing from one side to the other.*

The first approximation algorithm based on an SDP is due to Goemans and Williamson [**?**], for the MAXCUT problem. This problem can be expressed as an integer quadratic program:

$$\begin{aligned} \text{maximize} \quad & \sum_{(i,j)\in E} \frac{1 - v_i v_j}{2} \\ \text{subject to} \quad & v_i \in \{-1, 1\} \qquad \forall i \end{aligned}$$

Unless P = NP, we cannot solve this maximization problem efficiently. However, the authors in [**?**] observed that relaxing the integer quadratic program into an SDP and consequently rounding the SDP solution to obtain an approximate solution to the original integer quadratic program results to a good approximation algorithm for MAXCUT.

The most natural relaxation is

$$\begin{aligned} \text{maximize} \quad & \sum_{(i,j)\in E} \frac{1 - \langle v_i, v_j \rangle}{2} \\ \text{subject to} \quad & \|v_i\|^2 = 1 \qquad \forall i \end{aligned}$$

where the maximization is over vectors v_i instead of integer scalars. Solving the SDP gives a set of unit vectors in $\mathbb{R}^n$; since the vectors are not required to be collinear, the value of this relaxed program can only be higher than the value of the original quadratic integer program. Finally, a rounding procedure is needed to obtain a partition. Goemans and Williamson simply choose a uniformly random hyperplane through the origin and divide the vertices according to which side of the hyperplane the corresponding vectors lie. Straightforward analysis shows that this procedure achieves an expected approximation ratio (performance guarantee) of 0.87856. Assuming the Unique Games Conjecture, it can be shown that this approximation ratio is essentially optimal.

Since the original paper of Goemans and Williamson, SDPs have been applied to develop numerous approximation algorithms. Recently, Prasad Raghavendra [**?**] has developed a general framework for constraint satisfaction problems based on the Unique Games Conjecture.

Chapter 3

Unique Games on Expanding Instances

In this chapter we present two distinct algorithms that find good assignments for instances of Unique Games when the underlying graph has some significant expansion. The first algorithm is based on a spectral partitioning approach and performs well for arbitrary Γ-max-lin constraints on the underlying Unique Games graph. The second algorithm is based on semidefinite programming and covers the case where the constraints are arbitrary permutations. We note that even though there is a reduction from the case where the Unique Games instance has arbitrary constraints to the case where the constraints are Γ-max-lin [**?**], it does not preserve the expansion of the underlying graphs. Therefore the results of the second algorithm are not implied nor can be obtained by the first.

The chapter is organized as follows. The first part of the chapter (sections 3.2-3.3), is devoted to presenting the spectral approach which results in a spectral partitioning algorithm for expanding instances of Unique Games where the constraints are arbitrary Γ-max-lin. In section 3.2, we analyze the behavior of the SDP by Feige and Lovász [**?**] on random instances of unique games. We show that on random d-regular graphs with permutations chosen at random, the value of the SDP is very small with probability $1 - e^{-\Omega(d)}$. Hence, the SDP provides a proof of unsatisfiability for random unique games. In section 3.3 we show how the eigenvectors of a particular matrix may be used to recover good assignments to highly satisfiable instances of unique games in some special cases. We start in 3.3.1 by giving a spectral algorithm for recovering planted solutions. Given a random instance consistent with a given solution on $1 - \varepsilon$ fraction of the edges, our algorithm recovers a solution with value $1 - O(\varepsilon)$ with high probability at least $1 - e^{-\Omega(d)}$ over the inputs. Using similar arguments as in the planted solution case, we conclude in 3.3.2 with an algorithm that finds good solutions for a Γ-max-lin expanding unique game. We present both cases in a unified manner in order to emphasize the main ideas that were used in the analysis of the algorithm.

The second part of the chapter 3.4, describes a semidefinite programming-based algorithm that finds good assignments for instances of Unique Games when the underlying graph has some significant expansion and the constraints are arbitrary permutations. A new analysis of the standard SDP is introduced, which involves correlations among distant vertices. In

$$\text{Maximize} \quad \sum_{(u,v)\in E} \sum_{i=1}^{k} \langle \mathbf{u_i}, \mathbf{v}_{\pi_{\mathbf{uv}}(\mathbf{i})} \rangle \tag{3.1}$$

Subject to

$$\forall\ u \in V \quad \sum_{i=1}^{k} \|\mathbf{u_i}\|^2 = 1 \tag{3.2}$$

$$\forall\ u \in V \ \ \forall\ i \neq j \qquad \langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0 \tag{3.3}$$

$$\tag{3.4}$$

Figure 3.1: SDP1 for UNIQUE GAMES

section 3.5, we also show how this technique leads to a parallel repetition theorem for unique games when the graph is an expander.

3.1 Notation and Preliminaries

Unique Games. A *Unique Game* is defined in terms of a constraint graph $G = (V, E)$, a set of variables $\{x_u\}_{u\in V}$, one for each vertex u and a set of permutations (constraints) $\Pi_{uv} : [k] \to [k]$, one for each edge (u, v). An assignment to the variables is said to satisfy the constraint on the edge $(u, v) \in E$ if $\pi_{uv}(x_u) = x_v$. The edges are taken to be undirected and hence $\pi_{uv} = (\pi_{vu})^{-1}$. The goal is to assign a value from the set $[k]$ to each variable x_u so as to maximize the number of satisfied constraints.

An instance of Unique Games is Γ-max-lin when the constraints are of a very specific form, namely they are all linear equations over some abelian group Γ.

Khot [**?**] conjectured that it is NP-hard to distinguish between the cases when almost all the constraints of a unique game are satisfiable and when very few of the constraints are satisfiable. Formally, the statement of the conjecture is the following:

Conjecture 22 *(Unique Games Conjecture) For any constants $\varepsilon, \delta > 0$, for any $k > k(\varepsilon, \delta)$, 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 at most δ fraction of constraints are satisfiable.*

SDP relaxation for UNIQUE GAMES. Let $\mathcal{U} = (G(V, E), [k], \{\pi_{uv}\}_{(u,v)\in E})$ be a *UNIQUE GAMES* instance. We use standard SDP relaxations that appear in figures 3.1 and 3.2, which involve finding a vector assignment for each vertex. For every $u \in V$, we associate a set of k orthogonal vectors $\{\mathbf{u_1}, \cdots, \mathbf{u_k}\}$. In SDP1, the intention is that if $i_0 \in [k]$ is a label for vertex $u \in V$ then $u_{i_0} = 1$ and $u_i = 0$ for all $i \neq i_0$. In SDP2, the intention is that if $i_0 \in [k]$ is a label for vertex $u \in V$ then $u_{i_0} = \mathbf{1}$ and $u_i = \mathbf{0}$ for all $i \neq i_0$. Here, $\mathbf{1}$ is some fixed unit vector and $\mathbf{0}$ is the zero vector. Of course, in a general solution to the SDP this may no longer be true and $\{\mathbf{u_1}, \cdots, \mathbf{u_k}\}$ is just any set of orthogonal vectors.

$$\text{Maximize} \quad \mathbf{E}_{e=(u,v)\in E}\mathbf{E}_{i\in[k]} \left\langle \mathbf{u}_i, \mathbf{v}_{\pi_{uv}(i)} \right\rangle \tag{3.5}$$

Subject to

$$\forall\ u \in V \quad \mathbf{E}_{i\in[k]} \left\|\mathbf{u}_i\right\|^2 = 1 \tag{3.6}$$

$$\forall\ u \in V \ \ \forall\ i \neq j \qquad \langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0 \tag{3.7}$$

$$\forall\ u, v \in V \ \ \forall\ i, j \qquad \langle \mathbf{u}_i, \mathbf{v}_j \rangle \geq 0 \tag{3.8}$$

Figure 3.2: SDP2 for UNIQUE GAMES

3.2 Certifying Unsatisfiability for Random Instances of Unique Games

We look at the SDP for Unique-Games that appears in figure 3.1. The feasible region of the dual can be expressed as $Z \succeq 0$ where Z is an $nk \times nk$ matrix. We use Z_{uv} to denote the $k \times k$ block corresponding to the vertices u and v. The blocks are given by

$$Z_{uv} = \begin{cases} 0 & \text{if } (u,v) \notin E, u \neq v \\ -\frac{1}{2}\Pi_{uv} & \text{if } (u,v) \in E \\ Z_u & \text{if } u = v \end{cases}$$

where Π_{uv} is the permutation matrix corresponding to π_{uv} and Z_u is the (symmetric) matrix of all the variables corresponding to the vertex u. The off-diagonal entries of Z_u are $(Z_u)_{ij} = (Z_u)_{ji} = \frac{1}{2}x^u_{\{i,j\}}$ - a separate variable for each pair $\{i,j\}$ and vertex u. All the diagonal entries are the same, equal to a single variable $x^{(u)}$. The objective function of the whole SDP is $\sum_{u\in V} x^u$.

We will consider dual solutions with $x^u_{\{i,j\}} = 2d/k$ for all $u \in V$ and $i, j \in [k], i \neq j$. Also, we set $x^{(1)} = x^{(2)} = \ldots = x^{(n)} = \lambda + d/2k$. Here λ is taken to be an upper bound on the second eigenvalue. Note that the first eigenvalue of M is d since M can be thought of as the adjacency matrix of a d-regular graph on nk vertices. The objective value as $nd/2k + n\lambda$. Putting in these values for the variables, we will need to show that the following equation is satisfied.

$$\lambda I + \frac{d}{2k}J - \frac{1}{2}M \succeq 0$$

where I is the $nk \times nk$ identity matrix, J is a block diagonal matrix with $k \times k$ blocks of all 1s on the diagonal and M is a block matrix with $M_{uv} = \Pi_{uv}$ if $(u,v) \in E$ and 0 otherwise.

Let z denote the all vector with all coordinates $\frac{1}{\sqrt{nk}}$. Then z is the first eigenvector of M. We prove the following in the next section

Theorem 23 *Let M be a matrix generated according to a random d-regular graph and random permutations on each edge. Then, with probability $1 - e^{-\Omega(d)}$, $\lambda_2(M) \leq Cd^{3/4}$*

Hence, we take $\lambda = Cd^{3/4}$ which is a bound on the second eigenvalue[1]. Note that

[1]We believe that it is possible to improve this bound to even $C\sqrt{d}$ but this is not very important for our

z is the first eigenvector of both J and M. Since we can express any vector x and $\alpha z + \beta w$ with $w \perp z$, we have

$$\begin{aligned} x^T \left(\lambda I + \frac{d}{2k}J - \frac{1}{2}M\right) x &= (\alpha z + \beta w)^T \left(\lambda I + \frac{d}{2k}J - \frac{1}{2}M\right)(\alpha z + \beta w) \\ &= \lambda + \alpha^2 \frac{d}{2} + \beta^2 \frac{d}{2} w^T J w - \frac{1}{2}\left(\alpha^2 z^T M z + \beta^2 w^T M w\right) \end{aligned}$$

Since J is positive semidefinite, $z^T M z \leq d$ and $w^T M w \leq C d^{3/4}$, we have $x^T(\lambda I + \frac{\lambda_1}{2k}J - \frac{1}{2}M) \geq 0$ for every x. This gives that the value of the SDP for random d-regular graphs is $\frac{|E|}{k} + \frac{|E|}{d^{1/4}}$ with high probability.

3.2.1 Bounding the second eigenvalue for random d-regular graphs

We consider undirected random $2d$-regular graphs G_{2d} on n vertices constructed by choosing d permutations (over n elements) independently at random. For each of the chosen permutations σ and for each vertex u we add to the graph the edge $(u, \sigma(u))$. The unique game is then constructed for by then picking a random permutation π_{uv} (over k elements) for each edge $(u, v) \in E$.

The bound on the second eigenvalue is obtained in two steps. We first by first bound the expected value by examining the trace of a power of the matrix M. We then show a concentration bound using an application of Talagrand's inequality adapted from [**?**].

3.2.2 Bounding the mean

In the following argument, it will be convenient to consider the normalized matrices $M^* = (2d)^{-1}M$, $A^* = (2d)^{-1}A$. For any positive integer p, we have $Trace((M^*)^p) = \frac{1}{(2d)^p} Trace(M^p)$ and same for A^*. Let $\rho_1, \rho_2, \cdots, \rho_{nk}$ the eigenvalues of M^* in order of decreasing value. Clearly, $\rho_1 = 1$. Our next goal is to upper-bound the mean value of the quantity $\rho = max\{\rho_2, |\rho_n|\}$. Let p be a large positive integer to be fixed later.

Lemma 24

$$E[\rho] \leq (E[Trace((M^*)^{2p})] - 1)^{1/2p}$$

PROOF: Because Trace$((M^*)^{2p}) = \sum_{1 \leq i \leq nk} {\rho_i}^{2p}$ and because all the eigenvalues of a symmetric matrix are real, we have :

$$\rho^{2p} \leq Trace((M^*)^{2p}) - 1$$

Taking expectations over the probability space described above,(that is, over all $2d$-regular graphs and over all permutations of k elements within each non-zero block), we have

$$E[\rho] \leq E[\rho^{2p}]^{1/(2p)} \leq (E[Trace((M^*)^{2p})] - 1)^{1/2p}$$

by Jensen's inequality. □

---

purposes.

We next relate the value of $E[Trace((M^*)^{2p})]$ to $E[Trace((A^*)^{2p})]$.

Claim 25 *Let $A = [a_{ij}]$ be the adjacency matrix of a graph G and M be a block matrix with $M_{uv} = \Pi_{uv}$ if $(u, v) \in E$ and 0 otherwise. Then $E[Trace(M^{2p})] = Trace(A^{2p})$ where p is a positive integer and the expectation on the left hand side is taken over the choice of permutations.*

PROOF: Let S be a set containing all the sequences of $2p+1$ nodes of G that begin and end at the same node. I.e $S = \{uu_1 \cdots u_{2p}u\}$. Each $s \in S$ corresponds to a walk on G of length $2p$ that begins and ends at the same node and therefore also corresponds to a sequence of blocks of the matrix M above that begins and ends at the same block.

For any matrix $Q = [q_{ij}]$ and for any positive integer n we have

$$Trace(Q^n) = \sum_{i_1, i_2, \ldots i_n} q_{i_1 i_2} q_{i_2 i_3} \ldots q_{i_n i_1}$$

Observe that when Q is the adjacency matrix of a graph, each term in the above sum is 1 if $i_1, i_2, \cdots i_n, i_1$ is a path in the graph and 0 otherwise.

Thus, for the matrices A and M we have

$$\begin{aligned} Trace(A^{2p}) &= \sum_{u_1 u_2 \ldots u_{2p} u_1 \in S} a_{u_1 u_2} \ldots a_{u_{2p} u_1} \\ Trace(M^{2p}) &= \sum_{\substack{u_1, u_2, \ldots, u_{2p} u_1 \in S \\ i_1, i_2, \ldots, i_{2p} \in [k]}} m_{(u_1,i_1)(u_2,i_2)} \ldots m_{(u_{2p},i_{2p})(u_1,i_1)} \end{aligned}$$

where the tuple (u, i) corresponds to the index of the ith element of block u.

We can write each term $m_{(u,i)(v,j)} = a_{uv} \ldots \mathbb{I}_{\{\pi_{uv}(i)=j\}}$, where the random variable $\mathbb{I}_{\{\pi_{uv}(i)=j\}}$ is 1 when $\pi_{uv}(i) = j$ and 0 otherwise. We can now re-write the trace as

$$Trace(M^{2p}) = \sum_{u_1 u_2 \ldots u_{2p} u_1 \in S} a_{u_1 u_2} \ldots a_{u_{2p} u_1} \sum_{i_1, i_2, \ldots, i_{2p} \in [k]} \mathbb{I}_{\{\pi_{u_1 u_2}(i_1)=i_2\}} \ldots \mathbb{I}_{\{\pi_{u_{2p} u_1}(i_{2p})=i_1\}}$$

and, taking expectation over all permutations

$$E[Trace(M^{2p})] = \sum_{u_1 u_2 \ldots u_{2p} u_1 \in S} a_{u_1 u_2} \ldots a_{u_{2p} u_1} \sum_{i_1, i_2, \ldots, i_{2p} \in [k]} P[\pi_{u_1 u_2}(i_1) = i_2 \wedge \ldots \wedge \pi_{u_{2p} u_1}(i_{2p}) = i_1]$$

For multi-indices $U = u_1 u_2 \ldots u_{2p}$ and $I = i_1 i_2 \ldots i_{2p}$ let $E_{U,I}$ be the event $\{\pi_{u_1 u_2}(i_1) = i_2 \wedge \cdots \wedge \pi_{u_{2p} u_1}(i_{2p}) = i_1\}$. For a fixed U, the events $E_{U,I}$ where I takes all possible values consist of a partition of the whole probability space. Therefore with this notation,

$$E[Trace(M^{2p})] = \sum_U a_{u_1 u_2} \ldots a_{u_{2p} u_1} \sum_I P[E_{U,I}] = \sum_U a_{u_1 u_2} \ldots a_{u_{2p} u_1} = Trace(A^{2p})$$

□

Hence, to bound ρ, it suffices to bound $E[Trace(A^{2p})]$. The following lemma can be found in [**?**].

Lemma 26 *Let A^* as above and $p = (2-\varepsilon')log_{d/2}n$ a positive integer. Then*

$$E[Trace((A^*)^{2p})] \leq \frac{1}{n^{1-\varepsilon'}} + 1 + O(\frac{(logn)^4}{n})$$

Claim 27 *Let p be as above. Then for every $\varepsilon > 0$ we have the inequality :*

$$E[\rho] \leq (\frac{2}{d})^{1/4}(1+\varepsilon+o(1))$$

Proof: From claim 25 we have

$$E[Trace((M^*)^{2p})] = E[Trace((A^*)^{2p})]$$

Using lemma 26 we have

$$E[Trace((M^*)^{2p})] \leq (\frac{1}{n^{1-\varepsilon'}} + 1 + O(\frac{(logn)^4}{n}))$$

Hence,

$$E[\rho] \leq (E[Trace((M^*)^{2p})] - 1)^{1/2p} = (E[Trace((A^*)^{2p})] - 1)^{1/(2(2-\varepsilon')log_{d/2}n)}$$

$$\leq (\frac{1}{n^{1-\varepsilon'}})^{\frac{1}{2(2-\varepsilon')log_{d/2}n}}(1+o(1)) = (\frac{2}{d})^{1/4}(1+\varepsilon+o(1))$$

Which follows by the appropriate choice of ε'. □

From the above calculations it follows that if λ is the second largest (in absolute value) eigenvalue of M, then

$$E[\lambda] = O(d^{3/4})$$

We note that it is also possible to bound $E[\lambda]$ by $O(\sqrt{d})$ by using the (more involved) bound on $Trace((A^*)^{2p})$ from [**?**].

3.2.3 Concentration of λ around the mean

We will next prove that with probability that tends to 1 as $n \to \infty$, λ deviates from its mean by at most $\sqrt{d}$. For that we will first prove concentration of λ around its median, and then use elementary probability techniques to show that the expectation and the median of λ are very close. Namely, we will prove the following theorem :

Theorem 28 *The probability that λ_2 deviates from its median by more than t is at most $4e^{-t^2/128}$. The same estimate holds for the probability that λ_{kn} deviates from its median by more than t. Therefore $Pr[|\lambda - \mu(\lambda)| \geq t] \leq 2e^{-t^2/128}$, where $\mu(\lambda)$ denotes the median of λ.*

For that reason, we will use Talagrand's inequality in a similar manner as in [**?**].

Theorem 29 *(Talagrand's Inequality) Let $\Omega_1, \Omega_2, \cdots, \Omega_m$ be probability spaces, and let Ω denote their product space. Let $\mathcal{A}$ and $\mathcal{B}$ be two subsets of Ω and suppose that for each $B = (B_1, \cdots, B_m) \in \mathcal{B}$ there is a real vector $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_m)$ such that for every $A = (A_1, \cdots, A_m) \in \mathcal{A}$ the inequality*

$$\sum_{i:A_i \neq B_i} \alpha_i \geq t(\sum_{i=1}^{m} {\alpha_i}^2)^{1/2}$$

holds. Then

$$Pr[\mathcal{A}]Pr[\mathcal{B}] \leq e^{-t^2/4}$$

We now apply Talagrand's inequality to prove theorem 28. We will show the case for λ_2, but the same proof easily carries out for λ_{kn}. Some notation follows:
Let $\binom{m=n+1}{2}$ and consider the product space Ω of the blocks M_{ij},$1 \leq i, j, \leq n$ where each block is a $k \times k$ permutation matrix. We identify each element of Ω with the vector consisting of the corresponding m $k \times k$ blocks. Instead of i, j we will use indices u, v for the block of M corresponding to vertices u, v. Let μ denote the median of λ_2.
Let $\mathcal{A} = \{M|\lambda_2(M) \leq \mu\}$ and $\mathcal{B} = \{M|\lambda_2(M) \geq \mu + t\}$. By definition of the median, $Pr[\mathcal{A}] \geq 1/2$.

For any vector $f = (f(1), \cdots, f(nk)) \in \mathbb{R}^{nk}$ we will denote by $f_i \in \mathbb{R}^k$,$1 \leq i \leq n$ the vector that corresponds to the i-th block of k coordinates of f, i.e. $f_i = (f((i-1)k), f((i-1)k+1), \cdots, f(ik))$. Let $\|f\|$ be the euclidean norm of f.

Proof:(Of theorem 28) Fix a vector $B \in \mathcal{B}$. Let $f^{(1)}, f^{(2)}$ denote the first and second unit eigenvector of B. We define the following cost vector $\alpha = (a_{uv})$ for B.

$$\alpha_{uu} = (\|f_u^{(1)}\| + \|f_u^{(2)}\|)$$

$$\alpha_{uv} = \sqrt{2\alpha_{uu}\alpha_{vv}}, v \neq u$$

Let $D = \{(u, v)|A_{uv} \neq B_{uv}\}$. We will show that

$$\sum_{(u,v)\in D} \alpha_{uv} \geq c \cdot t \cdot (\sum_{1 \leq u \leq v \leq n} \alpha_{uv}^2)^{1/2}$$

Note that

$$\sum_{1 \leq u \leq v \leq n} \alpha_{uv}^2 = (\sum \alpha_{uu})(\sum \alpha_{vv}) = (\|f^{(1)}\|^2 + \|f^{(2)}\|^2)^2 = 4$$

Let $z = c_1 f^{(1)} + c_2 f^{(2)}$ be a unit vector (i.e. ${c_1}^2 + {c_2}^2 = 1$) which is perpendicular to the first eigenvector of A. Note that such a vector can always be found, since the orthogonality of $f^{(1)}$ and $f^{(2)}$ implies that the subspace $span\{f^{(1)}, f^{(2)}\}$ is 2-dimensional.

Then

$$z^T A z \leq \lambda_2(A) \leq \mu$$

and

$$z^T A z \geq \lambda_2(B) \geq \mu + t$$

which implies

$$\begin{aligned} t \leq z^T(B-A)z \leq \sum_{(u,v)\in D} {z_u}^T(B_{uv}-A_{uv})z_v &\leq \sum_{(u,v)\in D:(B_{uv}-A_{uv})_{ij}\neq 0} |z_{ui}||z_{vj}| \\ &\leq \sum_{(u,v)\in D} \sqrt{2\|z_u\|^2}\sqrt{2\|z_v\|^2} \\ &\leq \sum_{(u,v)\in D} 2\sqrt{\alpha_{uu}}\sqrt{\alpha_{vv}} = \sqrt{2}\sum_{(u,v)\in D}\alpha_{uv} \end{aligned}$$

The fourth inequality holds because each coordinate appears at most twice (each block is a permutation matrix). By combining the above, we obtain

$$\sum_{(u,v)\in D}\alpha_{uv} \geq \frac{t}{4\sqrt{2}}(\sum {\alpha_{vv}}^2)^{1/2} \quad \Longrightarrow \quad Pr[B] \leq 2e^{\frac{-t^2}{128}}$$

□

We conclude by showing that the eigenvalues are also concentrated around their expectation. Namely,

Theorem 30 $Pr[|\lambda - E[\lambda]| \geq t] \leq e^{-(1-o(1))t^2/128}$

To prove this, we show that the expectation and the median of eigenvalues are very close. We show the result for λ_2 but the result holds for all eigenvalues (with different constants in the exponent).

Claim 31 $E[\lambda_2] - \mu \leq 8\sqrt{2\pi}$

Proof:

$$E[\lambda_2] - \mu \leq E[|\lambda_2 - \mu|] = \int_0^\infty P[|\lambda_2 - \mu| > t]dt \leq \int_0^\infty 2e^{\frac{-t^2}{128}}dt = 8\sqrt{2\pi}$$

□

3.3 Recovering Solutions by Spectral Methods

For a given instance of unique games on a graph $G = (V, E)$, let M denote (as before) the $nk \times nk$ symmetric matrix such that the $k \times k$ block M_{uv} is equal to the permutation matrix Π_{uv} if $(u, v) \in E$ and 0 otherwise. We shall now show how the eigenvectors of

M may be used to recover good assignments to highly satisfiable instances of unique games in some special cases.

Specifically, we handle the cases when the instances are random regular graphs with random constraints, and also when the constraints are arbitrary Γ-max-lin instances and the underlying graph has some significant expansion. The properties used in both cases are the eigenvalue gap of the underlying graph and small number of eigenvectors of M with high eigenvalue.

We give the analysis for d-regular graphs in both cases to give a unified treatment. While our arguments for random graphs work give better bounds for regular graphs, the ones for expanding Γ-max-lin instances can easily be generalized to non-regular graphs by considering the eigenvectors of the matrix $D - M$ instead of M. Here D denotes an $nk \times nk$ diagonal matrix with $D_{uu} = deg(u) \cdot I$. If we think of M as the adjacency matrix of graph with vertex set $V \times [k]$ and each edge of G replaced by a matching, then $D - M$ can be thought of as the Laplacian matrix of that graph.

We construct an "almost satisfiable" instance according to the following model, which captures both the cases mentioned above:

- Pick a d-regular graph $G = (V, E)$ according to some distribution $\mathcal{D}_G$.
- To every $u \in V$, assign a value $A(u) \in [k]$.
- For every edge $(u, v) \in E$, pick a constraint π_{uv} *consistent with* $A(u)$ *and* $A(v)$ from some distribution $\mathcal{D}_{uv}$. Let M be tha matrix of this completely satisfiable game. We denote the game by (G, k, M).
- Let an adversary pick any $\varepsilon|E|$ edges and replace their constraints by arbitrary constraints. Let the new matrix be $\hat{M}$ and let $(G, k, \hat{M})$ denote the perturbed game.

The above model captures the random model with planted solutions if we take $\mathcal{D}_G$ to be the distribution over random d-regular graphs and $\mathcal{D}_{uv}$ to be uniform over permutations consistent with $A(u)$ and $A(v)$. The second case can be realized by taking D_G as any arbitrary distribution over graphs with second eigenvalue (say) at most $(1 - \gamma)d$ and $\mathcal{D}_{uv}$ as arbitrary Γ-max-lin constraints.

Let W the span of the eigenvectors of $\hat{M}$ with eigenvalue at least $(1 - 2\varepsilon)d$. The algorithm simply looks at a set $S \subseteq W$ of polynomialy many candidate vectors and "reads-off" an assignment as described below. The set S is chosen differently in each case.

`Recover-Solution`$_S(G, k, \hat{M})$

- For each $x \in S$, construct an assignment A_x by assigning to each vertex u, the index corresponding to the largest entry in the block $(x_{u1}, \ldots, x_{uk})$ i.e. $A(u) = \text{argmax}_i x_{ui}$.
- Out of all assignments A_x for $x \in S$, choose the one satisfying the maximum number of constraints.

To choose S, we will look at the analog of W for the matrix M. Let Y denote the span of eigenvectors of M with eigenvalue at least $(1 - \varepsilon)$. We will first show that if G has a large eigenvalue gap, then every vector in W is close to some vector in Y. We then identify some "nice" vectors in W such that the algorithm works for any vector which close to some

nice vector. We then identify a set $S \subseteq W$ such that at least one vector in S is close to a nice vector.

To show that the eigenspaces W and Y are close, we use the following claim which essentially appears in [**?**] as the sin θ theorem. We give the proof below for self-containment.

Claim 32 *Let w be a unit length eigenvector of $\hat{M}$ with eigenvalue $\hat{\lambda} \geq (1-2\varepsilon)d$ and let λ_s denote the largest eigenvalue of M which is smaller than $(1-2\varepsilon)d$. Then, w can be written as $\alpha y + \beta y_\perp$ with $|\beta| \leq \left\|(M-\hat{M})w\right\| / (\hat{\lambda} - \lambda_s)$*

Proof: We have

$$(M-\hat{M})w \;=\; \alpha My + \beta My_\perp - \hat{\lambda}w \;=\; \alpha(My - \hat{\lambda}y) + \beta(My_\perp - \hat{\lambda}y_\perp)$$

Since $(M-\hat{\lambda}I)y$ and $(M-\hat{\lambda}I)y_\perp$ are in orthogonal eigenspaces, we have

$$\left\|(M-\hat{M})w\right\|^2 = \alpha^2 \left\|(M-\hat{\lambda}I)y\right\|^2 + \beta^2 \left\|(M-\hat{\lambda}I)y_\perp\right\|^2 \geq \beta^2 \left\|(M-\hat{\lambda}I)y_\perp\right\|^2$$

However, $\left\|(M-\hat{\lambda})y_\perp\right\| \geq (\hat{\lambda} - \lambda_s)$ which proves the claim. □

Hence, to prove that the space Y does not change by much due to the perturbation, we simply need to bound $\left\|(M-\hat{M})w\right\|$. We shall also need the fact that w is somewhat "uniform" over each block. To formalize this, let $\overline{w}$ be the n-dimensional vector such that $\overline{w}_u = \|w_u\|$ where w_u is the k-dimensional vector $(w_{u1}, \ldots, w_{uk})^T$. We then show that $\overline{w}$ is very close to the all-one's vector $\mathbf{1}$.

Claim 33 *If w is an eigenvector of $\hat{M}$ with eigenvalue more than $(1-2\varepsilon)d$ and G has second eigenvalue less than $(1-\gamma)d$, then $\overline{w}$ can be written as $a\mathbf{1} + b\mathbf{1}_\perp$ with $|b| \leq \sqrt{\frac{2\varepsilon}{\gamma}}$*

Proof: Since, w corresponds to a large eigenvalue, we have that

$$(1-2\varepsilon)d \;\leq\; (\hat{w})^T \hat{M} \hat{w} \;\leq\; \sum_{u,v} \|w_u\| \, A_{uv} \, \|w_v\| \;=\; (\overline{w})^T A \overline{w}$$

Writing $\overline{w}$ as $\frac{a}{\sqrt{n}}\mathbf{1} + b\mathbf{1}_\perp$, we get

$$\begin{aligned} & (\overline{w})^T A \overline{w} \;\leq\; a^2 d + b^2(1-\gamma)d \\ \implies & (1-2\varepsilon)d \;\leq a^2 d + b^2(1-\gamma)d \;\implies\; |b| \leq \sqrt{\frac{2\varepsilon}{\gamma}} \end{aligned}$$

□

Using the above, and the fact that the matrix M is only perturbed in ε fraction of the edges, we can now bound $\left\|(M-\hat{M})w\right\|$ as follows.

Claim 34 $\left\|(M-\hat{M})w\right\| \leq 5\sqrt{\frac{\varepsilon}{\gamma}}$

PROOF: Define the $n \times n$ matrix R as $R_{uv} = 1$ when the block $(M - \hat{M})_{uv}$ has any non-zero entries, and $R_{uv} = 0$ otherwise. Note that if $(M - \hat{M})_{uv}$ is non-zero, then it must be the difference of two permutation matrices. Thus, for all v $\left\|(M - \hat{M})_{uv} w_v\right\| \leq 2R_{uv} \|w_v\|$. We have that

$$\begin{aligned}\left\|(M - \hat{M})w\right\| &= \sqrt{\sum_u \left\|\sum_v (M - \hat{M})_{uv} w_v\right\|^2} \leq \sqrt{\sum_u \left(\sum_v \left\|(M - \hat{M})_{uv} w_v\right\|\right)^2} \\ &\leq \sqrt{\sum_u \left(\sum_v 2R_{uv} \|w_v\|\right)^2} \\ &\leq 2\|R\overline{w}\|\end{aligned}$$

To estimate $\|R\overline{w}\|$, we break it up as

$$\|R\overline{w}\| \leq \frac{a}{\sqrt{n}} \|R \cdot \mathbf{1}\| + b \|R \cdot \mathbf{1}_\perp\|$$

Since R has at most d 1s in any row, $b\|R \cdot \mathbf{1}_\perp\| \leq \sqrt{\frac{2\varepsilon}{\gamma}} d$. Also, $\|R \cdot \mathbf{1}\| = \sqrt{\sum_u (\sum_v R_{uv})^2}$. Since R has a total of εnd 1s, this expression is maximized when it has d 1s in εn rows. This gives $\frac{1}{\sqrt{n}} \|R \cdot \mathbf{1}\| \leq \sqrt{\varepsilon} d$. Combining with the above, we have that

$$\left\|(M - \hat{M})w\right\| \leq 2\sqrt{\varepsilon} d + 2\sqrt{\frac{2\varepsilon}{\gamma}} d \leq 5\sqrt{\frac{\varepsilon}{\gamma}}$$

□

Combining the above bound with claim 32, we get that any unit-length vector $w \in W$ can be expressed as $\alpha y + \beta y_\perp$ where $y \in Y$ and $|\beta| \leq 5\sqrt{\frac{\varepsilon}{\gamma}} \cdot \frac{1}{(1-2\varepsilon)d - \lambda_s}$. Recall that λ_s was the largest eigenvalue of M smaller than $(1 - 2\varepsilon)d$. We now obtain bounds on λ_s and define the set S of candidate vectors separately for each case.

3.3.1 The Planted solution model on random graphs

Since G is a d-regular graph and each block of M is a permutation matrix, the first eigenvector of M (with eigenvalue d) is the vector $\frac{1}{nk}\mathbf{1}$. It is easy to verify that the following vector y is orthogonal to $\mathbf{1}$ and also has eigenvalue d.

$$y_{ui} = \begin{cases} \frac{k-1}{\sqrt{nk(k-1)}} & \text{if } i = A(u) \\ \frac{-1}{\sqrt{nk(k-1)}} & \text{otherwise} \end{cases}$$

The following claim shows that w.h.p. all other eigenvalues of the matrix M are small and hence y is the only vector orthogonal to $\mathbf{1}$ with eigenvalue more than $(1 - 2\varepsilon)d$.

Claim 35 *With high probability over the choice of M, $\lambda_i(M) \leq O(\sqrt{d})$ for all $i \geq 3$.*

PROOF: Let z be a vector perpendicular to both $\mathbf{1}$ and y such that $\|z\| = 1$. Then, we must have that

$$\sum_u \sum_i z_{ui} = 0 \quad \text{and} \quad \sum_u \left((k-1) z_{uA(u)} - \sum_{i \neq A(u)} z_{ui} \right) = 0$$

which implies

$$\sum_u z_{uA(u)} = \sum_{i \neq A(u)} z_{ui} = 0$$

We now define z_1 as $(z_1)_{ui} = y_{ui}$ for all $i \neq A(u)$ and $(z_1)_{uA(u)} = 0$. Also, let $z_2 = y - y_1$. Then for every u, $(z_2)_{uA(u)} = z_{uA(u)}$ is the only non-zero coordinate of z_2. Also $\|z_1\|, \|z_2\| \leq 1$. We have,

$$\|Mz\| = \|M(z_1 + z_2)\| = \|Mz_1 + Mz_2\| \leq \|Mz_1\| + \|Mz_2\|$$

However, since all constraints are satisfied by the assignment $x_u = A(u)$, $\|Mz_2\| = \left\|Az_2^G\right\|$, where z_2^G is an n-dimensional "projection" of z_2 on the graph by setting $(z_2^G)_u = (z_2)_u$, and A is the adjacency matrix of the graph. From the above equations we have that $\sum_u z_{uA(u)} = 0$, which means that z_2^G is perpendicular to the first eigenvector of A. Thus, w.h.p.

$$\|Mz_2\| = \left\|Az_2^G\right\| \leq O(\sqrt{d} \left\|z_2^G\right\|) \leq O(\sqrt{d})$$

We now consider a new game with matrix M_{k-1} with alphabet size $k-1$ obtained by deleting the value $A(u)$ for each u. Note that this is a completely random unique game for alphabet size $k-1$, since we chose constraints for M randomly after fixing $\pi_{uv}(A(u)) = A(v)$. Finally, it remains to notice that $\|Mz_1\| = \left\|M_{k-1} z_1^{(k-1)}\right\|$, where $z_1^{(k-1)}$ is the $n(k-1)$-dimensional projection of z_1 obtained by deleting coordinates $z_{uA(u)}$ for all u. We also have

$$\sum_{u,i} (z_1^{(k-1)})_{ui} = \sum_{u, i \neq A(u)} z_{ui} = 0$$

which gives that $z_1^{(k-1)}$ is perpendicular to the first eigenvector of M_{k-1} and hence by the previous eigenvalue estimates,

$$\|Mz_1\| = \left\|M_{k-1} z_1^{(k-1)}\right\| \leq O(\sqrt{d})$$

□

From the above, we get that w.h.p. $\lambda_s \leq O(\sqrt{d})$. Also, if the underlying graph G is random, then its second eigenvalue is $O(\sqrt{d})$ and γ is $1 - o(1)$. Combining this with claims 32 and 34, we see that every vector $w \in W$ can be expressed as $w = \alpha y + \beta y_\perp$, with $|\beta| \leq 6\sqrt{\varepsilon}$. Also, this gives $\alpha \geq 1 - 6\sqrt{\varepsilon}$.

To choose S, note that $\hat{M}$ has at least one eigenvector orthogonal to $\mathbf{1}$ with eigenvalue more than $1 - 2\varepsilon$, since $y \perp \mathbf{1}$ and $y^T \hat{M} y \geq (1 - \frac{k-1}{k}\varepsilon)d \geq (1 - 2\varepsilon)d$. Also, the dimension of W can be at most 2 since every unit vector in W must be close to a unit

vector in Y and w.h.p. Y has dimension 2. Let $w \in W$ be the eigenvector of $\hat{M}$ orthogonal to $\mathbf{1}$. We take $S = \{w, -w\}$.

Also, since $w \perp \mathbf{1}$, we can express $w = \alpha y + \beta y_\perp$ with both y and $y_\perp$ orthogonal to $\mathbf{1}$. Then, for one of the vectors w or $-w$, y must be the second eigenvector of the matrix M as described earlier. We now show that the algorithm recovers the corresct assignment to most of the variables.

Claim 36 *Let $w = \alpha y + \beta y_\perp$ with $y_{ui} = (k-1)/\sqrt{nk(k-1)}$ if $i = A(u)$ and $y_{ui} = -1/\sqrt{nk(k-1)}$ otherwise. Then, for ε small enough, the coordinate $w_{uA(u)}$ has the maximum value within its block for at least $(1 - 99\varepsilon)n$ blocks u .*

Proof: Within each block u, in order for coordinate $A(u)$ to be no longer the maximum one, it must happen that for some j

$$\alpha \frac{k-1}{\sqrt{nk(k-1)}} + \beta \cdot (w_\perp)_{uA(u)} \leq -\frac{\alpha}{\sqrt{nk(k-1)}} + \beta \cdot (w_\perp)_{uj}$$

This gives

$$\begin{aligned} & (w_\perp)_{uj} - (w_\perp)_{uA(u)} \geq \frac{k}{\sqrt{nk(k-1)}} \cdot \frac{\alpha}{\beta} \\ \implies \quad & [(w_\perp)_{uj}]^2 + [(w_\perp)_{uA(u)}]^2 \geq \frac{1}{2}[(w_\perp)_{uj} - (w_\perp)_{uA(u)}]^2 \geq \frac{k}{2n(k-1)} \cdot \frac{\alpha^2}{\beta^2} \\ \implies \quad & \|(w_\perp)_u\|^2 \geq \frac{k}{2n(k-1)} \cdot \frac{(1-6\sqrt{\varepsilon})^2}{36\varepsilon} \end{aligned}$$

.

We call such a block "bad". Assume that there are ηn bad blocks. Then

$$1 \geq \sum_{\text{bad } u} \|(w_\perp)_u\|^2 \geq \eta n \cdot \frac{k}{2n(k-1)} \cdot \frac{(1-6\sqrt{\varepsilon})^2}{36\varepsilon} \quad \implies \quad \eta \leq \frac{72\varepsilon}{(1-6\sqrt{\varepsilon})^2} \leq 99\varepsilon$$

□

Therefore, for all but at most 99ε fraction of the blocks, the maximum coordinate remains at the same place. The assigment recovered by our algorithm then fails to satisfy at most $99\varepsilon nd$ constraints corresponding to these blocks and εnd constraints perturbed initially. Thus, the solution violates at most $100\varepsilon nd = 200\varepsilon|E|$ constraints and has value at least $1 - 200\varepsilon$.

3.3.2 Expanding instances of Γ-max-lin

In the case of Γ-max-lin, for each edge (u, v) in the graph G, we have a constraint of the form $x_u - x_v = c_{uv}$, where x_u, x_v are variables taking values in Γ and $c_{uv} \in \Gamma$. Let k denote the size of the group Γ. As before, we consider a matrix $\hat{M}$ for the given instance, and think of it as an adversarial perturbation on ε-fraction of the edges of another matrix

M corresponding to a fully satisfiable instance. Let A be an assignment such that the values $x_u = A(u)$ satisfy all the constraints in the instance corresponding to M.

As in the previous analysis, we assume that the graph is d-regular with second eigenvalue at most $d(1-\gamma)$. We will be able to distinguish instances of Γ-max-lin in which $(1-\varepsilon)$ fraction of the constraints are satisfiable from those in which at most δ fraction of the cosntraints are satisfiable for $\gamma = \Omega(\varepsilon^{1/3})$.

For the matrix M, we define the eigenvectors $y^{(0)}, \ldots, y^{(k-1)}$ as

$$y^{(s)}_{ui} = \begin{cases} \frac{1}{\sqrt{n}} & \text{if} \quad i = A(u) + s \mod k \\ 0 & \text{otherwise} \end{cases}$$

Note that for Γ-max-lin, if $\forall u: \ x_u = A(u)$ is a satisfying assignment, then so is $\forall u: \ x_u = A(u)+s$. Hence, the vectors $y^{(0)}, \ldots, y^{(k-1)}$ correspond to satisfying assignments and are eigenvectors with eigenvalue d for the matrix M. We now show that any eigenvector which is orthogonal to all these vectors has eigenvalue at most $d(1-\gamma)$.

Claim 37 *Let x be a vector such that $x \perp y^{(s)} \forall s$. Then $x^T M x \leq (1-\gamma)d$.*

Proof: Since $x \perp y^{(s)} \forall s$, we have

$$\forall s \in \{0, \ldots, k-1\} \quad \sum_u x_{uA(u)+s} = 0$$

We then decompose x into $x^0, \ldots, x^{(k-1)}$, where

$$x^{(s)}_{ui} = \begin{cases} x_{ui} & \text{if} \quad i = A(u) + s \mod k \\ 0 & \text{otherwise} \end{cases}$$

It is immediate from the definition that $x = \sum_s x^{(s)}$ and that $\|x\|^2 = \sum_s \left\|x^{(s)}\right\|^2$. To bound the eigenvalue corresponding to x, note that

$$x^T M x = \sum_{s,t} (x^{(s)})^T M x^{(t)}$$

Let e_i denote the ith unit vector in k-dimensions. We can then write $x^{(s)}_u$ as $x_{uA(u)+s} e_{A(u)+s}$. Using this notation, we compute the terms in the above equation as

$$(x^{(s)})^T M x^{(t)} = \sum_{(u,v)\in E} (x^{(s)}_u)^T \Pi_{uv} (x^{(t)}_v) = \sum_{(u,v)\in E} x_{uA(u)+s} x_{vA(v)+t} \cdot (e_{A(u)+s})^T \Pi_{uv} e_{A(v)+t}$$

Since the permutation maps $A(u)$ to $A(v)$ and $A(u)+s$ to $A(v)+s$ for all s, $(x^{(s)})^T M x^{(t)} = 0$ when $s \neq t$. For the rest of the terms, we have

$$(x^{(s)})^T M x^{(s)} \ = \ \sum_{(u,v)\in E} x_{uA(u)+s} x_{vA(v)+s} \ \leq \ d(1-\gamma) \left\|x^{(s)}\right\|^2 \qquad (\text{Since } \sum_u x_{uA(u)+s} = 0)$$

Hence,

$$x^T M x \;=\; \sum_s (x^{(s)})^T M x^{(s)} \;\leq\; d(1-\gamma)\sum_s \left\|x^{(s)}\right\|^2 \;=\; d(1-\gamma)\,\|x\|^2$$

□

We take Y to be the span of $y^{(0)},\ldots,y^{(k-1)}$. From the above, we know that the next eigenvalue smaller than $(1-2\varepsilon)d$ for M is $\lambda_s \leq (1-\gamma)d$. Note that for all $s \in \{0,\ldots,k-1\}$, we have $(y^{(s)})^T \hat{M} y^{(s)} \geq d(1-\varepsilon)$. Let w be any unit-length eigenvector of $\hat{M}$, with eigenvalue at least $(1-2\varepsilon)d$. By claims 32 and 34, we can express w as $\sum_s \alpha_s y^{(s)} + \beta y_\perp$ with

$$|\beta| \;\leq\; 5\sqrt{\frac{\varepsilon}{\gamma}}d \cdot \frac{1}{(1-2\varepsilon)d-\lambda_s} \;\leq\; 5\sqrt{\frac{\varepsilon}{\gamma}}\cdot\frac{1}{\gamma-2\varepsilon} \;\leq\; 6\sqrt{\frac{\varepsilon}{\gamma^3}}$$

Note that this also implies that the eigenspace of vectors with eigenvalue greater than $(1-\varepsilon)d$ has dimension at most k (otherwise we would find a vector othogonal to $y^{(0)},\ldots,y^{(k-1)}$ which cannot be close to their span).

Hence, for $\gamma = \Omega(\varepsilon^{1/3})$, the eigenspace of the first k eigenvectors of $\hat{M}$ (W) is close to the eigenspace of the first k eigenvectors of M (Y). Also, Y contains the vectors $y^{(0)},\ldots,y^{(k-1)}$ which encode the solutions. As in claim 36 we can show that the algorithm works for any vector close to one of the vector $y^{(s)}$.

Claim 38 *If x is a vector such that $v = \alpha y^{(s)} + \beta y_\perp$ for some $y^{(s)}$ with $\alpha > 0$, then the coordinate $x_{uA(u)+s}$ is maximum in at least $(1-\frac{\beta^2}{\alpha^2}n)$ blocks.*

Proof: Within each block u, in order for coordinate $A(u)+s$ to be no longer the maximum one, it must happen that for some j

$$\alpha\frac{1}{\sqrt{n}} \leq \beta\cdot(y_\perp)_{uj}$$

However,this gives

$$\|(y_\perp)_u\| \;\geq\; (y_\perp)_{uj}^2 \;\geq\; \frac{\alpha^2}{\beta^2 n}$$

Since $\|y_\perp\| = 1$, this can only happen for at most $\frac{\beta^2}{\alpha^2}n$ blocks. □

To find a vector v close to one of the vectors $y^{(s)}$, we discretize the eigenspace of the first k eigenvectors of $\hat{M}$. Let $w^{(0)},\ldots,w^{(k-1)}$ be the eigenvectors. We define the set S as

$$S \;=\; \left\{ v = \sum_{s=0}^{k-1}\alpha_s w^{(s)} \;\mid\; \alpha_s \in \frac{1}{10\sqrt{k}}\mathbb{Z},\;\; \|v\| \leq 1 \right\}$$

S contains at least one vector v such that $v = \alpha y^{(s)} + \beta y_\perp$ for some s and $\beta \leq 1/10 + 6\sqrt{\varepsilon/\gamma^3} < 1/5$ for $\gamma > 20\varepsilon^{1/3}$. Thus, for this vector v, $\texttt{Recover-Solution}_S(G,k,\hat{M})$ recovers an assignment which agrees with $y^{(s)}$ in $(1-\frac{1}{24})$ fraction of the block. Hence, the assignment violates at most $\frac{1}{24}nd + \varepsilon nd < nd/20$ constaints. Since the total number of constraints is $nd/2$, this satisfies more than 90 percent of the constraints.

Finally, it remains to argue that the running time of the algorithm is polynomial. It can be calculated (see, for instance [**?**]) that the number of points in the set S is at most $e^{k \ln 90}$. Since $k = O(\log n)$ (this must hold for the long-code based reductions to be polynomial time), the number of points is polynomial in n. Hence, the algorithm runs in polynomial time.

3.4 An SDP-based Algorithm

In this section, we present another efficient algorithm for finding a good solution to the Unique Games problem when the constraint graph is an expander. This algorithm is based on a new analysis of the SDP relaxation for Unique Games that appears in figure 3.2, which we repeat below for convenience. The novelty of the algorithm lies in the fact that it involves correlations among distant vertices. The analysis also leads to a parallel repetition theorem for Unique Games when the underlying graph is an expander.

$$\text{Maximize} \quad \mathbf{E}_{e=(u,v)\in E}\mathbf{E}_{i\in[k]} \left\langle \mathbf{u}_i, \mathbf{v}_{\pi_{uv}(i)} \right\rangle \tag{3.9}$$

Subject to

$$\forall\ u \in V \quad \mathbf{E}_{i\in[k]} \left\|\mathbf{u}_i\right\|^2 = 1 \tag{3.10}$$

$$\forall\ u \in V \ \ \forall\ i \neq j \qquad \langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0 \tag{3.11}$$

$$\forall\ u, v \in V \ \ \forall\ i, j \qquad \langle \mathbf{u}_i, \mathbf{v}_j \rangle \geq 0 \tag{3.12}$$

Our proof will use the fact that the objective function (3.9) can be rewritten as

$$1 - \tfrac{1}{2}\mathbf{E}_{e=(u,v)\in E}\mathbf{E}_{i\in[k]} \left\|\mathbf{u}_i - \mathbf{v}_{\pi_{uv}(i)}\right\|^2 \tag{3.13}$$

Let $\mathcal{U} = (G(V,E), [k], \{\pi_{uv}\}_{(u,v)\in E})$ be a Unique Games instance, and let $\{\mathbf{u}_i\}_{u\in V, i\in[k]}$ be an optimal SDP solution. Assume wlog that its value is $1-\varepsilon$, since otherwise we know already that the instance is a NO instance. How do we extract a labeling from the vector solution?

Constraint (3.10) suggests an obvious way to view the vectors corresponding to vertex u as a distribution on labels, namely, one that assigns probability label i to u with probability $\frac{1}{k}\left\|\mathbf{u}_i\right\|^2$. The most naive idea for a rounding algorithm would be to use this distribution to pick a label for each vertex, where the choice for different vertices is made independently. Of course, this doesn't work since all labels could have equal probability under this distribution and thus the chance that the labels i, j picked for vertices u, v in an edge e satisfy $\pi_e(i) = j$ is only $1/k$.

More sophisticated roundings use the fact that if the SDP value is $1-\varepsilon$ for some small ε, then the vector assignments to the vertices of an average edge $e = (u, v)$ are highly correlated, in the sense that for "many" i, $\left\langle \overline{\mathbf{u}}_i, \overline{\mathbf{v}}_{\pi(i)} \right\rangle > 1 - \Omega(\varepsilon)$ where $\overline{\mathbf{u}}_i$ denotes the unit vector in the direction of $\mathbf{u}_i$. This suggests many rounding possibilities as explored in previous papers [**?**, **?**, **?**], but counterexamples [**?**] show that this edge-by-edge analysis can only go so far: high correlation for edges does not by itself imply that a good global assignment exists.

The main idea of this algorithm is to try to understand and exploit correlations in the vector assignments for vertices that are not necessarily adjacent. If u, v are not adjacent vertices we can try to identify the correlation between their vector assignments by noting that since the $\mathbf{v}_j$'s are mutually orthogonal, for each $\mathbf{u}_i$ there is at most one $\mathbf{v}_j$ such that $\langle \overline{\mathbf{u}}_i, \overline{\mathbf{v}}_j \rangle > 1/\sqrt{2}$. Thus we can set up a maximal partial matching among their labels where the matching contains label pairs (i, j) such that $\langle \overline{\mathbf{u}}_i, \overline{\mathbf{v}}_j \rangle > 1/\sqrt{2}$. The vector assignments to the two vertices can be thought of as highly correlated if the sum of squared ℓ_2 norm of all the $\mathbf{u}_i$'s (resp, all $\mathbf{v}_j$'s) involved in this matching is close to k. (This is a rough idea; see precise definition later.)

Our main contribution is to show that if the constraint graph is an expander then high correlation over edges necessarily implies high expected correlation between a randomly-chosen pair of vertices (which may be quite distant in the constraint graph). We also show that this allows us to construct a good global assignment. This is formalized below.

3.4.1 Rounding procedure and correctness proof

Now we describe our randomized rounding procedure $\mathcal{R}$, which outputs a labeling $\Lambda_{\text{alg}} \colon V \to [k]$. This uses a more precise version of the greedy matching outlined in the above overview. For a pair u, v of vertices (possibly nonadjacent), let $\sigma_{uv} \colon [k] \to [k]$ be a bijective mapping that maximizes $\mathbf{E}_{i \in [k]} \left\langle \mathbf{u}_i, \mathbf{v}_{\sigma_{uv}(i)} \right\rangle$; note that it can be efficiently found using max-weight bipartite matching. The procedure is as follows:

1. Pick a random vertex u.
2. Pick a label i for u from the distribution, where every label $i' \in [k]$ has probability $\frac{1}{k} \left\| \mathbf{u}_{i'} \right\|^2$.
3. Define $\Lambda_{\text{alg}}(v) := \sigma_{uv}(i)$ for every vertex $v \in V$.

(Of course, the rounding can be trivially derandomized since there are only nk choices for u, i.)

To analyse this procedure we define the *distance* $\rho(u, v)$ of a pair u, v of vertices as

$$\begin{aligned} \rho(u, v) &:= \frac{1}{2} \mathbf{E}_{i \in [k]} \left\| \mathbf{u}_i - \mathbf{v}_{\sigma_{uv}(i)} \right\|^2 \\ &= 1 - \mathbf{E}_{i \in [k]} \left\langle \mathbf{u}_i, \mathbf{v}_{\sigma_{uv}(i)} \right\rangle \qquad \text{(using (3.10))}. \end{aligned} \tag{3.14}$$

We think of two vertices u and v as highly correlated if $\rho(u, v)$ is small, i.e., $\mathbf{E}_{i \in [k]} \left\langle \mathbf{u}_i, \mathbf{v}_{\sigma_{uv}(i)} \right\rangle \approx 1$.

The following easy lemma shows that if the average vertex pair in G is highly correlated, then the above rounding procedure is likely to produce a good a labeling. Here we assume that G is a regular graph. Using standard arguments, all results can be generalized to the case of non-regular graphs. A proof of the lemma can be found in Section 40.

Lemma 39 (Global Corr. $\implies$ High Value) *The expected fraction of constraints satisfied by the labeling Λ_{alg} computed by the rounding procedure is*

$$\mathbf{E}_{\Lambda_{\mathrm{alg}} \leftarrow \mathcal{R}}[\, \mathsf{val}(\Lambda_{\mathrm{alg}})\,] \geq 1 - 3\varepsilon - 6\mathbf{E}_{u,v \in V}[\, \rho(u,v)\,].$$

It is easy to see that if the SDP value is $1-\varepsilon$ then the average correlation on edges is high. For an edge $e = (u,v)$ in G, let $\varepsilon_e := \frac{1}{2}\mathbf{E}_{i\in[k]} \left\|\mathbf{u}_i - \mathbf{v}_{\pi_{uv}(i)}\right\|^2$. Note, $\mathbf{E}_e[\varepsilon_e] = \varepsilon$. Then we have

$$\begin{aligned}\rho(u,v) = \frac{1}{2}\mathbf{E}_{i\in[k]} \left\|\mathbf{u}_i - \mathbf{v}_{\sigma_{uv}(i)}\right\|^2 &= 1 - \mathbf{E}_{i\in[k]} \left\langle \mathbf{u}_i, \mathbf{v}_{\sigma_{uv}(i)} \right\rangle \\ &\leq 1 - \mathbf{E}_{i\in[k]} \left\langle \mathbf{u}_i, \mathbf{v}_{\pi_{uv}(i)} \right\rangle = \varepsilon_e \\ &\qquad \text{(since } \sigma_{uv} \text{ is a max-weight matching).}\end{aligned}$$

As mentioned in the overview, we show that high correlation on edges implies (when the constraint graph is an expander) high correlation on the average pair of vertices. The main technical contribution in this proof is a way to view a vector solution to the above SDP as a vector solution for Sparsest Cut. This involves mapping any sequence of k vectors to a *single* vector in a nicely continuous way, which allows us to show that the distances $\rho(u,v)$ essentially behave like squared Euclidean distances.

Lemma 40 (Low Distortion Embedding of ρ)
For every positive even integer t and every SDP solution $\{\mathbf{u}_i\}_{u\in V, i\in[k]}$, there exists a set of vectors $\{\mathbf{V}_u\}_{u\in V}$ such that for every pair u, v of vertices

$$\tfrac{1}{2t}\left\|\mathbf{V}_u - \mathbf{V}_v\right\|^2 \leq \rho(u,v) \leq \left\|\mathbf{V}_u - \mathbf{V}_v\right\|^2 + O(2^{-t/2}).$$

Corollary 41 (Local Corr. $\implies$ Global Corr.)

$$\mathbf{E}_{u,v\in V}[\, \rho(u,v)\,] \leq 2t\varepsilon/\lambda + O(2^{-t/2}).$$

Proof: We use the following characterization of λ for regular graphs G

$$\lambda = \min \frac{\mathbf{E}_{(u,v)\in E}\left\|\mathbf{z}_u - \mathbf{z}_v\right\|^2}{\mathbf{E}_{u,v\in V}\left\|\mathbf{z}_u - \mathbf{z}_v\right\|^2}, \tag{3.15}$$

where the minimum is over all sets of vectors $\{\mathbf{z}_u\}_{u\in V}$. This characterization also shows that λ scaled by $n^2/|E|$ is a relaxation for the Sparsest Cut problem $\min_{\emptyset\neq S\subset V} |E(S,\overline{S})|/|S||\overline{S}|$ of G . Now using the previous Lemma we have

$$\begin{aligned}\mathbf{E}_{u,v\in V}[\, \rho(u,v)\,] &\leq \mathbf{E}_{u,v\in V}\left\|\mathbf{V}_u - \mathbf{V}_v\right\|^2 + O(2^{-t/2}) \\ &\leq \tfrac{1}{\lambda}\mathbf{E}_{(u,v)\in E}\left\|\mathbf{V}_u - \mathbf{V}_v\right\|^2 + O(2^{-t/2}) \\ &\leq \tfrac{2t}{\lambda}\mathbf{E}_{(u,v)\in E}[\, \rho(u,v)\,] + O(2^{-t/2}).\end{aligned}$$

□ By combining the Corollary 41 and Lemma 39, we can show the following theorem.

Theorem 42 *There is a polynomial time algorithm that computes a labeling Λ with*

$$\mathsf{val}(\Lambda) \geq 1 - O\left(\tfrac{\varepsilon}{\lambda}\log\left(\tfrac{\lambda}{\varepsilon}\right)\right)$$

if the optimal value of the SDP *in Figure 3.2 for $\mathcal{U}$ is $1-\varepsilon$.*

PROOF: By Corollary 41 and Lemma 39, the labeling Λ_{alg} satisfies a $1 - O(t\varepsilon/\lambda + 2^{-t/2})$ fraction of the constraints of $\mathcal{U}$. If we choose t to be an integer close to $2\log(\lambda/\varepsilon)$, it follows that $\mathsf{opt}(\mathcal{U}) \geq 1 - O(\frac{\varepsilon}{\lambda}\log(\frac{\lambda}{\varepsilon}))$. Since the rounding procedure $\mathcal{R}$ can easily be derandomized, a labeling Λ with $\mathsf{val}(\Lambda) \geq 1 - O(\frac{\varepsilon}{\lambda}\log(\frac{\lambda}{\varepsilon}))$ can be computed in polynomial time. □

We can show that the integrality gap (in terms of expansion) implied above is tight up to a logarithmic factor. The next theorem can be derived using the techniques in [**?**, **?**].

Theorem 43 *For every $\varepsilon > 0$ small enough and for every n large enough, there is a* UNIQUE GAMES *instance $\mathcal{U}_\varepsilon$ on $\Theta(\log(n))$ labels and a constraint graph with $\lambda = \Omega(\varepsilon)$, such that (1) $\mathsf{opt}(\mathcal{U}_\varepsilon) \leq 1/\log^{\varepsilon} n$, and (2) there is an* SDP *solution for $\mathcal{U}_\varepsilon$ of value at least $1 - O(\varepsilon)$.*

The next theorem shows that, assuming UGC, the approximation guarantee of Theorem 42 cannot be improved by more than a constant factor.

Theorem 44 *Assuming* UGC*, for every $\varepsilon, \delta > 0$, there exists $k = k(\varepsilon, \delta)$ such that for a* UNIQUE GAMES *instance $\mathcal{U} = (G(V,E), [k], \{\pi_{uv}\}_{(u,v)\in E})$ it is* NP*-hard to distinguish between*

- *YES Case:* $\mathsf{opt}(\mathcal{U}) \geq 1 - \varepsilon$,
- *NO Case:* $\mathsf{opt}(\mathcal{U}) \leq \delta$ *and* $\lambda > \Omega(\varepsilon)$.

3.4.2 Proof of Lemma 39

We consider the labeling Λ_{alg} computed by the randomized rounding procedure $\mathcal{R}$. Recall that $\Lambda_{\mathrm{alg}}(v) = \sigma_{uv}(i)$ where the vertex u is chosen uniformly at random and the label i is chosen with probability proportional to $\|\mathbf{u}_i\|^2$. For notational ease we assume that σ_{uu} is the identity permutation and σ_{uv} is the inverse permutation of σ_{vu}. The following claim gives an estimate on the probability that the constraint between an edge $e = \{v, w\}$ is satisfied by Λ_{alg}. Here we condition on the choice of u.

Claim 45 *For every vertex u and every edge $e = (v, w)$, $\mathbf{Pr}_{\Lambda_{\mathrm{alg}}}[\Lambda_{\mathrm{alg}}(w) \neq \pi_{v,w}(\Lambda_{\mathrm{alg}}(v)) \mid u] \leq 3 \cdot (\rho(u,v) + \varepsilon_e + \rho(w,u))$.*

PROOF: We may assume that both σ_{uv} and σ_{uw} are the identity permutation. Let $\pi = \pi_{vw}$. First note $\mathbf{Pr}_{\Lambda_{\mathrm{alg}}}[\Lambda_{\mathrm{alg}}(w) \neq \pi(\Lambda_{\mathrm{alg}}(v)) \mid u] = \mathbf{E}_{i\in[k]}\left[\|\mathbf{u}_i\|^2 \chi_{i\neq\pi(i)}\right]$, where $\chi_{\mathcal{E}}$ denotes the indicator random variable for an event $\mathcal{E}$. By orthogonality of the vectors $\{\mathbf{u}_i\}_{i\in[k]}$, it follows that

$$\begin{aligned}\mathbf{E}_{i\in[k]}\left[\|\mathbf{u}_i\|^2 \chi_{i\neq\pi(i)}\right] &\leq \tfrac{1}{2}\mathbf{E}_{i\in[k]}\left[\left(\|\mathbf{u}_i\|^2 + \left\|\mathbf{u}_{\pi(i)}\right\|^2\right)\chi_{i\neq\pi(i)}\right] \\ &= \tfrac{1}{2}\mathbf{E}_{i\in[k]}\left\|\mathbf{u}_i - \mathbf{u}_{\pi(i)}\right\|^2.\end{aligned}$$

By triangle inequality, $\left\|\mathbf{u}_i - \mathbf{u}_{\pi(i)}\right\| \leq \|\mathbf{u}_i - \mathbf{v}_i\| + \left\|\mathbf{v}_i - \mathbf{w}_{\pi(i)}\right\| + \left\|\mathbf{w}_{\pi(i)} - \mathbf{u}_{\pi(i)}\right\|$. Now we square both sides of the inequality and take expectations, $\mathbf{E}_{i\in[k]}\left\|\mathbf{u}_i - \mathbf{u}_{\pi(i)}\right\|^2 \leq 3\mathbf{E}_{i\in[k]}\|\mathbf{u}_i - \mathbf{v}_i\|^2 + 3\mathbf{E}_{i\in[k]}\left\|\mathbf{v}_i - \mathbf{w}_{\pi(i)}\right\|^2 + 3\mathbf{E}_{i\in[k]}\left\|\mathbf{w}_{\pi(i)} - \mathbf{u}_{\pi(i)}\right\|^2 = 6\rho(u,v) + 6\varepsilon_e + 6\rho(w,u)$. □

Proof:[of Lemma 39] From Claim 45 it follows

$$\mathbf{E}_{\Lambda_{\text{alg}}}[\,\mathsf{val}(\Lambda_{\text{alg}})\,] \geq 1 - 3\mathbf{E}_{u\in V}\mathbf{E}_{e=(vw)\in E}\left[\rho(u,v) + \varepsilon_e + \rho(w,u)\right].$$

Since G is a regular graph, both (u,v) and (w,u) are uniformly distributed over all pairs of vertices. Hence $\mathbf{E}_{\Lambda_{\text{alg}}}[\,\mathsf{val}(\Lambda_{\text{alg}})\,] \geq 1 - 3\varepsilon - 6\mathbf{E}_{u,v\in V}[\,\rho(u,v)\,]$. □

3.4.3 Proof of Lemma 40: the tensoring trick

Let t be an integer greater than or equal to 4, and $\{\mathbf{u}_i\}_{u\in V, i\in[k]}$ be an SDP solution for $\mathcal{U}$. Define $\overline{\mathbf{u}}_i = \frac{1}{\|\mathbf{u}_i\|}\mathbf{u}_i$ and $\mathbf{V}_u = \frac{1}{\sqrt{k}}\sum_{i\in[k]}\|\mathbf{u}_i\|\,\overline{\mathbf{u}}_i^{\otimes t}$, where $\otimes t$ denotes t-wise tensoring. Notice that the vectors $\mathbf{V}_u$ are unit vectors. Consider a pair u, v of vertices in G. The following claim implies the lower bound on $\rho(u,v)$ in Lemma 40.

Claim 46 $\|\mathbf{V}_u - \mathbf{V}_v\|^2 \leq t \cdot \mathbf{E}_{i\in[k]}\left\|\mathbf{u}_i - \mathbf{v}_{\sigma_{uv}(i)}\right\|^2$

Proof: Since $\mathbf{V}_u$ is a unit vector for each u, it suffices to prove $\langle \mathbf{V}_u, \mathbf{V}_v\rangle \geq 1 - t\rho(u,v)$. Let $\sigma = \sigma_{uv}$. By Cauchy-Schwarz,

$$\tfrac{1}{k}\textstyle\sum_i \|\mathbf{u}_i\|\|\mathbf{v}_{\sigma(i)}\| \leq \tfrac{1}{k}\big(\sum_i \|\mathbf{u}_i\|^2\big)^{1/2}\big(\sum_i \|\mathbf{v}_{\sigma(i)}\|^2\big)^{1/2} \leq 1.$$

Thus there is some $\alpha \geq 1$ such that the following random variable X is well-defined: it takes value $\langle \overline{\mathbf{u}}_i, \overline{\mathbf{v}}_{\sigma(i)}\rangle$ with probability $\alpha\cdot\frac{1}{k}\|\mathbf{u}_i\|\|\mathbf{v}_{\sigma(i)}\|$. By Jensen's Inequality, $(\mathbf{E}[X])^t \leq \mathbf{E}[X^t]$. Hence,

$$\begin{aligned} 1 - \rho(u,v)t \leq (1-\rho(u,v))^t &= \left(\mathbf{E}_{i\in[k]}\left[\|\mathbf{u}_i\|\|\mathbf{v}_{\sigma(i)}\|\left\langle \overline{\mathbf{u}}_i, \overline{\mathbf{v}}_{\sigma(i)}\right\rangle\right]\right)^t \\ &= (\mathbf{E}[X/\alpha])^t \leq (\mathbf{E}[X])^t/\alpha \\ &\leq \mathbf{E}[X^t/\alpha] = \langle \mathbf{V}_u, \mathbf{V}_v\rangle. \end{aligned}$$

This proves the claim. □

Matching between two label sets. In order to finish the proof of Lemma 40, it remains to prove the upper bound on $\rho(u,v)$ in terms of the distance $\|\mathbf{V}_u - \mathbf{V}_v\|^2$. For this part of the proof, it is essential that the vectors $\mathbf{V}_u$ are composed of (high) tensor powers of the vectors $\mathbf{u}_i$. For a pair u, v of vertices, consider the following set of label pairs

$$M = \{(i,j) \in [k]\times[k] \mid \langle \overline{\mathbf{u}}_i, \overline{\mathbf{v}}_j\rangle^2 > 1/2\}.$$

Since $\{\overline{\mathbf{u}}_i\}_{i\in[k]}$ and $\{\overline{\mathbf{v}}_j\}_{j\in[k]}$ are sets of ortho-normal vectors, M as bipartite graph between the labels for u and the labels for v is a (partial) matching, that is, every label for u has at

most one neighbor among the labels for v. Let σ be an arbitrary permutation of $[k]$ that agrees with the M on the matched labels, i.e., for all $(i,j) \in M$, we have $\sigma(i) = j$. The following claim shows the upper bound on $\rho(u,v)$ of Lemma 40.

Claim 47

$$\frac{1}{2}\mathbf{E}_{i\in[k]} \left\|\mathbf{u}_i - \mathbf{v}_{\sigma(i)}\right\|^2 \leq \frac{1}{2}\left\|\mathbf{V}_u - \mathbf{V}_v\right\|^2 + O(2^{-t/2}).$$

PROOF: Let $\delta = \left\|\mathbf{V}_u - \mathbf{V}_v\right\|^2$. Note that

$$\tfrac{1}{k}\sum_{i,j} \|\mathbf{u}_i\|\|\mathbf{v}_j\| \left\langle \overline{\mathbf{u}}_i, \overline{\mathbf{v}}_j \right\rangle^t = 1 - \delta/2. \tag{3.16}$$

We may assume that σ is the identity permutation. Then, $\rho(u,v)$ is at most

$$\begin{aligned}
\tfrac{1}{2}\mathbf{E}_{i\in[k]} \left\|\mathbf{u}_i - \mathbf{v}_i\right\|^2 &= 1 - \mathbf{E}_{i\in[k]} \left\langle \mathbf{u}_i, \mathbf{v}_i \right\rangle \\
&\leq 1 - \tfrac{1}{k}\sum_{i\in[k]} \|\mathbf{u}_i\| \left\|\mathbf{v}_i\right\| \left\langle \overline{\mathbf{u}}_i, \overline{\mathbf{v}}_i \right\rangle^t \\
&\qquad\qquad \text{(using } \left\langle \mathbf{u}_i, \mathbf{v}_i \right\rangle \geq 0\text{)} \\
&= \delta/2 + \tfrac{1}{k}\sum_{i\neq j} \|\mathbf{u}_i\|\|\mathbf{v}_j\| \left\langle \overline{\mathbf{u}}_i, \overline{\mathbf{v}}_j \right\rangle^t \\
&\qquad\qquad \text{(by (3.16))} \\
&= \delta/2 + \left\langle \mathbf{p}, A\,\mathbf{q} \right\rangle,
\end{aligned}$$

where $p_i = \frac{1}{\sqrt{k}}\|\mathbf{u}_i\|$, $q_j = \frac{1}{\sqrt{k}}\|\mathbf{v}_j\|$, $A_{ii} = 0$, and for $i \neq j$, $A_{ij} = \left\langle \overline{\mathbf{u}}_i, \overline{\mathbf{v}}_j \right\rangle^t$. Since both $\mathbf{p}$ and $\mathbf{q}$ are unit vectors, $\left\langle \mathbf{p}, A\,\mathbf{q} \right\rangle$ is bounded by the largest singular value of A. As the matrix A has only non-negative entries, its largest singular value is bounded by the maximum sum of a row or a column. By symmetry, we may assume that the first row has the largest sum among all rows and columns. We rearrange the columns in such a way that $A_{11} \geq A_{12} \geq \ldots \geq A_{1k}$. Since $\overline{\mathbf{u}}_1$ is a unit vector and $\{\overline{\mathbf{v}}_j\}_{j\in[k]}$ is a set of orthonormal vectors, we have $\sum_j \left\langle \overline{\mathbf{u}}_1, \overline{\mathbf{v}}_j \right\rangle^2 \leq 1$. Hence, $\left\langle \overline{\mathbf{u}}_1, \overline{\mathbf{v}}_j \right\rangle^2 \leq 1/j$ and therefore $A_{1j} \leq (1/j)^{t/2}$. On the other, every entry of A is at most $2^{-t/2}$, since all pairs (i,j) with $\left\langle \overline{\mathbf{u}}_i, \overline{\mathbf{v}}_j \right\rangle^2 > 1/2$ participate in the matching M, and hence, $A_{ij} = 0$ for all i,j with $\left\langle \overline{\mathbf{u}}_i, \overline{\mathbf{v}}_j \right\rangle^2 > 1/2$. It follows that the sum of the first row can be upper bounded by

$$\sum_{j\in[k]} A_{1j} \leq A_{11} + \sum_{j\geq 2}^{\infty} (\tfrac{1}{j})^{t/2} \leq 2^{-t/2} + \sum_{j\geq 2}^{\infty} (\tfrac{1}{j})^{t/2} = O(2^{-t/2}).$$

We conclude that the largest singular value of A is at most $O(2^{-t/2})$, and thus $\rho(u,v)$ can be upper bounded by $\delta/2 + O(2^{-t/2}) = \frac{1}{2}\left\|\mathbf{V}_u - \mathbf{V}_v\right\| + O(2^{-t/2})$, as claimed. □

3.4.4 Stronger relaxations

In this section, we consider stronger SDP relaxations for UNIQUE GAMES and for SPARSEST CUT. A systematic way to obtain stronger relaxations is provided by SDP hierarchies. We choose to state our results in terms of Lasserre's SDP hierarchy [**?**, **?**]. The

results in this section apply only to a special case of UNIQUE GAMES, called ΓMAX2LIN. We say a UNIQUE GAMES instance $\mathcal{U} = (G(V,E), [k], \{\pi_{uv}\}_{(u,v)\in E})$ has ΓMAX2LIN form, if the label set $[k]$ can be identified with the group $\mathbb{Z}_k$ in such a way that every constraint permutation π_{uv} satisfies $\pi_{uv}(i+s) = \pi_{uv}(i) + s \in \mathbb{Z}_k$ for all $s, i \in \mathbb{Z}_k$. In other words, π_{uv} encodes a constraint of the form $x_u - x_v = c_{uv} \in \mathbb{Z}_k$. The ΓMAX2LIN property implies that we can find an optimal SDP solution $\{\mathbf{u}_i\}_{i\in[k]}$ for $\mathcal{U}$ that is *shift-invariant*, i.e., for all $s \in \mathbb{Z}_k$ we have $\langle \mathbf{u}_{i+s}, \mathbf{v}_{j+s}\rangle = \langle \mathbf{u}_i, \mathbf{v}_j\rangle$. In particular, every vector $\mathbf{u}_i$ has unit norm.

Alternative Embedding for ΓMAX2LIN. The following lemma can be seen as alternative to Lemma 40. We emphasize that the lemma only holds for ΓMAX2LIN instances and shift-invariant SDP solutions.

Lemma 48 *Let Λ_{opt} be a labeling for $\mathcal{U}$ with $\mathsf{val}(\Lambda_{\mathrm{opt}}) = 1-\varepsilon$. Then the set of vectors $\{\mathbf{V}_u\}_{u\in V}$ with $\mathbf{V}_u = \mathbf{u}_{\Lambda_{\mathrm{opt}}(u)}$ has the following two properties:*

1. *$\rho(u,v) \le \frac{1}{2}\|\mathbf{V}_u - \mathbf{V}_v\|^2$ for every pair u, v of vertices*
2. *$\frac{1}{2}\mathbf{E}_{(u,v)\in E}\|\mathbf{V}_u - \mathbf{V}_v\|^2 \le \varepsilon + 2\varepsilon$*

Together with Lemma 39, the above lemma implies that the randomized rounding procedure $\mathcal{R}$ computes a labeling that satisfies at least a $1 - O(\varepsilon/\lambda)$ fraction of the constraints of $\mathcal{U}$, whenever $\mathsf{opt}(\mathcal{U}) \ge 1-\varepsilon$. In this sense, the above lemma allows to prove the main result of this paper for the special case of ΓMAX2LIN.

PROOF: Item 1 holds, since, by shift invariance,

$$\begin{aligned}\rho(u,v) = \tfrac{1}{2}\mathbf{E}_{i\in[k]}\left\|\mathbf{u}_i - \mathbf{v}_{\sigma_{uv}(i)}\right\|^2 &= \tfrac{1}{2}\left\|\mathbf{u}_{\Lambda_{\mathrm{opt}}(u)} - \mathbf{v}_{\sigma_{uv}(\Lambda_{\mathrm{opt}}(u))}\right\|^2 \\ &\le \tfrac{1}{2}\left\|\mathbf{u}_{\Lambda_{\mathrm{opt}}(u)} - \mathbf{v}_{\Lambda_{\mathrm{opt}}(v)}\right\|^2.\end{aligned}$$

Here we could assume, again by shift invariance, that $\left\|\mathbf{u}_i - \mathbf{v}_{\sigma_{uv}(i)}\right\|^2 = \min_j \|\mathbf{u}_i - \mathbf{v}_j\|^2$ for all i.

It remains to verify Item 2. By shift invariance,

$$\varepsilon_{uv} = \tfrac{1}{2}\mathbf{E}_{i\in[k]}\left\|\mathbf{u}_i - \mathbf{v}_{\pi_{uv}(i)}\right\|^2 = \tfrac{1}{2}\left\|\mathbf{u}_{\Lambda_{\mathrm{opt}}(u)} - \mathbf{v}_{\pi_{uv}(\Lambda_{\mathrm{opt}}(u))}\right\|^2.$$

Hence, if Λ_{opt} satisfies the constraint on an edge $(u,v) \in E$, then $\frac{1}{2}\|\mathbf{V}_u - \mathbf{V}_v\|^2 = \varepsilon_{uv}$. On the other hand, $\frac{1}{2}\|\mathbf{V}_u - \mathbf{V}_v\|^2 \le 2$ because every vector $\mathbf{V}_u$ has unit norm. Finally, since a $1-\varepsilon$ fraction of the edges is satisfied by Λ_{opt},

$$\mathbf{E}_{(u,v)\in E}\frac{1}{2}\|\mathbf{V}_u - \mathbf{V}_v\|^2 \le (1-\varepsilon)\cdot\mathbf{E}_{(u,v)\in E}[\,\varepsilon_{uv}\,] + \varepsilon\cdot 2.$$

□

Stronger Relaxations for Sparsest Cut. Let r be a positive integer. Denote by $\mathcal{I}$ the set of all subsets of V that have cardinality at most r. For every subset $I \in \mathcal{I}$, we have

$$\text{Minimize} \quad \frac{\mathbf{E}_{(u,v)\in E} \|\mathbf{z}_u - \mathbf{z}_v\|^2}{\mathbf{E}_{u,v\in V}\|\mathbf{z}_u - \mathbf{z}_v\|^2} \tag{3.17}$$

Subject to

$$\forall\ I, J \in \mathcal{I},\ \forall\ I', J' \in \mathcal{I} \qquad \langle \mathbf{x}_I, \mathbf{x}_J\rangle = \langle \mathbf{x}_{I'}, \mathbf{x}_{J'}\rangle \quad \text{if } I \cup J = I' \cup J' \tag{3.18}$$

$$\forall\ u \in V,\ \forall\ i \in [k] \qquad \mathbf{x}_{\{u\}} = \mathbf{z}_u \tag{3.19}$$

$$\|\mathbf{x}_\emptyset\|^2 = 1 \tag{3.20}$$

Figure 3.3: Stronger relaxation for Sparsest Cut.

a variable $\mathbf{x}_I$. We consider a strengthening of the spectral relaxation for Sparsest Cut (Figure 3.3).

The variables $\mathbf{x}_I$ are intended to have values $\mathbf{0}$ or $\mathbf{1}$, where $\mathbf{1}$ is some fixed unit vector. If the intended cut is $(S, \overline{S})$, we would assign $\mathbf{1}$ to all variables $\mathbf{x}_{\{u\}} = \mathbf{z}_u$ with $u \in S$. The variables $\mathbf{x}_I$ are relaxations of boolean variables x_I. The intended value of x_I is the product of the variables x_t, $t \in I$.

Let $z_r(G)$ denote the optimal value of the SDP in Figure 3.3. We have

$$\lambda \leq z_1(G) \leq \ldots \leq z_n(G) = \tfrac{n^2}{|E|} \min_{\emptyset \neq S \subset V} \frac{|E(S,\S)|}{|S||\S|}.$$

It can also be seen that the relaxation $z_3(G)$ is at least as strong as the relaxation for Sparsest Cut considered in [**?**]. The relaxations $z_r(G)$ are inspired by Lasserre's SDP hierarchy [**?**, **?**].

The proof of the following theorem is similar to the proof of Theorem 42. The main difference is that we use Lemma 48, instead of Lemma 40, in order to show that local correlation implies global correlation. By strengthening the SDP for Unique Games, the vectors $\mathbf{V}_u$ obtained from Lemma 48 can be extended to a solution for the stronger SDP for Sparsest Cut in Figure 3.3. This allows us to replace the parameter λ by the parameter $z_r(G)$ in the below theorem.

Theorem 49 *There is an algorithm that computes in time $(kn)^{O(r)}$ a labeling Λ with*

$$\mathsf{val}(\Lambda) \geq 1 - O(\varepsilon / z_r(G))$$

if $\mathsf{opt}(\mathcal{U}) \geq 1 - \varepsilon$ and $\mathcal{U}$ has ΓMAX2LIN form.

3.5 Parallel Repetition for Expanding Unique Games

In this section, we consider *bipartite* unique games, i.e., Unique Games instances $\mathcal{U} = (G(V, W, E), [k], \{\pi_{vw}\}_{(v,w)\in E})$ such that $G(V, W, E)$ is a bipartite graph with bipartition (V, W). A bipartite unique game can be seen as a 2-prover, 1-round proof system [**?**].

The two parts V, W correspond to the two provers. The edge set E corresponds to the set of questions asked by the verifier to the two provers and π_{vw} is the accepting predicate for the question corresponding to the edge (v, w).

In this section, we give an upper bound on the *amortized value* $\overline{\omega}(\mathcal{U}) = \sup_r \mathsf{opt}(\mathcal{U}^{\otimes r})^{1/r}$ of bipartite unique game $\mathcal{U}$ in terms of the expansion of its constraint graph. Here $\mathcal{U}^{\otimes r}$ denotes the game obtained by playing the game $\mathcal{U}$ for r rounds in parallel. We follow an approach proposed by Feige and Lovász [**?**]. Their approach is based on the SDP in Figure 3.4, which is a relaxation for the value of a bipartite unique game. Let $\overline{\sigma}(\mathcal{U})$ denote the value of this SDP relaxation. The following theorem is a consequence of the fact $\overline{\sigma}(\mathcal{U}^{\otimes r}) = \overline{\sigma}(\mathcal{U})^r$.

Theorem 50 ([?]) *For every bipartite unique game $\mathcal{U}$, $\overline{\omega}(\mathcal{U}) \leq \overline{\sigma}(\mathcal{U})$.*

We observe that the SDP in Figure 3.2 cannot be much stronger than the relaxation $\overline{\sigma}(\mathcal{U})$. The proof mostly uses standard arguments.

Lemma 51 *If $\overline{\sigma}(\mathcal{U}) = 1 - \varepsilon$ then the value of the SDP in Figure reffig:SDPtwo is at least $1 - 2\varepsilon$.*

$$\text{Maximize} \quad \mathbf{E}_{(v,w)\in E}\mathbf{E}_{i\in[k]} \left\langle \mathbf{v}_i, \mathbf{w}_{\pi_{vw}(i)} \right\rangle \tag{3.21}$$

Subject to

$$\forall\ v \in V,\ w \in W,\ i, j \in [k] \qquad \langle \mathbf{v}_i, \mathbf{w}_j \rangle \geq 0 \tag{3.22}$$

$$\forall\ v \in V,\ v' \in V \qquad \sum_{i,i'} \left|\langle \mathbf{v}_i, \mathbf{v}'_{i'} \rangle\right| \leq k \tag{3.23}$$

$$\forall\ w \in W,\ w' \in W \qquad \sum_{j,j'} \left|\langle \mathbf{w}_j, \mathbf{w}'_{j'} \rangle\right| \leq k \tag{3.24}$$

Figure 3.4: Feige-Lovasz SDP for Unique Games

Theorem 52 *If $\mathcal{U}$ is 2-prover 1-round unique game on alphabet $[k]$ with value at most $1-\varepsilon$, then the value $\mathcal{U}$ played in parallel for r rounds is at most $(1 - \Omega(\varepsilon \cdot \lambda / \log \frac{1}{\varepsilon}))^r$, where G is the graph corresponding to the questions to the two provers. In particular, the amortized value $\overline{w}(\mathcal{U})$ is at most $1 - \Omega(\varepsilon \cdot \lambda / \log \frac{1}{\varepsilon})$.*

Proof: Following the approach in [**?**], it is sufficient to show $\overline{\sigma}(\mathcal{U}) \leq 1 - \Omega(\varepsilon\lambda / \log \frac{1}{\varepsilon})$. Suppose that $\overline{\sigma}(\mathcal{U}) = 1 - \varepsilon$. Then by Lemma 51, the value of the SDP in Figure 3.2 is at least $1 - 2\varepsilon$. By Theorem 42, it follows that $\mathsf{opt}(\mathcal{U}) \geq 1 - O(\varepsilon \log \frac{\lambda}{\varepsilon} / \lambda)$. On the other hand, $\mathsf{opt}(\mathcal{U}) \leq 1 - \varepsilon$. Hence, $\varepsilon = O(\varepsilon \log \frac{\lambda}{\varepsilon} / \lambda)$ and therefore $\varepsilon = \Omega(\lambda\varepsilon \log \frac{1}{\varepsilon})$, as claimed. □

3.6 Conclusion and Future Directions

Resolving the Unique Games conjecture is perhaps one of the most notorious and important open problems in complexity theory. However, it is shown that any semidefinite

programming approach is doomed to fail in the general case, since there are *integrality gap instances* where the value of the SDP deviates a lot from the value of the game ([**?**], [**?**]). The following important question comes up:

Question 53 *Is there a polynomial time spectral algorithm that finds a highly satisfying assignment in case there exists one, for the general case of Unique Games?*

Even without aiming at resolving the Conjecture, there are several interesting questions to be asked: What are other cases of constraint graphs where we can design an efficient algorithm like in the expander case? For example, could we come up with such an algorithm for hypercube graphs or quotients of them?

Chapter 4

Subgraph Sparsification and Applications

In this chapter we consider a variation of the spectral sparsification problem where we are required to keep a subgraph of the original graph. Formally, given a union of two weighted graphs G and W and an integer k, we are asked to find a k-edge weighted graph W_k such that $G+W_k$ is a good spectral sparsifer of $G+W$. We will refer to this problem as the subgraph (spectral) sparsification. We present a nontrivial condition on G and W such that a good sparsifier exists and give a polynomial time algorithm to find the sparsifer.

As a significant application of our technique, we show that for each positive integer k, every n-vertex weighted graph has an $(n-1+k)$-edge spectral sparsifier with relative condition number at most $\frac{n}{k}\log n\,\tilde{O}(\log\log n)$ where $\tilde{O}()$ hides lower order terms. Our bound is within a factor of $\tilde{O}(\log\log n)$ from optimal. This nearly settles a question left open by Spielman and Teng about ultrasparsifiers, which is a key component in their nearly linear-time algorithms for solving diagonally dominant symmetric linear systems.

We also present another application of our technique to spectral optimization in which the goal is to maximize the algebraic connectivity of a graph (e.g. turn it into an expander) with a limited number of edges.

The chapter is organized as follows. In section 4.2, we give a deterministic polynomial time algorithm for constructing W'. The algorithm follows the lines of [**?**]and relies heavily on algebraic techniques. In sections 4.4 and 4.3 we give two significant applications of the *"subgraph sparsification"* result. In section 4.3, we show how to use the above result in order to construct nearly-optimal *ultrasparsifiers* for every graph. In section 4.4, we apply the result to the problem of finding k edges to add to a graph G in order to maximize its algebraic connectivity. In [**?**] the problem was proved to be APX-Hard and the question of finding an approximation algorithm for it has been heavily studied in the past. We give two approximation algorithms for it that use the results of section 4.2 and which, in several interesting cases, have very good performance guarantee.

4.1 Notation and Preliminaries

Graph Sparsifiers. A sparsifier of a graph $G = (V, E, w)$ is a d-sparse graph H that is similar to G in some useful way. (We say that a graph is d-sparse if it has at most dn edges). Many notions of similarity have been considered. For example, Chew's [**?**] spanners have the property that the distance between every pair of vertices in H is approximately the same as in G. Benczur and Karger's [**?**] cut-sparsifiers have the property that the weight of the boundary of every set of vertices is approximately the same in G as in H. We will mainly be interested in the spectral notion of similarity introduced by Spielman and Teng [**?**], [**?**]: we say that H is a κ-approximation of G if for all $x \in \mathbf{R}^V$,

$$x^T \mathcal{L}_G x \leq x^T \mathcal{L}_H x \leq \kappa x^T \mathcal{L}_G x \tag{4.1}$$

where $\mathcal{L}_G$ and $\mathcal{L}_H$ are the Laplacian matrices of G and H.

Equivalently, for such $H \subseteq G$ we will use the notation $H \preceq G \preceq \kappa H$ to imply that equation 4.1 holds.

In the case where G is the complete graph, excellent spectral sparsifiers are supplied by Ramanujan Graphs [**?**], [**?**]. These are d-regular graphs H all of whose non-zero Laplacian eigenvalues lie between $d - 2\sqrt{d-1}$ and $d + 2\sqrt{d-1}$. Thus, if we take a Ramanujan graph on n vertices and multiply the weight of every edge by $n/(d - 2\sqrt{d-1})$, we obtain a graph that $\frac{d+2\sqrt{d-1}}{d-2\sqrt{d-1}}$-approximates the complete graphs

In [**?**] the authors showed that every graph can be approximated at least this well by a graph with only twice as many edges. Namely, they showed that

Theorem 54 *For every $d > 1$, every undirected graph $G = (V, E, w)$ on n vertices contains a weighted subgraph $H = (V, F, \tilde{w})$ with $\lceil d(n-1) \rceil$ edges (i.e. average degree at most $2d$) that satisfies:*

$$x^T \mathcal{L}_G x \leq x^T \mathcal{L}_H x \leq \frac{d + 2\sqrt{d-1}}{d - 2\sqrt{d-1}} \cdot x^T \mathcal{L}_G x$$

Ultrasparsifiers. We say that a graph is k-ultra-sparse if it has at most n-1+k edges. We note that a spanning tree is 0-ultra-sparse.

An *ultra-sparsifier* of a graph $G = (V, E, w)$ is a d-sparse graph $U \subseteq G$ that approximates G in some useful way. In this chapter, we will use the notion of ultrasparsifiers as it appears in [**?**]. Namely, U is a (κ, N)- ultrasparsifier of G if it has the following properties

- $U \preceq G \preceq \kappa \cdot U$
- U has less than $n - 1 + N$ edges.

4.2 Matrix Sparsifiers

In this section, we prove an analog of the sparsification theorem of Batson, Spielman, and Srivastava [**?**].

Definition 55 *(**Graph Patch**) Let G be a (weighted) graph. A graph W on the vertices of G is a (k, T, λ^*)-patch for G if the following properties hold*[1]*,*

1. $\lambda_{k+1}(\mathcal{L}_G \mathcal{L}^\dagger_{G+W}) \equiv \lambda_{k+1}((\mathcal{L}^\dagger_{G+W})^{1/2} \mathcal{L}_G (\mathcal{L}^\dagger_{G+W})^{1/2}) \geq \lambda^*$;
2. $\mathrm{tr}(\mathcal{L}_W \mathcal{L}^\dagger_{G+W}) \leq T$.

We prove that for every patch, there exists a "patch sparsifier" supported on $O(k)$ edges. Specifically, we prove the following theorem.

Claim 56 *Let $W = (V, E_W, \{w_e\}_{e \in E_W})$ be a (k, T, λ^*)-patch for G with edge weights w_e and $N \geq 8k$. Then there is a weighted graph $W_k = (V, E_{W_k}, \{\tilde{w}_e\}_{e \in E_{W_k}})$ with edge weights $\tilde{w}_e$ such that*

1. *W_k has at most N edges; $E_{W_k} \subseteq E_W$.*
2. *$c_1 \min(N/T, 1)\lambda^* \mathcal{L}_{G+W} \preceq \mathcal{L}_{G+W_k} \preceq c_2 \mathcal{L}_{G+W}$, for some absolute constants c_1 and c_2.*
3. *The total weight of edges, $\sum_{e \in E_{W_k}} \tilde{w}_k$, is at most $\min(1, N/T) \sum_{e \in E_W} w_e$.*

We say that W_k is a patch sparsifier of W with respect to G.

The claim will follow immediately from the following theorem, which is is of independent interest. We will also show another (related) application of this theorem in Section 4.4.

Theorem 57 *Suppose we are given a positive definite $n \times n$ matrix X and a sequence of matrices $Y_i = v_i v_i^T$ $(i = 1, \ldots, m)$ with*

$$X + \sum_{i=1}^{m} Y_i = M^*,$$

and $\lambda_{\max}(M^*) \leq 1$. Additionaly, suppose each matrix Y_i has cost $cost_i$ and $\sum_{i=1}^m cost_i = 1$. Let $\lambda^* = \lambda_{k+1}(X)$, and $T = \lceil \mathrm{tr}(M^* - X) \rceil$. Then for every $N > 8k$ there exists a set of weights w_i with $|\{w_i : w_i \neq 0\}| = N$ such that the matrix $M = X + \sum_{i=1}^m w_i Y_i$ satisfies,

$$c_1 \min(N/T, 1) \cdot \lambda^* \cdot \lambda_{\min}(M^*) \leq \lambda_{\min}(M) \leq \lambda_{\max}(M) \leq c_2,$$

where c_1 and c_2 are some absolute constants, and $\sum_{i=1}^m w_i cost_i \leq \min(1, N/T)$.

Proof Overview. Our proof closely follows the approach of Batson, Spielman, and Srivastava [**?**]. We construct matrix M in N steps; at each step we choose an index i and weight w_i and add $w_i Y_i$ to the sum $X + \sum_{i=1}^m w_i Y_i$. Recall that Batson, Spielman, and Srivastava define two "barriers" l and u and maintain the property that all eigenvalues of M lie between l and u. At each step, they increase l and u and update matrix M so that

[1] we have $\lambda_{k+1}(\mathcal{L}_G \mathcal{L}^\dagger_{G+W}) = \lambda_{k+1}((\mathcal{L}^\dagger_{G+W})^{1/2} \mathcal{L}_G (\mathcal{L}^\dagger_{G+W})^{1/2})$, since $\lambda_i(AB) = \lambda_i(BA)$ for every two square matrices A and B

this property still holds. Finally, the ratio between u and l becomes very close to 1, which means that $\lambda_{\min}(M)$ is very close to $\lambda_{\max}(M)$. During this process, they keep track not only of the smallest and largest eigenvalues of M but of all n eigenvalues to avoid accumulation of eigenvalues in neighborhoods of l and u. To this end, they define two potential functions, the lower potential function $\Phi_l(M) = \sum_{i=1}^{n} \frac{1}{\lambda_i(M)-l}$ and the upper potential function $\Phi^u(M) = \sum_{i=1}^{n} \frac{1}{u-\lambda_i(M)}$, and then ensure that $\Phi_l(M)$ and $\Phi^u(M)$ do not increase over time. That guarantees that all eigenvalues of M stay far away from l and u.

In our proof, however, we cannot keep an eye on all eigenvalues. After each step, only one eigenvalue increases, and thus we need $\theta(n)$ steps to increase all eigenvalues participating in the definition of $\Phi_l(M)$. But our goal is to "patch" X in roughly k steps. So we focus our attention only on k smallest and T largest eigenvalues.

Let S be the eigenspace of X corresponding to k smallest eigenvalues, and P_S be the projection onto S. We define the lower potential function as follows,

$$\Phi_l(A) = \operatorname{tr}(P_S(A - lI)P_S)^\dagger = \sum_{i=1}^{k} \frac{1}{\lambda_i(A|_S) - l},$$

where $A|_S$ denotes the restriction of A to the space S ($A|_S$ is a $k \times k$ matrix). Note that the space S is fixed, and the eigenvector corresponding to the smallest eigenvalue will not necessarily lie in S after a few steps. We want to ensure that after N steps,

$$\sum_{i=1}^{m} w_i Y_i\big|_S \succeq c\min(N/T, 1) \sum_{i=1}^{m} Y_i\big|_S = c\min(N/T, 1)(M^* - X)\big|_S,$$

or in other words, $\lambda_{\min}((Z(\sum_{i=1}^{m} w_i Y_i)Z)|_S) \geq c\min(N/T, 1)$, where $Z = \left((P_S(M^* - X)P_S)^\dagger\right)^{1/2}$. To this end, we show how to update M and l so that $\Phi_l(Z(\sum_{i=1}^{m} w_i Y_i)Z)$ does not increase, and l equals $c\min(N/T, 1)$ after N steps. It remains to lower bound $\lambda_{\min}(M)$ in the entire space. We know that all eigenvalues of X (and therefore, of M) in $S^\perp$ are at least λ^*. We show that that together with an upper bound on $\lambda_{\max}(M)$ implies that $\lambda_{\min}(M) \geq c_1 \min(N/T, 1) \cdot \lambda^* \lambda_{\min}(M^*)$ (the product of the lower bounds on $\lambda_{\min}$ in spaces S and $S^\perp$ divided by the upper bound on $\lambda_{\max}$).

Similarly, we amend the definition of the upper potential function. Since we need to bound $\lambda_{\max}$ in the entire space, we cannot restrict $\Phi^u(M)$ to a fixed subspace. For a matrix A, we consider the eigenspace of A corresponding to its largest T eigenvalues. Denote it by $L_A(A)$; denote the projection onto $L(A)$ by $P_{L(A)}$. Then

$$\Phi^u(A) = \operatorname{tr}(P_{L(A)}(uI - A)^{-1}P_{L(A)}) = \operatorname{tr}(P_{L(A)}(uI - A)P_{L(A)})^\dagger = \sum_{i=n-T+1}^{N} \frac{1}{u - \lambda_i(A)}.$$

Note that both definitions of $\Phi^u(A)$ — in terms of regular inverse and in terms of pseudoinverse — are equivalent since $L(A)$ is an invariant subspace of A. However, $\Phi_l(A)$ is not equal to $\operatorname{tr}(P_S(A - lI)^{-1}P_S)$ in general since S is not necessarily an invariant subspace of A.

Our algorithm and analysis are similar to those of Batson, Spielman, and Srivastava [**?**]. However, several complications arise because we are controlling eigenvalues in different subspaces and, moreover, one of these subspaces, $L(A)$, is not fixed.

Let us summarize the proof. We construct the matrix M iteratively in N steps. Let $A^{(q)}$ be the matrix and $w_i^{(q)}$ be the weights after q steps. We define an auxiliary matrix $B^{(q)}$ as $Z(A^{(q)} - X)Z$. We have,

$$A^{(q)} = X + \sum_i w_i^{(q)} Y_i; \quad B^{(q)} = \sum_i w_i^{(q)} Z Y_i Z = Z(A^{(q)} - X)Z.$$

We will ensure that the following properties hold after each step (for some values of constants l_0, δ_L, u_0, δ_U, ε_L, ε_U, which we will specify later).

1. $\Phi_{l_0}(B^{(0)}) \le \varepsilon_L$ and $\Phi^{u_0}(A^{(0)}) \le \varepsilon_U$.

2. Each matrix $A^{(q)}$ and $B^{(q)}$ is obtained by a rank-one update of the previous one:

   $$A^{(q+1)} = A^{(q)} + tY_i, \quad B^{(q+1)} = B^{(q)} + tZY_iZ$$

   for some i.

3. Lower and upper potentials do not increase. Namely, for every $q = 0, 1, \ldots, N$,

   $$\Phi^{u_0+(q+1)\delta_U}(A^{(q+1)}) \le \Phi^{u_0+q\delta_U}(A^{(q)}) \le \varepsilon_U \text{ and } \Phi_{l+(q+1)\delta_L}(B^{(q+1)}) \le \Phi_{l_0+q\delta_L}(B^{(q)}) \le \varepsilon_L.$$

4. At each step q, $\lambda_{\min}(B^{(q)}\big|_S) > l \equiv l_0 + q\delta_L$ and $\lambda_{\max}(A^{(q)}) < u \equiv u_0 + q\delta_U$. In particular, this condition ensures that all terms in the definitions of upper and lower potentials are positive.

5. At each step q, the total cost is at at most $q/\max(N, T)$: $\sum w_i^{(q)} cost_i \le q/\max(N, T)$.

We present the complete proof in Sections 4.2.2 and 4.2.3. In Section 4.2.2, we first find conditions under which we can update $A^{(q)}$ and u (Lemma 64), and $B^{(q)}$ and l (Lemma 65). Then we show that both conditions can be simultaneously satisfied (Lemma 66). In Section 4.2.1, we prove several theorems that we need later to deal with a non-fixed subspace $L(A)$. Finally, in Section 4.2.3, we combine all pieces of the proof together.

4.2.1 Some Basic Facts about Matrices

Sherman–Morrison Formula

We use the Sherman–Morrison Formula, which describes the behavior of the inverse of a matrix under rank-one updates. We first state the formula for regular inverse [**?**], and then we show that a similar expression holds for the pseudoinverse.

Lemma 58 (Sherman–Morrison Formula) *If A is a nonsingular $n \times n$ matrix and $Y = vv^T$ is a rank-one update, then*

$$(A+Y)^{-1} = A^{-1} - \frac{A^{-1}YA^{-1}}{1 + A^{-1} \bullet Y}$$

Lemma 59 *If A is a symmetric (possibly singular) $n \times n$ matrix, $Y = vv^T$ is a rank-one update, then*

$$(A+PYP)^\dagger = A^\dagger - \frac{A^\dagger Y A^\dagger}{1 + A^\dagger \bullet Y},$$

where P is the orthogonal projection on $\mathrm{Im}(A)$.

Proof: Let $\overline{v} = Pv$ and $\overline{Y} = PYP = \overline{v}\overline{v}^T$. Note that $A^\dagger Y A^\dagger = A^\dagger \overline{Y} A^\dagger$, since $PA^\dagger = P$, and

$$A^\dagger \bullet \overline{Y} = \mathrm{tr}\, A^\dagger \overline{Y} = \mathrm{tr}\, A^\dagger (PYP) = \mathrm{tr}(PA^\dagger P)Y = A^\dagger \bullet Y.$$

We need to verify that

$$(A+\overline{Y})\left(A^\dagger - \frac{A^\dagger \overline{Y} A^\dagger}{1 + A^\dagger \bullet \overline{Y}}\right) = \left(A^\dagger - \frac{A^\dagger \overline{Y} A^\dagger}{1 + A^\dagger \bullet \overline{Y}}\right)(A+\overline{Y}) = P.$$

Since A is a symmetric matrix, $AA^\dagger = A^\dagger A = P$. Since $P^2 = P$, $P\overline{Y}P = \overline{Y}$ and $\overline{Y}A^\dagger\overline{Y} = \overline{v}\overline{v}^T A \overline{v}\overline{v}^T = \overline{v}(A \bullet \overline{Y})\overline{v}^T = (A \bullet \overline{Y})\overline{Y}$. We calculate,

$$\begin{aligned}(A+\overline{Y})\left(A^\dagger - \frac{A^\dagger \overline{Y} A^\dagger}{1 + A^\dagger \bullet \overline{Y}}\right) &= AA^\dagger + \overline{Y}A^\dagger - \frac{\overline{P}}{YA^\dagger + \overline{Y}A^\dagger\overline{Y}A^\dagger} 1 + A^\dagger \bullet \overline{Y} \\ &= P + \overline{Y}A^\dagger - \frac{(1 + A^\dagger \bullet \overline{Y})\overline{Y}A^\dagger}{1 + A^\dagger \bullet \overline{Y}} = P + \overline{Y}A^\dagger - \overline{Y}A^\dagger = P.\end{aligned}$$

Similarly,

$$(A^\dagger - \frac{A^\dagger \overline{Y} A^\dagger}{1 + A^\dagger \bullet \overline{Y}})(A+\overline{Y}) = P.$$

□

Majorization

Lemma 60 *(Majorization) For every positive semidefinite matrix A, every projection matrix P, and every $r \in \{1, \ldots, n\}$*

$$\sum_{i=n-r+1}^{n} \lambda_i(A) \geq \sum_{i=n-r+1}^{n} \lambda_i(PAP). \tag{4.2}$$

In particular, $\lambda_{\max}(A) \geq \lambda_{\max}(PAP)$.

Proof: Let $e_1, \ldots, e_n$ be an orthonormal eigenbasis of A so that e_i has eigenvalue $\lambda_i(A)$. Similarly, let $\tilde{e}_1, \ldots, \tilde{e}_n$ be an orthonormal eigenbasis of PAP so that $\tilde{e}_i$ has eigenvalue

$\lambda_i(PAP)$. Write

$$\tilde{e}_i = \sum_{j=1}^{n} \langle e_j, \tilde{e}_i \rangle e_j.$$

Note that if $\lambda_i(PAP) \neq 0$ then $\tilde{e}_i \in \mathrm{Im}(PAP) \subseteq \mathrm{Im}(P)$ and $P\tilde{e}_i = \tilde{e}_i$. Then

$$\lambda_i(PAP) = \tilde{e}_i^T PAP\tilde{e}_i = \tilde{e}_i A\tilde{e}_i = \sum_{j=1}^{n} \langle e_j, \tilde{e}_i \rangle^2 \lambda_j(A).$$

If $\lambda_i(PAP) = 0$ then trivially

$$\lambda_i(PAP) = 0 \leq \sum_{j=1}^{n} \langle e_j, \tilde{e}_i \rangle^2 \lambda_j(A).$$

Therefore,

$$\sum_{i=n-r+1}^{n} \lambda_i(PAP) \leq \sum_{i=n-r+1}^{n} \sum_{j=1}^{n} \langle e_j, \tilde{e}_i \rangle^2 \lambda_j(A) = \sum_{j=1}^{n} \left(\sum_{i=n-r+1}^{n} \langle e_j, \tilde{e}_i \rangle^2 \right) \lambda_j(A).$$

That is, $\sum_{i=n-r+1}^{n} \lambda_j(PAP)$ is at most the sum of $\lambda_j(A)$ with weights $\sum_{i=n-r+1}^{n} \langle e_j, \tilde{e}_i \rangle^2$. The total weight of all $\lambda_1(A), \ldots, \lambda_n(A)$ is r:

$$\sum_{i=n-r+1}^{n} \underbrace{\sum_{j=1}^{n} \langle e_j, \tilde{e}_i \rangle^2}_{\|\tilde{e}_i\|^2} = \sum_{i=n-r+1}^{n} \|\tilde{e}_i\|^2 = r.$$

The weight of each eigenvalue $\lambda_j(A)$ in the sum is at most 1:

$$\sum_{i=n-r+1}^{n} \langle e_j, \tilde{e}_i \rangle^2 \leq \sum_{i=1}^{n} \langle e_j, \tilde{e}_i \rangle^2 = 1.$$

Therefore, the sum does not exceed the sum of the r largest eigenvalues $\sum_{i=n-r+1}^{n} \lambda_r(A)$. □

Corollary 61 *For every positive semidefinite matrix A, every projection matrix P and $u > \lambda_{\max}(A)$, the following inequality holds.*

$$\Phi^u(PAP) = \sum_{i=n-T+1}^{n} \frac{1}{u - \lambda_i(PAP)} \leq \sum_{i=n-T+1}^{n} \frac{1}{u - \lambda_i(A)} = \Phi^u(A) \tag{4.3}$$

Proof: The statement follows from the Karamata Majorization Inequality. The inequality claims that for every two non-increasing sequences that satisfy (4.2) and for every increasing

convex function f,

$$\sum_{i=n-k+1}^{n} f(\lambda_i(A)) \geq \sum_{i=n-k+1}^{n} f(\lambda_i(PAP)).$$

Plugging in $f(x) = \frac{1}{u-x}$ (defined on $(0, u)$), we obtain the desired inequality. □

Lemma 62 *Let A be a positive semidefinite matrix such that $A \preceq I_n$. Assume $\mathrm{Tr}(A) \leq r \in \mathbb{N}$. Then for every positive semidefinite matrix M, $A \bullet M \leq \sum_{i=N-r+1}^{N} \lambda_i(M)$.*

PROOF: By von Neumann's inequality [**?**], $A \bullet M = \mathrm{tr}(AM) \leq \sum_{i=1}^{n} \lambda_i(A)\lambda_i(M)$. Since $\sum_{i=1}^{n} \lambda_i(A) \leq r$ and all $\lambda_i(A) \leq 1$, we can easily see that the above product achieves its maximum when the largest r eigenvalues of A are 1 and the rest are 0. In this case, we have, $A \bullet M \leq \sum_{i=1}^{n} \lambda_i(A)\lambda_i(M) = \sum_{i=n-r+1}^{n} \lambda_i(M)$. □

As a corollary we get the following result.

Corollary 63 *Let X, M^* and T be as in Theorem 57. Then for any positive semidefinite matrix U, we have $U \bullet (M^* - X) \leq \sum_{i=n-T+1}^{n} \lambda_i(U)$.*

4.2.2 Barrier Shifts

In this section, we analyze how we can update matrices $A^{(q)}$ and $B^{(q)}$, and increment barriers l and r so that the upper and lower potentials do not increase. Let us think of $\Phi^u(A)$ as a function of an n^2 dimensional vector (consisting of entries of A). Then in the first approximation $\Phi^{u+\delta_U}(A+tY) \approx \Phi^{u+\delta_U}(A)+tY \bullet U$, where U is the gradient of $\Phi^{u+\delta_U}$ at A (U is an $n \times n$ matrix). Thus the potential function does not increase, $\Phi^{u+\delta_U}(A+tY) \leq \Phi^u(A)$, roughly when $tY \bullet \frac{U}{\Phi^u(A)-\Phi^{u+\delta_U}(A)} \leq 1$. Similarly, $\Phi_{l+\delta_L}(B + tY) \leq \Phi_l(B)$, roughly when $tY \bullet \frac{L}{\Phi_{l+\delta_L}(B)-\Phi_l(B)} \geq 1$, where L is the gradient of $\Phi_{l+\delta_L}$ at B. Following [**?**], we make these statements precise (we need to take into account lower order terms). We define matrices U_A and L_B,

$$U_A = \frac{((u + \delta_U)I - A)^{-2}}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + ((u + \delta_U)I - A)^{-1};$$
$$L_B = \frac{(P_S(B - (l + \delta_L)I)P_S)^{\dagger 2}}{\Phi_{l+\delta_L}(B) - \Phi_l(B)} - (P_S(B - (l + \delta_L)I)P_S)^{\dagger}$$

Lemma 64 *(Upper Barrier Shift) Suppose $\lambda_{max}(A) < u$ and $Y = vv^T$ is a rank-one update. If $U_A \bullet Y \leq \frac{1}{t}$ then $\Phi^{u+\delta_U}(A + tY) \leq \Phi^u(A)$ and $\lambda_{\max}(A + tY) < u + \delta_U$.*

PROOF: Let $u' = u + \delta_U$ and $P = P_{L(A+tY)}$. By the Sherman–Morrison formula (Lemma (58)), we can write the updated potential as:

$$\begin{aligned}
\Phi^{u+\delta_U}(A+tY) &= \operatorname{tr} P(u'I - A - tY)^{-1}P = \operatorname{tr} P\left((u'I-A)^{-1} + \frac{t(u'I-A)^{-1}Y(u'I-A)^{-1}}{1-t(u'I-A)^{-1}\bullet Y}\right)P \\
&= \operatorname{tr} P(u'I-A)^{-1}P + \operatorname{tr}\frac{tP(u'I-A)^{-1}Y(u'I-A)^{-1}P}{1-t(u'I-A)^{-1}\bullet Y} \\
&\le \Phi^{u+\delta_U}(PAP) + \frac{t(u'I-A)^{-2}\bullet Y}{1-t(u'I-A)^{-1}\bullet Y} \\
&\le \Phi^{u+\delta_U}(A) + \frac{t(u'I-A)^{-2}\bullet Y}{1-t(u'I-A)^{-1}\bullet Y} \\
&= \Phi^{u}(A) - (\Phi^{u}(A) - \Phi^{u+\delta_U}(A)) + \frac{(u'I-A)^{-2}\bullet Y}{1/t-(u'I-A)^{-1}\bullet Y}
\end{aligned}$$

Here, we used Corollary 61 for the inequality on line 4.

Substituting $U_A \bullet Y \le 1/t$ gives $\Phi^{u+\delta_U}(A+tY) \le \Phi^u(A)$. The statement about $\lambda_{\max}$ follows from continuity of eigenvalues. □

Lemma 65 *(Lower Barrier Shift) Suppose $\lambda_{\min}(B|_S) > l + \delta_L$ and $Y = vv^T$ is a rank-one update. If $L_B \bullet Y \ge 1/t$ then $\Phi_{l+\delta_L}(B+tY) \le \Phi_l(B)$ and $\lambda_{\min}((B+tY)|_S) > l + \delta_L$.*

PROOF: We proceed as in the proof for the upper potential. Let $l' = l + \delta_L$ and $P = P_S$. By the Sherman–Morrison formula for the pseudoinverse (Lemma 59), we have:

$$\begin{aligned}
\Phi_{l+\delta_L}(B+tY) &= \operatorname{tr}(P(B+tY-l'I)P)^\dagger = \operatorname{tr}(P(B-l'I)P + tPYP)^\dagger \\
&= \operatorname{tr}(P(B-l'I)P)^\dagger - \frac{t\operatorname{tr}((P(B-l'I)P)^\dagger Y(P(B-l'I)P)^\dagger)}{1+t(P(B-l'I)P)^\dagger\bullet Y} \\
&= \Phi_l(B) + (\Phi_{l+\delta_L}(B) - \Phi_l(B)) - \frac{t(P(B-l'I)P)^{\dagger 2}\bullet Y}{1+t(P(B-l'I)P)^\dagger\bullet Y}
\end{aligned}$$

Note that matrix U_A is positive semidefinite. Rearranging shows that $\Phi_{l+\delta_L}(B+Y) \le \Phi_l(B)$ when $L_A(\pi) \ge 1/t$. It is immediate that $\lambda_{\min}(P_S(A+t\pi\pi^T)P_S) > l + \delta_L$ since $\lambda_{\min}(P_SAP_S) > l + \delta_L$. □

Now we prove that we can choose Y_i and t so that conditions of both lemmas are satisfied.

Lemma 66 *(Both Barriers) If $\Phi^u(A) \le \varepsilon_U$ and $\Phi_l(B) \le \varepsilon_L$ and $\varepsilon_U, \varepsilon_L, \delta_U, \delta_L$ satisfy*

$$0 \le \frac{1}{\delta_U} + \varepsilon_U + \max(N,T) \le \frac{1}{\delta_L} - \varepsilon_L,$$

and X, Y_i, $cost_i$, Z, T and N as in Theorem 57, $M^* - X$ is non-singular on S, then there exists i and positive t for which

$$L_B \bullet (ZY_iZ) \ge 1/t \ge U_A \bullet Y_i, \text{ and} \tag{4.4}$$

$$cost_i \cdot t \le 1/\max(N,T). \tag{4.5}$$

We will use the following lemma

Lemma 67 $\sum_{i=1}^{m} U_A \bullet Y_i \leq \frac{1}{\delta_U} + \varepsilon_U$ *and* $\sum_{i=1}^{m} L_B \bullet (ZY_iZ) \geq \frac{1}{\delta_L} - \varepsilon_L$.

PROOF: 1. We use Corollary 63 to bound the Frobenius product of Y_i with each of the two summands in the definition of U_A (note that they are positive semidefinite), we get

$$\begin{aligned}
\sum_{i=1}^{m} U_A \bullet Y_i &= U_A \bullet \sum_{i=1}^{m} Y_i = U_A \bullet (M^* - X) \\
&= \frac{((u+\delta_U)I - A)^{-2}}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} \bullet (M^* - X) + ((u+\delta_U)I - A)^{-1} \bullet (M^* - X) \\
&\leq \sum_{i=n-T+1}^{n} \lambda_i \left(\frac{((u+\delta_U)I - A)^{-2}}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} \right) + \sum_{i=n-T+1}^{n} \lambda_i \left(((u+\delta_U)I - A)^{-1} \right) \\
&= \frac{\sum_{i=n-T+1}^{n} \frac{1}{(u+\delta_U - \lambda_i(A))^2}}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + \sum_{i=n-T+1}^{n} \frac{1}{(u+\delta_U) - \lambda_i(A)}
\end{aligned}$$

Note that the first term is at most $1/\delta_U$, since

$$\begin{aligned}
\sum_{i=n-T+1}^{n} \frac{1}{(u+\delta_U - \lambda_i(A))^2} &\leq \sum_{i=n-T+1}^{n} \frac{1}{(u - \lambda_i(A))(u+\delta_U - \lambda_i(A))} \\
&= \frac{1}{\delta_U} \sum_{i=n-T+1}^{n} \left(\frac{1}{u - \lambda_i(A)} - \frac{1}{(u+\delta_U) - \lambda_i(A)} \right) = \frac{\Phi^u(A) - \Phi^{u+\delta_U}(A)}{\delta_U}
\end{aligned}$$

and the second term equals $\Phi^{u+\delta_U}(A)$. Thus $\sum_{i=1}^{m} U_A \bullet Y_i \leq \varepsilon_U + 1/\delta_U$.
2. Let P be the projection on $\mathrm{Im}(M^* - X)$. Since $(M^* - X)$ is non-singular on S, $PP_S = P_S$. We have,

$$\begin{aligned}
\sum_{i=1}^{m} L_B \bullet ZY_iZ &= L_B \bullet \sum_{i=1}^{m} ZY_iZ = L_B \bullet Z(M^* - X)Z = L_B \bullet P \\
&= \mathrm{tr} \left(\frac{(P_S(B - (l+\delta_L)I)P_S)^{\dagger 2}}{\Phi_{l+\delta_L}(B) - \Phi_l(B)} - (P_S(B - (l+\delta_L)I)P_S)^{\dagger} \right) \\
&= \frac{\sum_{i=1}^{k} \frac{1}{(\lambda_i(B|_S) - (l+\delta_L))^2}}{\Phi_{l+\delta_L}(B) - \Phi_l(B)} - \sum_{i=1}^{k} \frac{1}{\lambda_i(B|_S) - (l+\delta_L)} \\
&\geq 1/\delta_L - \varepsilon_L,
\end{aligned}$$

where the last line follows from Claim 3.6 in [**?**].

□

PROOF:(Of Lemma 66) For the previous lemma, we get: $\sum_{i=1}^{m}(U_A \bullet Y_i + \max(N,T)cost_i) \leq \frac{1}{\delta_U} + \varepsilon_U + \max(N,T) \leq L_B \bullet (ZY_iZ)$. Thus for some i, $U_A \bullet Y_i + \max(N,T)cost_i \leq L_B \bullet (ZY_iZ)$. Letting $t = (L_B \bullet (ZY_iZ))^{-1}$, we satisfy (4.4) and (4.5). □

4.2.3 Proof of Theorem 57

Now we are ready to prove Theorem 57. We assume that $M^* - X$ is non-singular on S (which we can ensure by an arbitrary small pertrubation).

We start with $A^{(0)} = X$, $B^{(0)} = 0$ and all weights $w_i^{(0)} = 0$. We define parameters as follows,

$$\begin{aligned} &\delta_L = 1/(2\max(N,T)), && \varepsilon_L = 1/(4\delta_L), && l_0 = -4k\delta_L, \\ &\delta_U = 4\delta_L, && \varepsilon_U = 1/(4\delta_L), && u_0 = 4T\delta_L + 1, \end{aligned}$$

so as to satisfy conditions of Lemma 66, $\Phi^u(A^{(0)}) = \Phi^u(X) = \sum_{i=1}^{T} \frac{1}{u_0 - \lambda_{n+1-i}(X)} \le T/(u_0 - 1) = \varepsilon_U$, $\Phi_l(B^{(0)}) = \sum_{i=1}^{k} \frac{1}{0 - l_0} = -k/l_0 = \varepsilon_L$, $1/\delta_U + \varepsilon_U + \max(N,T) = \frac{3}{2}\max(N,T) = 1/\delta_L - \varepsilon_L$. Then we iteratively apply Lemma 66. At iteration q, we find an index i and a positive t such that $L_{B^{(q)}}(ZY_iZ) \ge 1/t \ge U_{A^{(q)}}(Y_i)$, $cost_i \cdot t \le 1/\max(N,T)$, and increment the weight of matrix Y_i by t: $w_i^{(q+1)} = w_i^{(q)} + t$; update $l = l + \delta_L$ and $u = u + \delta_U$. The total cost increases by at most $1/\max(N,T)$. Finally, after N iterations we obtain matrices $A^{(N)}$ and $B^{(N)}$ with

$$\begin{aligned} \lambda_{\max}(A^{(N)}) &\le u_0 + N\delta_U = 2(N+T)/\max(N,T) + 1 \equiv \theta_{\max} \\ \lambda_{\min}(B^{(N)}\big|_S) &\ge l_0 + N\delta_L = (N/2 - 2k)/\max(N,T) \equiv \theta_{\min}. \end{aligned}$$

Now consider an arbitrary unit vector v. Let $v = v_S + v_{S^\perp}$, where $v_S \in S$ and $v_{S^\perp} \perp S$. Since $B^{(N)} \succeq \theta_{\min} P_S$ and $v_S \in S$,

$$\begin{aligned} v_S^T A^{(N)} v_S &= v_S^T (X + (P_S(M^* - X)P_S)^{1/2} B^{(N)} (P_S(M^* - X)P_S)^{1/2}) v_S \\ &\ge v_S^T (X + (P_S(M^* - X)P_S)^{1/2} \theta_{\min} P_S (P_S(M^* - X)P_S)^{1/2}) v_S \\ &= \theta_{\min} v_S^T M^* v_S + (1 - \theta_{\min}) v_S^T X v_S \ge \theta_{\min} \lambda_{\min}(M^*) \|v_S\|^2. \end{aligned}$$

On the other hand, $v_{S^\perp}^T A^{(N)} v_{S^\perp} \le \theta_{\max} \|v_{S\perp}\|$. Thus from the triangle inequality for the norm induced by $A^{(N)}$, we get

$$(v^T A^{(N)} v)^{1/2} \ge \theta_{\min}^{1/2} \lambda_{\min}(M)^{1/2} \|v_S\| - \theta_{\max}^{1/2} \|v_{S^\perp}\| \ge \theta_{\min}^{1/2} \lambda_{\min}(M)^{1/2} - (\theta_{\max}^{1/2} + \theta_{\min}^{1/2} \lambda_{\min}(M)^{1/2}) \|v_{S^\perp}\|.$$

On the other hand, since S is an eigenspace of X corresponding to k smallest eigenvalues,

$$(v^T A^{(N)} v)^{1/2} \ge (v^T X v)^{1/2} \ge (v_{S^\perp}^T X v_{S^\perp})^{1/2} \ge \lambda^{*1/2} \|v_{S^\perp}\|.$$

One of the two bounds above for $(v^T A^{(N)} v)^{1/2}$ increases and the other decreases as $\|v_{S^\perp}\|$ increases. They are equal when $\|v_{S^\perp}\| = \frac{\theta_{\min}^{1/2} \lambda_{\min}(M^*)^{1/2}}{\lambda^{*1/2} + \theta_{\max}^{1/2} + \theta_{\min}^{1/2} \lambda_{\min}(M^*)^{1/2}}$. Therefore,

$(v^T A^{(N)} v)^{1/2} \geq \frac{\theta_{\min}^{1/2} \lambda^{*1/2} \lambda_{\min}(M^*)^{1/2}}{\lambda^{*1/2} + \theta_{\max}^{1/2} + \theta_{\min}^{1/2} \lambda_{\min}(M^*)^{1/2}}$. We conclude that

$$\lambda_{\min}(A^{(N)}) = \min_{v:\|v\|=1} v^T A^{(N)} v \geq \frac{\theta_{\min} \lambda^* \lambda_{\min}(M^*)}{\left(\lambda^{*1/2} + \theta_{\max}^{1/2} + \theta_{\min}^{1/2} \lambda_{\min}(M^*)^{1/2}\right)^2}.$$

Plugging in the values of parameters, we get the statement of the theorem for $M = A^{(N)}$. The total cost is at most $N/\max(N,T) = \min(1, N/T)$.

Finally, we prove Claim 56. PROOF:[Claim 56] Let $V = \mathrm{Im}(\mathcal{L}_{G+W}) = \ker(\mathcal{L}_{G+W})^{\perp}$. Let $\mathcal{L}_e$ be the Laplacian of the edge e. Define

$$\begin{aligned} X &= \left((\mathcal{L}_{G+W}^{\dagger})^{1/2} \mathcal{L}_G (\mathcal{L}_{G+W}^{\dagger})^{1/2} \right) \Big|_V, \\ Y_e &= w_e \left((\mathcal{L}_{G+W}^{\dagger})^{1/2} \mathcal{L}_e (\mathcal{L}_{G+W}^{\dagger})^{1/2} \right) \Big|_V, \\ cost_e &= w_e / \left(\sum_{d \in E_W} w_d \right). \end{aligned}$$

Since $\mathcal{L}_G + \sum_{e \in E_W}^{m} w_e \mathcal{L}_e = \mathcal{L}_{G+W}$, we have $X + \sum_{e \in E_W} Y_e = I$. By the definition of the (k, T, λ^*)-patch, $\mathrm{tr}(I - X) \leq T$ and $\lambda^* \leq \lambda^{k+1}(X)$. We apply Theorem 57 to matrices X, Y_e and $M^* = I$. We obtain a set of weights ρ_e — supported on at most N edges — such that

$$c_1 \min(N/T, 1) \cdot \lambda^* \leq \lambda_{\min} \left(X + \sum_{e \in E_W} \rho_e Y_e \right) \leq \lambda_{\max} \left(X + \sum_{e \in E_W} \rho_e Y_e \right) \leq c_2,$$

Let $\tilde{w}_e = \rho_e w_e$. Weights $\tilde{w}_i$ define subgraph W_k with at most N edges. It follows that

$$c_1 \min(N/T, 1) \lambda^* \mathcal{L}_{G+W} \preceq \mathcal{L}_{G+W_k} \preceq c_2 \mathcal{L}_{G+W}.$$

The total weight of edges of W_k is $\sum_{e \in E_W} \rho_e w_e = (\sum_{e \in E_W} \rho_e cost_e) \sum_{d \in E_W} w_d \leq \min(1, N/T) \sum_{d \in E_W} w_d$. □

4.3 Constructing Nearly-Optimal Ultrasparsifiers

We now apply our subgraph sparsification to build ultrasparsifiers. Recall that a weighted graph U is a (κ, k)-ultrasparsifier of another graph G if $U \preceq G \preceq \kappa \cdot U$ and U has only $n - 1 + k$ edges, where n is the number of vertices in U and G. The main result of this section is the following theorem.

Theorem 68 *For any integer $k > 0$, every graph has an $(\frac{n}{k} \log n \tilde{O}(\log \log n), k)$–ultrasparsifier.*

Our basic idea to build a good ultrasparsifier U is quite simple. Without loss of generality, we can assume that G is connected and has $O(n)$ edges. Otherwise given a graph G, we can first find a linear size sparsifier using [**?**], for each of its connected components, and build a good ultrasparsifier for each component. Because U is only k edges aways from

a tree, our construction starts with good tree T. As it will be much more clear below, the quality of a tree is measured by its *stretch*, as introduced by Alon, Karp, Peleg and West [**?**].

Suppose T is a spanning tree of $G = (V, E, w)$. For any edge $e \in E$, let $e_1, \cdots, e_k \in F$ be the edges on the unique path in T connecting the endpoints of e. The *stretch* of e w.r.t. T is given by $\mathrm{st}_T(e) = w(e)(\sum_{i=1}^{k} \frac{1}{w(e_i)})$. The stretch of the graph G with respect to T is defined by $\mathrm{st}_T(G) = \sum_{e \in E} \mathrm{st}_T(e)$. Our construction will start with a spanning tree with the lowest possible stretch. By [**?**], we can in polynomial time grow a spanning tree T with

$$\mathrm{st}_T(G) = O(n \log n \log\log n (\log\log\log n)^3).$$

Remark 2 *For the sake of simplicity of the presentation, we will show the construction of ultrasparsifiers with $\Theta(k)$ edges. We note that by choosing the appropriate constants, the number of edges can be made exactly k.*

Let $\kappa = c_1 \cdot \mathrm{st}_T(G)/k$ for a sufficiently large constant c_1. Our job is to choose $\Theta(k)$ more weighted edges $\tilde{W}$ and set $U = T + \tilde{W}$ such that $c_2 \cdot U \preceq G \preceq \kappa \cdot U$, for a constant c_2. To this end, let $W = (1/(c_3\kappa)) \cdot G$, for some constant c_3. Then, $G = c_3\kappa \cdot W \preceq c_3\kappa \cdot (W + T)$. Also, because $T \preceq G$, we have $T + W \preceq (1 + 1/(c_3\kappa))G \preceq c_4 \cdot G$, for a constant c_4. Therefore, if we can find a $\Theta(k)$–edge subgraph $\tilde{W}$ of W such that $T + \tilde{W} \preceq \Theta(1) \cdot (T + W)$, we can then build a $n - 1 + \Theta(k)$ edge graph $U = T + \tilde{W}$ satisfying $c_2 \cdot U \preceq G \preceq \kappa \cdot U$ (if we choose our constants c_i's carefully).

To apply our subgraph sparsification results to construct $\tilde{W}$, we use the following structure result of Spielman and Woo ([**?**]: Theorem 2.1 and Corollary 2.2).

Lemma 69 *(Theorem 2.1 in [**?**]) (1) $Tr({\mathcal{L}_T^{\dagger}}^{1/2} \mathcal{L}_G {\mathcal{L}_T^{\dagger}}^{1/2}) = st_T(G)$. (2) For every $t > 0$, the number of eigenvalues of ${\mathcal{L}_T^{\dagger}}^{1/2} \mathcal{L}_G {\mathcal{L}_T^{\dagger}}^{1/2}$ greater than t is at most $st_T(G)/t$.*

We now use Lemma 69 to prove the following lemma, from which Theorem 68 follows directly.

Lemma 70 *W is a $(k, O(k), \Theta(1))$–patch for T.*

PROOF: Let $\lambda_i = \lambda_i((\mathcal{L}_{T+W}^{\dagger})^{1/2} \mathcal{L}_T (\mathcal{L}_{T+W}^{\dagger})^{1/2})$ be the i-th eigenvalue, and y_i be the corresponding eigenvector. Let $x_i = L_{T+W}^{1/2} y_i$. Then,

$$\lambda_i = \lambda_i((\mathcal{L}_{T+W}^{\dagger})^{1/2} \mathcal{L}_T (\mathcal{L}_{T+W}^{\dagger})^{1/2}) = \frac{x_i^T \mathcal{L}_T x_i}{x_i^T \mathcal{L}_T x_i + x_i^T \mathcal{L}_W x_i} = \frac{x_i^T \mathcal{L}_T x_i}{x_i^T \mathcal{L}_T x_i + x_i^T \mathcal{L}_G x_i/(c_3\kappa)},$$

implying

$$\frac{x_i^T \mathcal{L}_G x_i}{x_i^T \mathcal{L}_T x_i} = \frac{1 - \lambda_i}{\lambda_i} c_3\kappa = \left(\frac{1 - \lambda_i}{\lambda_i}\right) c_3 c_1 \frac{\mathrm{st}_T(G)}{k} = \frac{\mathrm{st}_T(G)}{\frac{k}{c_1 c_3} \frac{\lambda_i}{1 - \lambda_i}}$$

It follows from the definition of λ_i that $0 \leq \lambda_i < 1$. Hence, $(1 - \lambda_{i-1})/\lambda_{i-1} \geq (1 - \lambda_i)/\lambda_i$. By Courant—Fischer theorem and the property 2 of Lemma 69, we have $k \leq \frac{k}{c_1 c_3} \frac{\lambda_{k+1}}{1 - \lambda_{k+1}}$. Therefore, $\lambda_{k+1} \geq \frac{c_1 c_3}{1 + c_1 c_3} = \Theta(1)$. We also have,

$$\begin{aligned}\mathrm{tr}\left((\mathcal{L}_{T+W}^{\dagger})^{1/2}\mathcal{L}_W(\mathcal{L}_{T+W}^{\dagger})^{1/2}\right) &\leq \mathrm{tr}\left((\mathcal{L}_T^{\dagger})^{1/2}\mathcal{L}_W(\mathcal{L}_T^{\dagger})^{1/2}\right) = \frac{1}{c_3\kappa}\mathrm{tr}\left((\mathcal{L}_T^{\dagger})^{1/2}\mathcal{L}_G(\mathcal{L}_T^{\dagger})^{1/2}\right) \\ &\leq \frac{k}{c_3c_1\mathrm{st}_T(G)}\mathrm{st}_T(G) = \frac{k}{c_3c_1} = \Theta(k).\end{aligned}$$

We proved that W is a $(k, O(k), \Theta(1))$–patch for T. □

We next show that the parameters of the ultrasparsifiers we obtained are optimal, up to low order terms.

Theorem 71 *Let G be a Ramanujan d-regular expander graph, for some constant d. Let U a (κ, N) ultrasparsifier for G. Then $\kappa \geq \frac{n}{N}\log n$.*

Proof: Let T be a low-stretch spanning tree of G, as above. As mentioned in [**?**], $\mathrm{st}_T(G) = \Omega(m\log n)$ where m is the number of edges of the original graph. From lemma 69, and the conditions on the stretch of T we have $\mathrm{Tr}(\mathcal{L}_G{\mathcal{L}_T}^{\dagger}) = \mathrm{st}_T(G) \geq C\cdot n\log n$ for some constant C.

Since $x^T\mathcal{L}_G x = \Theta(1)$ for the expander, the above inequality implies that $\sum_{i=1}^{n}\frac{1}{x^T\mathcal{L}_T x} \geq n\log n$ where x_i are the eigenvectors of $\mathcal{L}_G(\mathcal{L}_T)^{\dagger}$. It is immediate from Markov's inequality that there exists some k such that ${x_k}^T\mathcal{L}_T x_k \leq \frac{C_1 k}{n\log n}$. Assume that for all $i \leq k$ we have ${x_i}^T\mathcal{L}_T x_i \leq {x_k}^T\mathcal{L}_T x_k \leq \frac{C_1 k}{n\log n}$. (Otherwise take $k' < k$ appropriately). Then also $\lambda_k(\mathcal{L}_T) \leq \frac{C_1 k}{n\log n}$. By the minmax theorem for eigenvalues this implies that adding $N = k-2$ edges to T will result to a graph U with $\lambda_2(\mathcal{L}_U) \leq \lambda_k(\mathcal{L}_T) \leq \frac{C_1 k}{n\log n}$. Thus any ultrasparsifier U with N edges will have

$$C_2 = \lambda_2(\mathcal{L}_G) \leq \kappa\lambda_2(\mathcal{L}_U) \leq \frac{C_1 k}{n\log n} \Rightarrow \kappa \geq \Omega(\frac{n\log n}{k}) = \Omega(\frac{n\log n}{N})$$

□

4.4 Maximizing Algebraic Connectivity by Adding few edges

In this section, we present an approximation algorithm for the following problem: given a graph $G = (V, E_{base})$, a set of candidate edges E_{cand}, and a parameter k, add at most k candidate edges to G so as to maximize its algebraic connectivity, that is, find a subset $E \subset E_{cand}$ that maximizes $\lambda_2(\mathcal{L}_{G+E})$. The problem was introduced by Ghosh and Boyd [**?**], who presented a heuristic for it. It is known that the problem is NP-hard [**?**]. But prior to this work, no approximation algorithm was known for it.

We use two upper bounds for the cost of the combinatorial solution in order to prove an approximation guarantee: one upper bound is the SDP value, λ_{SDP}, and the other is $\lambda_{k+2}(\mathcal{L}_G)$ (see Lemma 72). Note that neither of these two bounds are good approximations for the value of the optimum solution by themselves (for instance, if G consists of n isolated vertices, (V, E_{cand}) is an expander, $k < n$, then the value of the combinatorial solution is

0 but $\lambda_{SDP} \sim k/n$), but their combinations lead to a good upper bound for the optimum solution λ_{OPT}.

For clarity and simplicity of exposition, we assume here that (V, E_{base}) and (V, E_{cand}) are bounded degree graphs with the maximum degree Δ. Our algorithm uses a natural semidefinite relaxation that was also used by Ghosh and Boyd [**?**]. We introduce a variable w_e (the weight of the edge e) for each candidate edge $e \in E_{cand}$; add constraints that all edge weights are between 0 and 1, and the total weight is at most k. Then we require that $\lambda_2(\mathcal{L}_G + \sum_e w_e \mathcal{L}_e) \geq \lambda_{SDP}$ (where $\mathcal{L}_e$ is the Laplacian of the edge e). We do that by adding an SDP constraint $\mathcal{L}_G + \sum_e w_e \mathcal{L}_e \succeq \lambda_{SDP} P_{(1,\ldots,1)^\perp}$, where $P_{(1,\ldots,1)^\perp}$ is the projection on the space orthogonal to $(1,\ldots,1)^\perp$. We get the following SDP relaxation.

$$\begin{aligned} \text{maximize: } & \lambda_{SDP}, \\ \text{subject to: } & \mathcal{L}_G + \sum_{e \in E_{cand}} w_e \mathcal{L}_e \succeq \lambda_{SDP} \cdot P_{(1,\ldots,1)^\perp}, \\ & \sum_{e \in E_{cand}} w_e \leq k, \\ & 0 \leq w_e \leq 1 \text{ for every } e \in E_{cand}. \end{aligned}$$

We solve the semidefinite program and obtain solution $\{w_e\}_{e \in E_{cand}}$. The total weight of all edges is k, however, the number of edges involved, or the support of the solution could be significantly higher than k.

We use our algorithm to *sparsify* the SDP solution using Theorem 57. More precisely, we apply Theorem 57 with $X = \mathcal{L}_G/(4\Delta)$ and $Y_e = w_i \mathcal{L}_e/(4\Delta)$ restricted to the space $(1,\ldots,1)^\perp$, $N = 8k$, $T = \mathrm{tr}(\sum_e w_e \mathcal{L}_e)/(4\Delta) \leq k$ and $cost_i = w_i$ (we divide $\mathcal{L}_G$ and $\mathcal{L}_e$ by 4Δ to ensure that $\lambda_{\max}(X + \sum_e Y_i) \leq 1$). We get a set of weights ρ_e supported on at most $8k$ edges s.t.

$$\frac{1}{4\Delta}\lambda_2(\mathcal{L}_G + \sum_e \rho_e w_e \mathcal{L}_e) = \lambda_{\min}(X + \sum_e \rho_e Y_e) \geq c\lambda_{k+2}(X)\lambda_{\min}(X + \sum_e Y_e) \geq c\frac{1}{(4\Delta)^2}\lambda_{k+2}(\mathcal{L}_G)\lambda_{SDP}.$$

That is, we obtain a combinatorial weighted solution $\tilde{w}_e = \rho_i w_i$ whose value is at least $c\lambda_{k+2}(\mathcal{L}_G)\lambda_{SDP}/(4\Delta)$ (if $k+2 > n$, the value is at least $c\lambda_{SDP}$). We next show that $\lambda_{SDP} \geq \lambda_{OPT}$ and $\lambda_{k+2}(G) \geq \lambda_{OPT}$. Therefore, the value of the solution is at least $c\lambda_{OPT}^2/\Delta$.

Lemma 72 *The value of the optimal solution, λ_{OPT}, is at most $\lambda_{k+2}(\mathcal{L}_G)$.*

PROOF: Consider the optimal solution E. Let $\mathcal{L}_E$ be the Laplacian of the graph formed by E. Note that $\mathrm{rank}(\mathcal{L}_E) \leq |E| \leq k$, therefore, $\dim \ker \mathcal{L}_E \geq n - k$. Let S be the $k+1$-dimensional space spanned by the eigenvectors of $\mathcal{L}_G$ corresponding to $\lambda_2(\mathcal{L}_G), \ldots, \lambda_{k+2}(\mathcal{L}_G)$. Since $\dim S + \dim \ker E > n$, spaces S and $\ker \mathcal{L}_E$ have a non-trivial intersection. Choose a unit vector $v \in \ker S \cap \mathcal{L}_E$. We have $v(\mathcal{L}_G + \mathcal{L}_E)v^T \leq \lambda_{k+2}(\mathcal{L}_G) + 0 = \lambda_{k+2}(\mathcal{L}_G)$. Also v is orthogonal to the vector $(1,\ldots,1)^\perp$. Therefore, $\lambda_{OPT} = \lambda_2(\mathcal{L}_G + \mathcal{L}_E) \leq \lambda_{k+2}(\mathcal{L}_G)$. □

The edges in the support of $\tilde{w}_e$, $E = \{\tilde{w}_e : \tilde{w}_e \neq 0\}$, form a non-weighted combinatorial solution. Since $\lambda_{\max}(\mathcal{L}_X + \sum_e \tilde{w}_e \mathcal{L}_e) = O(\Delta)$, all weights $\tilde{w}_e$ are bounded by $O(\Delta)$, and thus the algebraic connectivity of $G + E$ is at least $c\lambda_{k+2}(\mathcal{L}_G)\lambda_{SDP}/\Delta^2$.

Theorem 73 *There is a polynomial time approximation algorithm that finds a solution of value at least $c\lambda_{OPT}^2/\Delta$ supported on at most $8k$ edges with total weight at most k. If $k \geq n$ the algorithm finds a constant factor approximation.*

We present two corollaries for special instances of the problem.

Corollary 74 *If it is possible to make G an expander by adding k edges (and thus $\lambda_{OPT} \sim \Delta$), then the algorithm finds a constant factor approximation.*

Note that if the graph formed by candidate edges is an expander then the value of the following SDP solution $w_e = k/|E_{cand}|$ for each edge $e \in E_{cand}$ is $\Omega(k/n)$, thus $\lambda_{SDP} \geq ck/n$.

Corollary 75 *If the graph formed by candidate edges is an expander, then the approximation algorithm from Theorem 73 finds a solution of value at least $c\frac{k}{n\Delta}\lambda_{OPT}$.*

Remark 3 *It is possible to get rid of the dependence on Δ in Theorem 73 and Corollary 75 and obtain approximation guarantees of $c\min(\lambda_{OPT}, \lambda_{OPT}^2)$ and $\frac{ck}{n}\lambda_{OPT}$ respectively. We omit the details in this extended abstract.*

4.5 Open Problems

We believe that there is a constant factor approximation algorithm for optimizing the algebraic connectivity as in section 4.4. We also believe that there is an almost linear time algorithm for constructing sparsifiers, patch sparsifiers and ultrasparsifiers.
Another important direction is to examine what other graph quantities can be approximated by replacing the original graph with a sparse one. For example, can one construct "sparsifier" that approximate *vertex expansion*? What about sparsifiers for directed graphs?

Chapter 5

Sparsest Cut on Quotients of the Hypercube

In this chapter, we present a simple construction and analysis of an $\Omega(\log \log N)$ integrality gap for the well-known Sparsest Cut semi-definite program (SDP). This holds for the uniform demands version (i.e. edge expansion). The same quantitative gap was proved earlier by Devanur, Khot, Saket, and Vishnoi [STOC 2006], following an integrality gap for non-uniform demands due to Khot and Vishnoi [FOCS 2005]. These previous constructions involve a complicated SDP solution and analysis, while our gap instance, vector solution, and analysis are somewhat simpler and more intuitive.

Furthermore, our approach is rather general, and provides a variety of different gap examples derived from quotients of the hypercube. It also illustrates why the lower bound is stuck at $\Omega(\log \log N)$, and why new ideas are needed in order to derive stronger examples.

5.1 Preliminaries

We first discuss some preliminary notions and theorems that will be used throughout the paper.

Asymptotic notation. For expressions A and B, we will use the notation $A \lesssim B$ to denote $A = O(B)$, and $A \approx B$ to denote the conjunction of $A \lesssim B$ and $A \gtrsim B$.

Sparsity of graphs. We will consider undirected graphs $G = (V, E)$ where every edge (u, v) has a non-negative weight $w(u, v)$. For any subset $E' \subseteq E$ of edges, we write $w(E') = \sum_{e \in E'} w(e)$. For two sets $S, T \subseteq V$, we write $E(S, T)$ for the set of edges with one endpoint in S and one in T.

For a subset $S \subseteq V$, we use

$$\Phi(S) = \frac{w(E(S, \overline{S}))}{|S||\overline{S}|}$$

to denote the *sparsity of* S. We then write $\Phi(G) = \min_{S \subseteq V} \Phi(S)$ for the *sparsest cut value of* G.

We will be particularly interested in graphs derived from the (unweighted) n-dimensional hypercube $Q_n = \left\{\frac{-1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right\}^n$. We will use Q_n to denote the set of vertices in the n-cube, and $E(Q_n)$ to denote the set of edges. The classical discrete isoperimetric inequality shows that if we write $S_i = \{x \in Q_n : x_i < 0\}$, then for every $i \in [n]$,

$$\Phi(Q_n) = \Phi(S_i) = \frac{4|E(S_i, \overline{S}_i)|}{|Q_n|^2} \approx |Q_n|^{-1}.$$

A well-known theorem of Kahn, Kalai, and Linial [**?**] then asserts the following.

Theorem 76 (KKL Theorem) *For any $S \subseteq Q_n$, there exists an $i \in [n]$ for which*

$$\frac{|E(S, \overline{S}) \cap E(S_i, \overline{S}_i)|}{|S||\overline{S}|} \gtrsim \frac{\log n}{n} \Phi(Q_n).$$

Weighted "quotients" of the cube. Let Γ be any group acting on $[n] = \{1, 2, \ldots, n\}$ by permutations. We can naturally extend Γ to act on Q_n via $\pi(x_1, \ldots, x_n) = (x_{\pi(1)}, \ldots, x_{\pi(n)})$ for any $\pi \in \Gamma$. For an element $u \in Q_n$, we use Γu to denote the Γ-orbit of u. We refer to a subset $S \subseteq Q_n$ as Γ-invariant if $\Gamma S = S$.

We define a weighted graph Q_n/Γ as follows. The vertices are simply those of Q_n, and the edges are $E(Q_n) \cup E'$, where $E' = \{(u, v) : u \in \Gamma v\}$. We define

$$w(e) = \begin{cases} 1 & e \in E(Q_n) \\ 2^{2n} & e \in E'. \end{cases}$$

The point of this choice is to ensure that $\Phi(Q_n/\Gamma) = \Phi(S)$ is always achieved by a Γ-invariant set S, since separating any Γ-orbit involves cutting an edge of very large value. (Note that, because we are only using weights which are polynomial in the graph size, our gap examples can easily be made unweighted.)

We recall that Γ is said to act *transitively* on $[n]$ if for every $i, j \in [n]$, there exists a permutation $\pi \in \Gamma$ with $\pi(i) = j$. From Theorem 76, one can easily derive the following.

Theorem 77 (Transitive actions) *If Γ acts transitively on $[n]$, then $\Phi(Q_n/\Gamma) \gtrsim \Phi(Q_n) \log n$.*

Proof: We know that $\Phi(Q_n/\Gamma) = \Phi(S)$ for some Γ-invariant set S. By Theorem 76, there exists an $i \in [n]$ for which

$$\frac{|E(S, \overline{S}) \cap E(S_i, \overline{S}_i)|}{|S||\overline{S}|} \gtrsim \frac{\log n}{n} \Phi(Q_n).$$

But for any other $j \in [n]$, there exists an action $\pi \in \Gamma$ with $\pi(i) = j$, hence

$$\frac{|E(\pi(S), \pi(\overline{S})) \cap E(S_j, \overline{S}_j)|}{|\pi(S)||\pi(\overline{S})|} = \frac{|E(S, \overline{S}) \cap E(S_i, \overline{S}_i)|}{|S||\overline{S}|},$$

implying that

$$\Phi(S) = \sum_{j=1}^{n} \frac{|E(S,\overline{S}) \cap E(S_j,\overline{S}_j)|}{|S||\overline{S}|} = n \cdot \frac{|E(S,\overline{S}) \cap E(S_i,\overline{S}_i)|}{|S||\overline{S}|} \gtrsim \Phi(Q_n) \log n.$$

□

The Sparsest Cut SDP. Given a weighted graph $G = (V, E)$, we recall the standard SDP relaxation of Sparsest Cut,

$$\mathsf{SDP}(G) = \min \left\{ \frac{\sum_{uv \in E} w(u,v)\|x_u - x_v\|^2}{\sum_{u,v \in V} \|x_u - x_v\|^2} : \|x_u - x_v\|^2 \leq \|x_u - x_w\|^2 + \|x_w - x_v\|^2 \ \forall u,v,w \in V \right\},$$

where the minimum is taken over all choices of vectors $\{x_u\}_{u \in V}$ lying in some finite-dimensional Euclidean space. It is well-known that $\mathsf{SDP}(Q_n) = \Phi(Q_n) \approx |Q_n|^{-1}$.

We say that a vector solution $\{x_u\}_{u \in Q_n}$ is Γ-invariant if $x_u = x_{\pi(u)}$ for all $u \in Q_n$ and $\pi \in \Gamma$. Observe that a Γ-invariant solution for the Sparsest Cut SDP on Q_n/Γ has value

$$\frac{\sum_{uv \in E(Q_n)} \|x_u - x_v\|^2}{\sum_{u,v \in Q_n} \|x_u - x_v\|^2},$$

since all elements of a Γ-orbit are mapped to the same vector.

Weak triangle inequalities and pseudometrics. For the sake of exposition, we will also define an "SDP value" for solutions satisfying a weak form of the triangle inequalities. We recall that for any set X, a non-negative, symmetric function $d : V \times V \to \mathbb{R}$ is called a *pseudometric on* V if it satisfies the triangle inequalities, i.e. $d(u,v) \leq d(u,w) + d(w,v)$ for all $u, v, w \in V$, and additionally $d(u,u) = 0$ for all $u \in V$.

For any $\beta \geq 1$, let

$$\mathsf{SDP}_\beta(G) = \min \left\{ \frac{\sum_{uv \in E} w(u,v)\|x_u - x_v\|^2}{\sum_{u,v \in V} \|x_u - x_v\|^2} : d(u,v) \leq \|x_u - x_v\|^2 \leq \beta d(u,v) \right\},$$

where the minimum is over all choices of vectors $\{x_u\}_{u \in V}$, and additionally over all pseudometrics d on V. Observe that $\mathsf{SDP}(G) = \mathsf{SDP}_1(G)$. One might also note that the Arora-Rao-Vazirani algorithm [**?**], and all known analyses derived from it, only use the weaker $\mathsf{SDP}_{O(1)}$ inequalities.

Tensoring. We recall that for two vectors $x, y \in \mathbb{R}^k$ and $t \in \mathbb{N}$, we have the tensored vectors $x^{\otimes t}, y^{\otimes t} \in \mathbb{R}^{k^t}$ which satisfy $\langle x^{\otimes t}, y^{\otimes t} \rangle = \langle x, y \rangle^t$.

Finally, we need the following tail inequality.

Lemma 78 (Hoeffding bound) *Let $X_1, X_2, \ldots, X_n$ be independent random variables with $\mathbb{E}X_i = 0$ for every $i \in [n]$. Then,*

$$\mathbf{Pr}\left[\left|\sum_{i=1}^{n} X_i\right| \geq L\right] \leq 2\exp\left(\frac{-L^2}{2\sum_{i=1}^{n} \|X_i\|_\infty^2}\right).$$

5.2 A simple example: Cyclic shifts

Consider the cyclic shift operator $\sigma : [n] \to [n]$ defined by $\sigma(i) = (i+1) \bmod n$, and let $\Gamma = \{\sigma^0, \sigma^1, \ldots, \sigma^{n-1}\}$ be the group of permutations generated by σ. By Theorem 77, we have $\Phi(Q_n/\Gamma) \gtrsim \Phi(Q_n) \log n$. On the other hand, we will now show that the "weak" SDP value of Q_n/Γ is approximately $\mathsf{SDP}(Q_n)$, thus exhibiting a (weak) SDP gap of $\Omega(\log n) = \Omega(\log\log|Q_n|)$. This will illustrate the main ideas behind our proof for general quotients, and the true SDP value will be analyzed in the next section.

Theorem 79 *For $n \in \mathbb{N}$, $\mathsf{SDP}_{16}(Q_n/\Gamma) \lesssim \mathsf{SDP}(Q_n)$.*

Proof: For every $u \in Q_n$, we define the vector

$$x_u = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} (\sigma^i u)^{\otimes 8},$$

and put $\widetilde{x_u} = x_u/\|x_u\|$. Observe that

$$\langle x_u, x_v \rangle = \frac{1}{n} \sum_{i,j=0}^{n-1} |\langle \sigma^i u, \sigma^j v \rangle|^8 = \sum_{i=0}^{n-1} |\langle u, \sigma^i v \rangle|^8. \tag{5.1}$$

We now define a subset of "pseudorandom" vertices of Q_n whose orbits under Γ are not too self-correlated,

$$\mathcal{P} = \left\{ u \in Q_n : \sum_{i=0}^{n-1} |\langle u, \sigma^i u \rangle|^6 \le 1 + \frac{1}{4n} \right\}.$$

Note that, by Cauchy-Schwarz, for $u, v \in \mathcal{P}$, we have

$$\sum_{i=0}^{n-1} |\langle u, \sigma^i v \rangle|^6 \le \sqrt{\sum_{i=0}^{n-1} |\langle u, \sigma^i u \rangle|^6} \sqrt{\sum_{i=0}^{n-1} |\langle v, \sigma^i v \rangle|^6} \le 1 + \frac{1}{4n}. \tag{5.2}$$

(To see this, observe that $\sum_{i=0}^{n-1} |\langle u, \sigma^i v \rangle|^6$ is an inner product, as in (5.1).)

Most vertices are pseudorandom. For any $u \in Q_n$, we can write

$$\langle u, \sigma u \rangle = \sum_{\substack{1 \le i \le n \\ i \text{ even}}} u_i u_{\sigma(i)} + \sum_{\substack{1 \le i \le n \\ i \text{ odd}}} u_i u_{\sigma(i)} = T + T',$$

where each u_i appears exactly once in each of the sums T and T'. It is easy to see that a similar decomposition holds for $\langle u, \sigma^i u \rangle$ for any $i \in \{1, 2, \ldots, n-1\}$.

Therefore by Lemma 78, we have

$$\mathbf{Pr}_{u \in Q_n} \left[|\langle u, \sigma^i u \rangle| \ge 2t/\sqrt{n} \right] \le \mathbf{Pr} \left[|T| \ge t/\sqrt{n} \right] + \mathbf{Pr} \left[|T'| \ge t/\sqrt{n} \right] \le 4e^{-t^2/2}, \tag{5.3}$$

since each of T and T' is a sum of i.i.d. uniform elements of $\{\pm\frac{1}{n}\}$. Setting $t = n^{1/3}/2$ and taking a union bound over $i = 1, 2, \ldots, n-1$ yields

$$\mathbf{Pr}_{u\in Q_n}\left[\sum_{i=0}^{n-1} |\langle u, \sigma^i v\rangle|^6 > 1 + \frac{1}{4n}\right] \leq 4ne^{-n^{2/3}/8} \leq n^{-2}, \tag{5.4}$$

for n sufficiently large, hence $|\mathcal{P}| \geq |Q_n|(1 - n^{-2})$.

The SDP value. Fix some $u_0 \in \mathcal{P}$. Our final SDP solution will consist of the vectors $\{x'_u\}_{u\in Q_n}$ with $x'_u = \widetilde{x_u}$ for $u \in \mathcal{P}$ and $x'_u = \widetilde{x_{u_0}}$ otherwise. Thus we will only need to verify the weak triangle inequalities for $\{\widetilde{x_u}\}_{u\in\mathcal{P}}$. It is clear that our proposed SDP solution is Γ-invariant.

For an edge $(u, v) \in E(Q_n)$, using (5.1), we have

$$\langle x_u, x_v\rangle \geq |\langle u, v\rangle|^8 = \left(1 - \frac{2}{n}\right)^8 \geq 1 - \frac{16}{n}.$$

Hence for $u, v \in \mathcal{P}$ with $(u, v) \in E(Q_n)$, we have $\|\widetilde{x_u} - \widetilde{x_v}\|^2 = O(1/n)$. In particular,

$$\sum_{(u,v)\in E(Q_n)} \|x'_u - x'_v\|^2 \lesssim \frac{|E(Q_n)|}{n} + 4|E(Q_n \setminus \mathcal{P})| \lesssim \frac{|E(Q_n)|}{n}, \tag{5.5}$$

since $|Q_n \setminus \mathcal{P}| \leq |Q_n|/n^2$.

On the other hand, if we choose $u, v \in Q_n$ at random, then for any $i \in [n]$, using Lemma 78,

$$\mathbf{Pr}_{u,v\in Q_n}\left[|\langle u, \sigma^i v\rangle| \geq t/\sqrt{n}\right] \leq 2e^{-t^2/2}.$$

Setting $t \approx \sqrt{\log n}$ and taking a union bound over all $i \in [n]$ shows that for n sufficiently large, $\mathbf{Pr}_{u,v\in Q_n}[|\langle x_u, x_v\rangle| \geq \frac{1}{4}] \leq \frac{1}{2}$. In particular,

$$\sum_{u,v\in Q_n} \|x'_u - x'_v\|^2 \geq \sum_{u,v\in\mathcal{P}} \|\widetilde{x_u} - \widetilde{x_v}\|^2 \approx \sum_{u,v\in\mathcal{P}} \|x_u - x_v\|^2 \gtrsim |\mathcal{P}|^2 \gtrsim |Q_n|^2.$$

Combining the preceding line with (5.5) shows that the value of the potential SDP solution $\{x'_u\}_{u\in Q_n}$ is $O(|Q_n|^{-1}) = O(\mathsf{SDP}(Q_n))$.

Verifying the weak triangle inequalities. We are thus left to verify the weak triangle inequalities for $\{\widetilde{x_u}\}_{u\in\mathcal{P}}$. To this end, we will define a cyclic shift-invariant metric d on Q_n and then show that for $u, v \in \mathcal{P}$, we have $d(u, v) \approx \|\widetilde{x_u} - \widetilde{x_v}\|^2$.

Let $\lambda(u, v) = \max\{|\langle u, \sigma^i v\rangle| : i \in [n]\}$ and put $d(u, v) = 1 - \lambda(u, v)^8$. It is clear that $d(u, v) = d(\sigma u, v) = d(u, \sigma v)$. Next, observe that for any $u, v, w \in Q_n$, we have

$$1 + \langle u, v\rangle \geq \langle u, w\rangle + \langle v, w\rangle,$$

since the inequality $1 + xy \geq xz + yz$ for $x, y, z \in \{-1, 1\}$ is straightforward to verify. Observing that $u^{\otimes 8}, v^{\otimes 8}, w^{\otimes 8} \in Q_{n^8}$, it follows that

$$1 + |\langle u, v\rangle|^8 \geq |\langle u, w\rangle|^8 + |\langle v, w\rangle|^8. \tag{5.6}$$

Now suppose that $i, j \in \mathbb{N}$ are such that $\lambda(u, w) = |\langle \sigma^i u, w\rangle|$ and $\lambda(v, w) = |\langle \sigma^j v, w\rangle|$. In that case, we have

$$\begin{aligned} 1 + \lambda(u, v)^8 &\geq 1 + |\langle \sigma^i u, \sigma^j v\rangle|^8 \\ &\geq |\langle \sigma^i u, w\rangle|^8 + |\langle \sigma^j v, w\rangle|^8 \\ &= \lambda(u, w)^8 + \lambda(v, w)^8, \end{aligned}$$

where the second inequality is simply (5.6). Rearranging shows that the preceding inequality is precisely $d(u, v) \leq d(u, w) + d(v, w)$, i.e. that d satisfies the triangle inequality.

We are thus left to show that $1 - \lambda(u, v)^8 \approx 1 - \langle \widetilde{x_u}, \widetilde{x_v}\rangle$ for $u, v \in \mathcal{P}$. If $\lambda(u, v) = 1$, then both expressions are 0, so we may assume that $\lambda(u, v) \neq 1$. One direction is easy: Using the fact that if $\lambda(u, v) \neq 1$, then $\lambda(u, v)^8 \leq \lambda(u, v) \leq 1 - \frac{2}{n}$, we have

$$\begin{aligned} 1 - \langle \widetilde{x_u}, \widetilde{x_v}\rangle &\leq 1 - (1 + \tfrac{1}{4n})^{-1} \langle x_u, x_v\rangle \\ &\leq 1 - (1 + \tfrac{1}{4n})^{-1} \lambda(u, v)^8 \\ &\leq 1 - (1 - \tfrac{1}{4n}) \lambda(u, v)^8 \\ &\leq 2\left[1 - \lambda(u, v)^8\right]. \end{aligned}$$

Now, the key to satisfying the (weak) triangle inequalities is the following simple calculation:

$$\langle \widetilde{x_u}, \widetilde{x_v}\rangle \leq \langle x_u, x_v\rangle = \sum_{i=0}^{n-1} |\langle u, \sigma^i v\rangle|^8 \leq \lambda(u, v)^2 \sum_{i=0}^{n-1} |\langle u, \sigma^i v\rangle|^6 \leq (1 + \tfrac{1}{4n}) \lambda(u, v)^2,$$

where in the last inequality, we have used $u, v \in \mathcal{P}$. Thus assuming $\langle \widetilde{x_u}, \widetilde{x_v}\rangle = 1 - \delta$, we get

$$\lambda(u, v)^8 \geq \left((1 - \delta)\left(1 - \frac{1}{4n}\right)\right)^4 \geq 1 - 4\left(\delta + \frac{1}{4n}\right),$$

but $\lambda(u, v) \leq 1 - \frac{2}{n}$, hence $\delta \geq \frac{1}{4n}$ so that $\lambda(u, v)^8 \geq 1 - 8\delta$, implying $1 - \lambda(u, v)^8 \leq 8(1 - \langle \widetilde{x_u}, \widetilde{x_v}\rangle)$. □

5.3 General quotients

In the present section, we derive SDP solutions for "pseudorandom" subsets of general quotient constructions. Unlike the previous section, we will ensure that these solutions satisfy the full triangle inequalities.

5.3.1 Metrics and kernels

Fix a subgroup Γ acting on $[n]$ by permutations. We let $\psi_\Gamma = \max\{|\Gamma u| : u \in Q_n\}$ be the maximum size of any Γ-orbit. For $u, v \in Q_n$, we define

$$\lambda(u,v) = \max_{\pi\in\Gamma} |\langle u, \pi v\rangle|,$$

and for every $t \in \mathbb{N}$,

$$\alpha_t(u,v) = \sum_{\pi\in\Gamma} |\langle u, \pi v\rangle|^{2t},$$

and

$$\overline{\alpha_t}(u,v) = \frac{\alpha_t(u,v)}{\sqrt{\alpha_t(u,u)\,\alpha_t(v,v)}}.$$

Finally, we define two distance functions on Q_n corresponding to λ and α_t, respectively. For $s, t \in \mathbb{N}$, define

$$\begin{aligned} \rho_{s,t}(u,v) &= 1 - \left(\frac{1}{2} + \frac{1}{2}\lambda(u,v)^{2t}\right)^s \\ K_{s,t}(u,v) &= 1 - \left(\frac{1}{2} + \frac{1}{2}\overline{\alpha_t}(u,v)\right)^s. \end{aligned}$$

Lemma 80 *For every $t \in \mathbb{N}$, both α_t and $\overline{\alpha_t}$ are positive semi-definite kernels on Q_n. For every $s \in \mathbb{N}$, the same is true for $(u,v) \mapsto \left(\frac{1}{2} + \frac{1}{2}\overline{\alpha_t}(u,v)\right)^s$.*

PROOF: If we define $f : Q_n \to \mathbb{R}^{n^{2t}}$ by $f(u) = |\Gamma|^{-1/2}\sum_{\pi\in\Gamma}(\pi u)^{\otimes 2t}$ then $\alpha_t(u,v) = \langle f(u), f(v)\rangle$ and $\overline{\alpha_t}(u,v) = \left\langle \frac{f(u)}{\|f(u)\|_2}, \frac{f(v)}{\|f(v)\|_2}\right\rangle$. For the final implication, note that the sum of two PSD kernels is PSD, and also a positive integer power of a PSD kernel is PSD. □

From Lemma 80 and the fact that $0 \le \overline{\alpha_t}(u,v) \le 1$ for all $u, v \in Q_n$, one verifies that $K_{s,t}$ is a negative-definite kernel on Q_n, i.e. there exists a system of (unit) vectors $\{x_u\}_{u\in Q_n}$ such that $\|x_u - x_v\|^2 = K_{s,t}(u,v)$.

It is clear that both functions $\rho_{s,t}$ and $K_{s,t}$ are Γ-invariant in both coordinates. We will now show that $\rho_{s,t}$ is a metric. In Section 5.3.2, we will show that $K_{s,t}(u,v) \approx \rho_{s,t}(u,v)$ for "pseudorandom" $u, v \in Q_n$. This will motivate our analysis of the metrical properties of $K_{s,t}$ in Section 5.4.

Lemma 81 *If $0 \le a \le b \le c \le 1$ and $1+a \ge b+c$, then for any $r \ge 1$, $a^r - b^r - c^r \ge a - b - c$. In particular, for any $a, b, c \in [0,1]$, $1 + a \ge b + c$ implies $1 + a^r \ge b^r + c^r$.*

PROOF: We may assume that $a \ne 1$. In this case, write b and c as a convex combination of a and 1 as follows: $b = \frac{1-b}{1-a}a + (1 - \frac{1-b}{1-a})$ and $c = \frac{1-c}{1-a}a + (1 - \frac{1-c}{1-a})$. Now, using the fact that $x - x^r$ is concave for $x \in [0,1]$ and $r \ge 1$, write

$$(b - b^r) + (c - c^r) \ge \frac{1-b}{1-a}(a - a^r) + \frac{1-c}{1-a}(a - a^r) \ge \frac{2-b-c}{1-a}(a - a^r) \ge a - a^r,$$

where the final inequality follows from $1 + a \geq b + c$. To verify the second claim of the lemma, note that if $a > b$ or $a > c$, then $1 + a^r \geq b^r + c^r$ holds trivially. □

Corollary 82 *Let X be any set, $U : X \times X \to [0,1]$, and $s \geq 1$. If $D'(x,y) = 1 - (\frac{1}{2} + \frac{1}{2}U(x,y))^s$ is a pseudometric on X, then so is $D(x,y) = 1 - (\frac{1}{2} + \frac{1}{2}U(x,y))^{s'}$ for any $s' \geq s$.*

PROOF: The triangle inequality for D on $x, y, z \in X$ reduces to verifying

$$1 + (\tfrac{1}{2} + \tfrac{1}{2}U(x,y))^{s'} \geq (\tfrac{1}{2} + \tfrac{1}{2}U(x,z))^{s'} + (\tfrac{1}{2} + \tfrac{1}{2}U(y,z))^{s'}.$$

Since $s' \geq s$, Lemma 81 implies that this reduces to the triangle inequality for D'. □

Lemma 83 *For every $s, t \in \mathbb{N}$, $\rho_{s,t}$ is a pseudometric on Q_n.*

PROOF: By Corollary 82, it suffices to prove this for $\rho_{1,t}$. It's clear that for any $u \in Q_n$, $\rho_{1,t}(u,u) = 0$ because $\lambda(u,u) = 1$. Now fix $u, v, w \in Q_n$. The triangle inequality $\rho_{1,t}(u,v) \leq \rho_{1,t}(u,w) + \rho_{1,t}(v,w)$ reduces to verifying

$$1 + \lambda(u,v)^{2t} \geq \lambda(u,w)^{2t} + \lambda(v,w)^{2t}. \tag{5.7}$$

Suppose that $\lambda(u,w) = |\langle \pi u, w \rangle|$ and $\lambda(v,w) = |\langle v, \pi' w \rangle|$. Then,

$$\begin{aligned} \lambda(u,v)^{2t} &\geq |\langle \pi u, \pi' v \rangle|^{2t} \\ &\geq |\langle \pi u, w \rangle|^{2t} + |\langle \pi' v, w \rangle|^{2t} - 1 \qquad (5.8) \\ &= \lambda(u,w)^{2t} + \lambda(v,w)^{2t} - 1, \end{aligned}$$

where (5.8) follows just as in (5.6). □

Before turning to the precise relationship between $K_{s,t}$ and $\rho_{s,t}$, we calculate $\rho_{s,t}(u,v)$ for edges and for random pairs in Q_n.

Lemma 84 (Edges) *If $u, v \in E(Q_n)$, then $\rho_{s,t}(u,v) \leq \frac{2st}{n}$.*

PROOF: Observe that

$$\lambda(u,v)^{2t} \geq \left(1 - \frac{2}{n}\right)^{2t} \geq 1 - \frac{4t}{n},$$

hence $\rho_{s,t}(u,v) \leq 1 - (1 - \frac{2t}{n})^s \leq \frac{2st}{n}$. □

The next lemma is a straightforward application of Lemma 78 and a union bound.

Lemma 85 (Random pairs) *Suppose that $u, v \in Q_n$ are chosen independently and uniformly at random. Then,*

$$\mathbf{Pr}\left[\lambda(u,v)^{2t} \geq L\right] \leq 2\psi_G \exp\left(\frac{-L^{1/t}n}{2}\right).$$

In particular, for any $s, t \in \mathbb{N}$, if $\psi_\Gamma \leq 2^{0.1n}$, then

$$\mathbf{Pr}[\rho_{s,t}(u,v) \geq \tfrac{1}{4}] \geq \mathbf{Pr}[\lambda(u,v)^{2t} \leq \tfrac{1}{2}] \geq \tfrac{1}{2}.$$

5.3.2 Pseudorandom orbits and $\rho_{s,t} \approx K_{s,t}$

For $r \in \mathbb{N}$, define

$$\mathcal{P}_r(\eta) = \{u \in Q_n : \alpha_r(u,u) \leq 1+\eta\}$$

as the set of all elements whose Γ-orbits are not too self-correlated. Note that, by Cauchy-Schwarz, $u, v \in \mathcal{P}_r(\eta)$ implies $\alpha_r(u,v) \leq \sqrt{\alpha_r(u,u)\,\alpha_r(v,v)} \leq 1+\eta$.

The next lemma is central. It says that if $\alpha_t(u,v)$ is large and u, v are pseudorandom, then the contribution to $\alpha_t(u,v)$ comes mainly from a single large "matching" term, i.e. u is strongly correlated with some element of Γv.

Lemma 86 *Let $t > r$ and $\delta \in [0,1]$. If $u, v \in \mathcal{P}_r(\eta)$ and $\alpha_t(u,v) \geq 1-\delta$, then*

$$\lambda(u,v)^{2(t-r)} \quad \geq \quad 1-\delta-\eta.$$

Proof: We have,

$$\alpha_t(u,v) \leq \lambda(u,v)^{2t-2r} \sum_{\pi\in\Gamma} |\langle u, \pi v\rangle|^{2r} = \lambda(u,v)^{2(t-r)}\alpha_r(u,v) \leq (1+\eta)\lambda(u,v)^{2(t-r)}.$$

It follows that $\lambda(u,v)^{2(t-r)} \geq \frac{1-\delta}{1+\eta} \geq 1-\delta-\eta$. □

Theorem 87 (Weak triangle inequality for $K_{s,t}$) *For every $r, s \in \mathbb{N}$ and $u, v \in \mathcal{P}_r(\frac{1}{4n})$,*

$$\rho_{s,2r}(u,v) \approx K_{s,2r}(u,v),$$

where the implicit constant is independent of the given parameters.

Proof: Let $\eta = \frac{1}{4n}$ and $t = 2r$, and suppose that $u, v \in \mathcal{P}_r(\eta)$. If $\lambda(u,v) = 1$, then $\overline{\alpha_t}(u,v) = 1$ as well, hence $\rho_{s,t}(u,v) = K_{s,t}(u,v)$.

Now suppose that $\lambda(u,v) \neq 1$. In that case,

$$\lambda(u,v)^{2t} \leq \left(1-\tfrac{2}{n}\right)^{2t} \leq 1-\tfrac{2}{n}. \tag{5.9}$$

Assume that $\overline{\alpha_t}(u,v) = 1-\delta$ for some $\delta \in [0,1]$. Then, $\alpha_t(u,v) \geq \overline{\alpha_t}(u,v) \geq 1-\delta$, so Lemma 86 implies that $\lambda(u,v)^{2t} \geq (1-\delta-\eta)^2 \geq 1-2(\delta+\eta)$, and from (5.9), we conclude that $\delta \geq \frac{3}{4n}$. This, in turn, implies that $\eta \leq \delta/3$, which gives $\lambda(u,v)^{2t} \geq 1-3\delta$.

Finally, we observe that

$$\overline{\alpha_t}(u,v) \geq (1-\eta)\alpha_t(u,v) \geq (1-\delta/3)\alpha_t(u,v) \geq (1-\delta/3)\lambda(u,v)^{2t},$$

hence $\lambda(u,v)^{2t} \leq (1-\delta)(1+\delta/3) \leq 1-\frac{2\delta}{3}$. We have thus shown that $1-\lambda(u,v)^{2t}$ and $1-\overline{\alpha_t}(u,v)$ are within an $O(1)$ factor for all $u, v \in \mathcal{P}_r(\eta)$.□

Verification of the full triangle inequalities occurs in Appendix 5.4.

5.4 Triangle inequalities

In this section, we verify that $K_{22,t}$ is a pseudometric on $\mathcal{P}_r(\frac{1}{(4n)^2})$ for $t = O(r)$. In other words, the corresponding vectors form a valid SDP solution.

Theorem 88 *For some $t = O(r)$, $K_{22,t}$ is a pseudometric on $\mathcal{P}_r(\frac{1}{(4n)^2})$.*

PROOF: Let $\eta = \frac{1}{(4n)^2}$, and fix $u, v, w \in \mathcal{P}_r(\eta)$. To prove triangle inequality for $K_{s,t}$, it suffices to show that

$$1 + (\tfrac{1}{2} + \tfrac{1}{2}\overline{\alpha_t}(u,v))^s \geq (\tfrac{1}{2} + \tfrac{1}{2}\overline{\alpha_t}(u,w))^s + (\tfrac{1}{2} + \tfrac{1}{2}\overline{\alpha_t}(v,w))^s.$$

If both $\overline{\alpha_t}(u,w), \overline{\alpha_t}(v,w) \leq \frac{15}{16}$, then for $s = 22$, both terms are the right hand side are at most $\frac{1}{2}$, and the inequality is trivially satisfied. So we assume that $\overline{\alpha_t}(u,w) \geq \frac{15}{16}$ for the remainder of the proof.

By Corollary 82, to prove triangle inequality for $K_{22,t}$, it suffices to prove the same inequality for $K_{1,t}$ or $K_{2,t}$, i.e. one of the following inequalities.

$$\begin{aligned} 3 + \overline{\alpha_t}(u,v)[2 + \overline{\alpha_t}(u,v)] &\geq \overline{\alpha_t}(u,w)[2 + \overline{\alpha_t}(u,w)] + \overline{\alpha_t}(v,w)[2 + \overline{\alpha_t}(v,w)] \\ 1 + \overline{\alpha_t}(u,v) &\geq \overline{\alpha_t}(u,w) + \overline{\alpha_t}(v,w). \end{aligned}$$

Clearly both of these hold if $\lambda(u,w) = 1$ or if $\lambda(w,v) = 1$, so we assume this is not the case, and we are left to prove one of the following.

$$3 + \alpha_t(u,v)[2 + \alpha_t(u,v)] \geq \alpha_t(u,w)[2 + \alpha_t(u,w)] + \alpha_t(v,w)[2 + \alpha_t(v,w)] + 5\eta \quad (5.10)$$

$$1 + \alpha_t(u,v) \geq \alpha_t(u,w) + \alpha_t(v,w) + 2\eta, \quad (5.11)$$

recalling that $\overline{\alpha_t}(u,v) \leq \alpha_t(u,v) \leq (1+\eta)\overline{\alpha_t}(u,v)$ for all $u, v \in \mathcal{P}_r(\eta)$. We remark that this loss in η will be acceptable beacuse when two points $u, v \in Q_n$ are distinct, they have $|\langle u, v\rangle| \leq 1 - \frac{4}{n}$, giving us $\approx \frac{1}{n}$ slack when the orbits of u, v, and w are distinct.

Case I (Strong matching): $\lambda(u,w), \lambda(v,w) \geq 1 - \frac{1}{2t}$.

Let $\lambda(v,w) = 1 - \delta, \lambda(u,w) = 1 - \varepsilon$, and observe that $\lambda(u,v) \geq 1 - (\delta + \varepsilon)$ by (5.7). Also, since $\lambda(u,w) \neq 1$ and $\lambda(w,v) \neq 1$, we have $\delta, \varepsilon \geq \frac{4}{n}$, and in particular $\eta \leq \varepsilon\delta$. We will verify (5.11). Write,

$$\alpha_t(v,w) \leq (1-\delta)^{2t} + \left(\alpha_r(v,w) - (1-\delta)^{2r}\right)^{t/r} \leq (1-\delta)^{2t} + (\eta + 2r\delta)^{t/r}, \quad (5.12)$$

and similarly $\alpha_t(u,w) \leq (1-\varepsilon)^{2t} + (\eta + 2r\varepsilon)^{t/r}$.

Using the preceding inequalities, to prove (5.11), it suffices to show that

$$1 + (1 - (\delta+\varepsilon))^{2t} - (1-\delta)^{2t} - (1-\varepsilon)^{2t} \geq (\eta + 2r\delta)^{t/r} + (\eta + 2r\varepsilon)^{t/r} + 5\eta. \quad (5.13)$$

But we have,

$$\begin{aligned}
1+(1-(\delta+\varepsilon))^{2t}-(1-\delta)^{2t}-(1-\varepsilon)^{2t} &= \sum_{i=2}^{2t}(-1)^i\binom{2t}{i}\left[\sum_{j=1}^{i-1}\binom{i}{j}\delta^j\varepsilon^{i-j}\right] \\
&\geq 2\binom{2t}{2}\delta\varepsilon-\binom{2t}{3}3\delta\varepsilon(\delta+\varepsilon) \\
&= t(2t-1)\delta\varepsilon\left([1-2(t-1)\delta]+[1-2(t-1)\varepsilon]\right) \\
&\geq 2t(2t-1)\delta\varepsilon\left(1-\frac{2(t-1)}{2t}\right) \\
&= (2t-1)\delta\varepsilon \\
&\geq ((2r+1)\delta)^{t/r}+((2r+1)\varepsilon)^{t/r}+5\varepsilon\delta,
\end{aligned}$$

where the final inequality holds for some $t = O(r)$ chosen large enough. This proves (5.13), recalling that $\eta \leq \varepsilon\delta$.

Case II (Weak matching): $\lambda(v,w) \leq 1-\frac{1}{2t}$.

Suppose that $\alpha_t(u,w) = 1-\delta$. Our aim is to prove (5.10), which we write as

$$2\left(\alpha_t(v,w)-\alpha_t(u,v)\right)+\left(\alpha_t(v,w)-\alpha_t(u,v)\right)\left(\alpha_t(v,w)+\alpha_t(u,v)\right) \leq \delta(4-\delta)-2\eta. \quad (5.14)$$

Note that since $\overline{\alpha_t}(u,w) \geq \frac{15}{16}$, we have $\delta \leq \frac{1}{16}$. Furthermore, by Lemma 86, we have $\lambda(u,w) \geq 1-\frac{\delta+\eta}{2(t-r)}$. In particular, for $t = O(r)$ chosen large enough, we have $\lambda(u,w) \geq 1-\frac{1}{2t}$, which explains why cases I and II are exhaustive.

Now, if $\alpha_t(v,w) \geq 0.65$, then Lemma 86 implies $\lambda(v,w) \geq 1-\frac{0.35+\eta}{2(t-r)} \geq 1-\frac{0.45}{t}$ for $t \geq 2r$, which contradicts our assumption. We conclude that $\alpha_t(v,w) \leq 0.65$. In this case, we may assume that $\alpha_t(u,v) \leq 0.7$, since otherwise (5.11) is trivially satisfied, thus we have $\alpha_t(u,v), \alpha_t(v,w) \leq 0.7$.

The main idea in the "weak matching" case is to show that $\alpha_t(u,v) \gtrsim \alpha_t(v,w)$, but we cannot rely on a single "matched pair" (i.e. the triangle inqualities for λ) to do this. Instead, we argue that $\alpha_t(u,v)$ receives a large contribution on average.

To this end, write $\lambda(u,w) = 1-\beta$, and let $\pi_0 \in \Gamma$ be such that $|\langle \pi_0 u, w\rangle| = \lambda(u,w)$. Then,

$$\alpha_t(u,v) = \sum_{\pi\in\Gamma}|\langle \pi_0 u, \pi v\rangle|^{2t} \geq \sum_{\pi\in\Gamma}\left[\max(0, |\langle \pi_0 u, w\rangle| + |\langle w, \pi v\rangle| - 1)\right]^{2t} \geq \sum_{\pi\in\Gamma}\left[\max(0, |\langle w, \pi v\rangle| - \beta)\right]^{2t}.$$

Let $I = \{\pi \in \Gamma : |\langle w, \pi v\rangle| \geq \beta\}$, and observe that

$$\sum_{\pi\notin I}|\langle w, \pi v\rangle|^{2t} \leq \beta^{2t-2r}\sum_{\pi\notin I}|\langle w, \pi v\rangle|^{2r} \leq \beta^{2(t-r)}\alpha_r(w,v) \leq \beta^{2(t-r)}(1+\eta).$$

Therefore,

$$\begin{aligned}
\alpha_t(u,v) &\geq \sum_{\pi\in I}(|\langle w,\pi v\rangle| - \beta)^{2t} \\
&\geq \sum_{\pi\in I}|\langle w,\pi v\rangle|^{2t}\left(1-\frac{\beta}{|\langle w,\pi v\rangle|^{2t}}\right)^{2t} \\
&\geq \sum_{\pi\in I}|\langle w,\pi v\rangle|^{2t}\left(1-\frac{2\beta t}{|\langle w,\pi v\rangle|^{2t}}\right) \\
&\geq \left(\sum_{\pi\in\Gamma}|\langle w,\pi v\rangle|^{2t}\right) - (1+\eta)\left[\beta^{2(t-r)} - 2\beta t\right] \\
&\geq \alpha_t(w,v) - (1+\eta)\left[\delta^{2(t-r)} - (\delta+\eta)\frac{t}{t-r}\right].
\end{aligned}$$

Plugging this into (5.14) and using $\alpha_t(u,v), \alpha_t(v,w) \leq 0.7$ yields,

$$3.4(1+\eta)\left(\delta^{2(t-r)} + (\delta+\eta)\frac{t}{t-r}\right) \leq \delta(4-\delta^2) - 2\eta.$$

Now, since $\lambda(u,w) \neq 1$, we have $\lambda(u,w) \leq 1 - \frac{4}{n}$, and using Lemma 86 gives $\delta \geq \frac{2t}{n}$; in particular, $\eta \leq \delta/16$. Combining this with $\delta \leq \frac{1}{16}$, it suffices to prove

$$3.7\left(2\delta^{2(t-r)} + \delta\frac{t}{t-r}\right) \leq 3.8\delta,$$

which certainly holds for some choice of $t = O(r)$. □

5.4.1 Integrality gaps

We now discuss the consequences of Theorem 88 for integrality gaps.

Theorem 89 *Let Γ be any group acting on $[n]$ with $\psi_\Gamma \leq 2^{0.1n}$. If $|\mathcal{P}_r(\frac{1}{(4n)^2})| \geq |Q_n|(1 - n^{-2})$, then*

$$\mathsf{SDP}(Q_n/\Gamma) \leq O(r)\,\mathsf{SDP}(Q_n).$$

Proof: Let $\mathcal{P} = \mathcal{P}_r(\frac{1}{(4n)^2})$. Let $C \geq 1$ be such that $K_{22,Cr}$ is a pseudometric on $\mathcal{P}$, according to Theorem 88. By Lemma 80, $K_{22,Cr}$ is negative-definite kernel, i.e. there exists a system of vectors $\{x_u\}_{u\in Q_n}$ such that $\|x_u - x_v\|^2 = K_{22,Cr}(u,v)$.

Fix some arbitrary $u_0 \in \mathcal{P}$. We define a new solution by

$$x'_u = \begin{cases} x_u & u \in \mathcal{P} \\ x_{u_0} & u \notin \mathcal{P}. \end{cases}$$

Certainly $\{x'_u\}_{u\in Q_n}$ is a Γ-invariant vector solution that satisfies the triangle inequalites. We are left to compute the value of this solution.

First, for $(u,v) \in E(Q_n)$ with $u, v \in \mathcal{P}$, by Theorem 87 and Lemma 84, we have

$$\|x_u - x_v\|^2 = K_{22,Cr}(u,v) \approx \rho_{22,Cr}(u,v) = O(r/n).$$

Hence,

$$\sum_{uv \in E(Q_n)} \|x'_u - x'_v\|^2 \lesssim |E(Q_n)|\frac{r}{n} + 4|E(Q_n \setminus \mathcal{P})| \lesssim |E(Q_n)|\frac{r}{n},$$

using $|Q_n \setminus \mathcal{P}| \leq |Q_n|/n^2$.

On the other hand, using Theorem 87 and Lemma 85,

$$\sum_{u,v \in Q_n} \|x'_u - x'_v\|^2 \geq \sum_{u,v \in \mathcal{P}} K_{22,Cr}(u,v) \gtrsim \sum_{u,v \in \mathcal{P}} \rho_{22,Cr}(u,v) \gtrsim |\mathcal{P}|^2 \gtrsim |Q_n|^2.$$

This verifies that $\mathsf{SDP}(Q_n/\Gamma) \leq O(r)\,\mathsf{SDP}(Q_n)$. □

Using this, we can recover the best-known integrality gap.

Corollary 90 *If $\Gamma = \langle\sigma\rangle$ is the group generated by cyclic shifts, then $\mathsf{SDP}(Q_n/\Gamma) \lesssim \mathsf{SDP}(Q_n)$.*

PROOF: An argument similar to that of (5.4) shows that for n large enough and some $r = O(1)$, $|\mathcal{P}_r(\frac{1}{(4n)^2})| \geq |Q_n|(1 - n^{-2})$. □

The problem with averaging over orbits. Of course, one might hope that using techniques more sophisticated than Theorem 76, it is possible to find nice groups Γ for which $\Phi(Q_n/\Gamma) \gtrsim f(n)\Phi(Q_n)$, where $f(n) \gg \log n$. In this case, one could hope to derive stronger integrality gaps. Indeed, Bourgain and Kalai [**?**] exhibit primitive permutation groups Γ which yield such bounds. Unfortunately, the following lemma poses a problem.

Lemma 91 *For any group Γ acting on $[n]$, $\Phi(Q_n/\Gamma) \lesssim \Phi(Q_n)\log(n|\Gamma|)$.*

PROOF:[Proof sketch] Let $k \in \mathbb{N}$ and define $F_k : Q_n \to \{0,1\}$ by $F_k(u) = 1$ if there exist an $i \in [n]$ such that $u_i, u_{i+1}, \ldots, u_{i+k} < 0$. Let

$$S_k = \{u \in Q_n : F_k(v) = 1 \text{ for some } v \in \Gamma u\}. \tag{5.15}$$

It is clear that S_k is Γ-invariant. Now, there exists a $k \leq \log(n|\Gamma|)$ such that $|S_k|/|Q_n| \in [1/3, 2/3]$, since for a randomly chosen $u \in Q_n$, a fixed sequence will satisfy $u_i, u_{i+1}, \ldots, u_{i+k} < 0$ with probability 2^{-k}, and there are at most $n|\Gamma|$ such sequences under consideration in (5.15).

By a standard analysis, a randomly chosen $u \in S_k$ will, with high probability, have only $O(k)$ pivotal bits, implying that $|E(u, \overline{S}_k)|$ is typically $O(k) = O(\log(n|\Gamma|))$, which implies that $|E(S_k, \overline{S}_k)| \leq O(\log(n|\Gamma|))|S_k|$, and yields $\Phi(S_k) \leq O(\log(n|\Gamma|))|Q_n|^{-1} \approx O(\log(n|\Gamma|))\Phi(Q_n)$. □

The preceding lemma is problematic, because in order for $|\mathcal{P}_r(1/n^2)|$ to be almost everything, one has to take $r \gtrsim \frac{\log|\Gamma|}{\log n}$, This is because in a sum like

$$\alpha_t(u,v) = \sum_{\pi\in\Gamma} |\langle u, \pi v\rangle|^{2t}$$

the terms not corresponding to $\pi = \mathrm{id}$ can generally only be expected to be of order $(n^{-1/2})^{2t} = n^{-t}$, but there are possibly $|\Gamma|$ of these terms, implying that we need $t \approx \frac{\log|\Gamma|}{\log n}$ in order for these terms to have total magnitude $o(1)$. In the next section, we discuss how different vector solutions can be used with $r = O(1)$ for a specific example with $|\Gamma| \approx 2^{n^{\Omega(1)}}$.

5.5 Larger orbits: Permutations of the rows

In this section, we discuss $m \times n$ sign matrices with $m = \mathrm{poly}(n)$, where Γ includes all permutations of the rows, meaning that our previous SDP solutions would not be adequate (as the orbits are now huge). Still, we give a (weak) SDP solution with $\mathsf{SDP}_{O(1)}(Q_{mn}/\Gamma) \approx \mathsf{SDP}(Q_n)$. Unfortunately, it is not difficult to see that $\Phi(Q_{mn}/\Gamma) \approx \Phi(Q_n)\log n$, meaning that we again achieve only an $\Omega(\log\log N)$ integrality gap. It is possible that a hierarchical version of this construction could give larger gaps.

5.5.1 The metric

For every $m, n \in \mathbb{N}$, let $X_{m,n} = \left(\frac{1}{\sqrt{n}}\{-1,1\}^n\right)^m \subseteq \mathbb{R}^{mn}$ be the space of sequences $(A_1, A_2, \ldots, A_m)$ with each $A_i \in \{\frac{-1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}^n$. The symmetric group S_m acts in a natural way on $X_{m,n}$: For $\pi \in S_m$, we have $\pi(A) = \pi(A_1, \ldots, A_m) = (A_{\pi(1)}, \ldots, A_{\pi(m)})$. Let $\mathcal{X}_{m,n}$ be the set of orbits of $X_{m,n}$ under the S_m action. We define

$$\lambda_t(A,B) = \frac{1}{m}\max_{\pi:[m]\to[m]} \sum_{i=1}^m |\langle A_i, B_{\pi(i)}\rangle|^{2t},$$

where the maximum is over all bijections π.

Lemma 92 *For any $A, B, C \in X_{m,n}$ and any $t \in \mathbb{N}$, we have*

$$\lambda_t(A,B) \geq \lambda_t(A,C) + \lambda_t(B,C) - 1.$$

Proof: Let $\pi, \pi' : [m] \to [m]$ be such that $\lambda_t(A,C) = \frac{1}{m}\sum_{i=1}^m |\langle A_i, C_{\pi(i)}\rangle|^{2t}$ and $\lambda_t(B,C) = \frac{1}{m}\sum_{i=1}^m |\langle B_i, C_{\pi'(i)}\rangle|^{2t}$. Then letting $\sigma = (\pi')^{-1}\circ\pi$, we have

$$\begin{aligned}
\lambda_t(A,B) &\geq \frac{1}{m}\sum_{i=1}^m |\langle A_i, B_{\sigma(i)}\rangle|^{2t} \\
&\geq -1 + \frac{1}{m}\sum_{i=1}^m |\langle A_i, C_{\pi(i)}\rangle|^{2t} + \frac{1}{m}\sum_{i=1}^m |\langle B_{\sigma(i)}, C_{\pi(i)}\rangle|^{2t} \qquad (5.16)\\
&= -1 + \lambda_t(A,C) + \lambda_t(B,C).
\end{aligned}$$

□

Next we define, for every $s, t \in \mathbb{N}$, the distance function $\rho_{s,t}(A, B) = 1 - (\frac{1}{2} + \frac{1}{2}\lambda_t(A, B))^s$.

Claim 93 *For every $s, t \in \mathbb{N}$, $\rho_{s,t}$ is a metric on $\mathcal{X}_{m,n}$.*

PROOF: First, it's clear that $\rho_{s,t}(A, B) = \rho_{s,t}(\pi A, B)$ for all $\pi \in S_m$ and $A, B \in X_{m,n}$. Also, $\rho_{s,t}(A, A) = 0$ because $\lambda_t(A, A) = 1$.

Now, consider $A, B, C \in X_{m,n}$. The triangle inequality $\rho_{s,t}(A, B) \leq \rho_{s,t}(A, C) + \rho_{s,t}(B, C)$ reduces to verifying

$$1 + (\tfrac{1}{2} + \tfrac{1}{2}\lambda_t(A, B))^s \geq (\tfrac{1}{2} + \tfrac{1}{2}\lambda_t(A, C))^s + (\tfrac{1}{2} + \tfrac{1}{2}\lambda_t(B, C))^s.$$

Write this as

$$1 + x^s \geq y^s + z^s. \tag{5.17}$$

Then $x, y, z \in [0, 1]$ since $\lambda_t(A, B) \in [0, 1]$ for all $A, B \in X_{m,n}$. Combining this with the fact that $1 + x \geq y + z$ from Lemma 92, we conclude that (5.17) holds. □

Finally, we analyze the behavior of $\rho_{s,t}$ on "edges" of $X_{m,n}$ and on random pairs. If $A, A' \in X_{m,n}$, we write $A \sim A'$ if $\|A - A'\|_2^2 = \frac{4}{n}$ (i.e. the hamming distance between A and A' is one).

Lemma 94 (Edges) *If $A, A' \in X_{m,n}$ with $A \sim A'$, then $\rho_{s,t}(A, A') \leq \frac{2st}{mn}$.*

PROOF: Observe that

$$\lambda_t(A, A') \geq \frac{1}{m}\sum_{i=1}^{m} |\langle A_i, A'_i\rangle|^{2t} = \frac{1}{m}\left(m - 1 + \left(1 - \frac{2}{n}\right)^{2t}\right) \geq 1 - \frac{4t}{mn}.$$

hence $\rho_{s,t}(A, A') = 1 - (1 - \frac{2t}{mn})^s \leq \frac{2st}{mn}$. □

Lemma 95 (Random pairs) *Suppose that $A, B \in X_{m,n}$ are chosen independently and uniformly at random. Then*

$$\mathbf{Pr}\left[\lambda_t(A, B) \geq Ln^{-t}\right] \leq 2me^{-\frac{1}{2}L^{1/t}}.$$

In particular, for any $s, t \in \mathbb{N}$, we have $\mathbf{Pr}[\rho_{s,t}(A, B) \geq \frac{1}{4}] \geq \mathbf{Pr}[\lambda_t(A, B) \leq \frac{1}{2}] \geq \frac{1}{2}$.

5.5.2 An equivalent negative-definite kernel

We now define, for any $t \in \mathbb{N}$, two kernels. For $A, B \in X_{m,n}$, let

$$\alpha_t(A, B) = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{m} |\langle A_i, B_j\rangle|^{2t},$$

and

$$\overline{\alpha_t}(A, B) = \frac{\alpha_t(A, B)}{\sqrt{\alpha_t(A, A)\,\alpha_t(B, B)}}.$$

Lemma 96 *For every $t \in \mathbb{N}$, α_t and $\overline{\alpha_t}$ are both positive semi-definite kernels on $\mathcal{X}_{m,n}$.*

PROOF: Define maps $f, \overline{f} : X_{m,n} \to \mathbb{R}^{n^t}$ by $f(A) = \frac{1}{\sqrt{m}} \sum_{i=1}^m A_i^{\otimes 2t}$ and $\overline{f}(A) = f(A)/\|f(A)\|_2$. Then $\langle f(A), f(B)\rangle = \alpha_t(A,B)$ and $\langle \overline{f}(A), \overline{f}(B)\rangle = \overline{\alpha_t}(A,B)$. Clearly f is invariant under the S_m action on $X_{m,n}$. □

For every $s, t \in \mathbb{N}$, define a negative-definite kernel on $\mathcal{X}_{m,n}$ by

$$K_{s,t}(A,B) = 1 - \left(\frac{1}{2} + \frac{\overline{\alpha_t}(A,B)}{2}\right)^s.$$

For $r \in \mathbb{N}$, let

$$\mathcal{N}_r(\eta) = \left\{A \in X_{m,n} : |\langle A_i, A_j\rangle|^{2r} \le \frac{\eta}{m} \quad \forall i \neq j \in [m]\right\}$$

be the set of elements in $X_{m,n}$ with small self-correlation. In particular, $A \in \mathcal{N}_r(\eta)$ implies that $\alpha_r(A,A) \le 1 + \eta$. Using Cauchy-Schwarz, we have $\alpha_r(A,B) \le \sqrt{\alpha_r(A,A)\,\alpha_r(B,B)}$, hence $A, B \in \mathcal{N}_r(\eta)$ implies $\alpha_r(A,B) \le 1 + \eta$ as well.

Lemma 97 (Heavy matchings) *Suppose that $t \ge 2r$, $\eta \le \frac{1}{16}$, $\delta \in [0,1]$, and $A, B \in \mathcal{N}_r(\eta)$. Then $\alpha_t(A,B) \ge 1 - \delta$ implies that*

$$\lambda_t(A,B) \ge 1 - (10\delta + 2\eta)$$

PROOF: Define $\alpha_i = 1 - \sum_{j=1}^m |\langle A_i, B_j\rangle|^{2t}$ and $\beta_i = \max_{j\in[m]} |\langle A_i, B_j\rangle|$. Then,

$$1 - \alpha_i \le \beta_i^{2t-2r} \sum_{j=1}^m |\langle A_i, B_j\rangle|^{2r} \le \beta_i^{2(t-r)} \sqrt{\|A_i\|_2 \cdot \alpha_r(B,B)} \le \beta_i^{2(t-r)}(1+\eta),$$

so we have

$$\beta_i^{2t} \ge \left(\frac{1-\alpha_i}{1+\eta}\right)^{\frac{t}{t-r}} \ge 1 - \frac{t}{t-r}(\alpha_i + \eta) \ge 1 - 2(\alpha_i + \eta). \tag{5.18}$$

Now suppose that

$$\alpha_t(A,B) = \frac{1}{m}\sum_{i=1}^m (1 - \alpha_i) \ge 1 - \delta.$$

Let $S = \{i \in [m] : \alpha_i \le \frac{1}{8}\}$. Clearly $|S| \ge (1 - 8\delta)m$ since $\sum_{i=1}^m \alpha_i \le \delta m$. Define a mapping $\pi : S \to [m]$ by $\pi(i) = \operatorname{argmax}_{j\in[m]} |\langle A_i, B_j\rangle|^{2t}$.

We claim that π is injective. Observe that for $i \in S$, (5.18) implies that $\beta_i^{2t} \ge 1 - 2\left(\frac{1}{8} + \eta\right) \ge \frac{5}{8}$. So if $\pi(i) = \pi(j)$ for $i \neq j \in S$, then we have

$$|\langle A_i, A_j\rangle|^{2t} \ge |\langle A_i, B_{\pi(i)}\rangle|^{2t} + |\langle A_j, B_{\pi(i)}\rangle|^{2t} - 1 \ge \frac{1}{4},$$

which contradicts the fact that for $A \in \mathcal{N}_r(\eta)$, we have $|\langle A_i, A_j\rangle|^{2t} \le |\langle A_i, A_j\rangle|^{2r} \le \frac{\eta}{m} \le \frac{1}{16}$.

Since π is injective, it follows that

$$\lambda_t(A,B) \geq \frac{1}{m}\sum_{i\in S}\beta_i^{2t} \geq \frac{1}{m}\sum_{i\in S}(1-2(\alpha_i+\eta)) \geq \frac{|S|}{m} - 2(\delta+\eta) \geq 1-(10\delta+2\eta).$$

□

Even though $K_{s,t}$ may not be a metric, we show that it is always close to $\rho_{s,t}$.

Theorem 98 (Bi-lipschitz equivalence) *There exists a universal constant $C \geq 1$ such that for any $t \geq 2r$, the distance functions $K_{s,t}$ and $\rho_{s,t}$ are C-bi-lipschitz equivalent when restricted to $\mathcal{N}_r(\frac{1}{20mn})$.*

PROOF: If $A = \pi(B)$ for some $\pi \in S_m$, then clearly $\lambda_t(A,B) = \overline{\alpha_t}(A,B) = 1$, hence $\rho_{s,t}(A,B) = K_{s,t}(A,B) = 0$. Let $\eta = \frac{1}{20mn}$.

Consider $A, B \in \mathcal{N}_r(\eta)$ where A and B are in different equivalence classes of $\mathcal{X}_{m,n}$. Then clearly we have

$$\lambda_t(A,B) \leq \frac{1}{m}\left(m-1+\left(1-\frac{2}{n}\right)^{2t}\right) \leq 1-\frac{2}{mn}. \tag{5.19}$$

Now suppose that $\overline{\alpha_t}(A,B) = 1-\delta$ for some $\delta \in [0,1]$. In that case, $\alpha_t(A,B) \geq \overline{\alpha_t}(A,B) \geq 1-\delta$, so Lemma 97 implies that $\lambda_t(A,B) \geq 1-(10\delta+2\eta)$. From (5.19), we conclude that $\delta \geq \frac{1}{6mn}$. This, in turn, implies that $\eta \leq \delta/3$, which gives $\lambda_t(A,B) \geq 1-11\delta$.

Finally, we observe that

$$\overline{\alpha_t}(A,B) \geq (1-\eta)\alpha_t(A,B) \geq (1-\delta/3)\alpha_t(A,B) \geq (1-\delta/3)\lambda_t(A,B),$$

hence $\lambda_t(A,B) \leq (1-\delta)(1+\delta/3) \leq 1-\frac{2\delta}{3}$. We conclude that $1-\lambda_t(A,B)$ and $1-\overline{\alpha_t}(A,B)$ are within an $O(1)$ factor of each other for all $A, B \in \mathcal{N}_r(\eta)$. This immediately implies that $K_{s,t}(A,B)$ and $\rho_{s,t}(A,B)$ are within an $O(1)$ multiplicative factor. □

The final result of this section concerns how large one needs to choose r (and hence t) so that $\mathcal{N}_r(\frac{1}{20mn})$ contains most of the points of $X_{m,n}$.

Lemma 99 *Let $\eta = \frac{1}{20mn}$, and consider $A \in X_{m,n}$ chosen uniformly at random. For any $\tau = \tau(m,n)$, there exists a choice of $r \approx \frac{\log m}{\log n - \log\log\frac{m}{\tau}}$ for which*

$$\mathbf{Pr}\left[A \notin \mathcal{N}_r(\eta)\right] \leq \tau.$$

PROOF: Let $A_i, A_j \in \{\frac{-1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}^n$ be chosen independently at random, then

$$\mathbf{Pr}\left[A \notin \mathcal{N}_r(\eta)\right] \leq m^2\mathbf{Pr}\left[|\langle A_i, A_j\rangle| \geq \left(\frac{1}{20m^2n}\right)^{1/2r}\right] \leq 2m^2\exp\left(\frac{-n}{2(20m^2n)^{1/r}}\right).$$

Simplifying yields the desired conclusion. □

The point is that we can choose any $m = \text{poly}(n)$ and $\tau = 2^{-n^{0.1}}$, and we still only need $r = O(1)$.

5.6 Discussion and future directions: PSD flows and triangle inequalities

In this section, we discuss the question of whether $\mathsf{SDP}_{O(1)}(G) \approx \mathsf{SDP}(G)$ for every graph G, i.e. whether the weak triangle inequalities can always be converted to strong triangle inequalities with only an $O(1)$ loss. This is most nicely stated in the setting of the SDP dual.

Let $G = (V, E)$ be a finite, undirected graph, and for every pair $u, v \in V$, let $\mathcal{P}_{uv}$ be the set of all paths between u and v in G. Let $\mathcal{P} = \bigcup_{u,v \in V} \mathcal{P}_{uv}$. A *flow in* G is a mapping $F : \mathcal{P} \to \mathbb{R}_{\geq 0}$. We define, for every vertex $(u, v) \in E$, the *congestion on* (u, v) as

$$C_F(u, v) = \sum_{p \in \mathcal{P} : (u,v) \in p} F(p).$$

For any $u, v \in V$, let $F[u, v] = \sum_{p \in \mathcal{P}_{uv}} F(p)$ be the amount of flow sent between u and v.

The standard "maximum concurrent flow" problem is simply

$$\mathsf{mcf}(G) = \text{maximize } \Big\{ D : \forall u, v, F[u, v] \geq D \text{ and } \forall (u, v) \in E, C_F(u, v) \leq 1. \Big\}$$

If we define the symmetric matrix

$$A_{u,v} = F[u, v] - D + \mathbf{1}_{\{(u,v) \in E\}} - C_F(u, v),$$

then certainly every feasible flow of value D satisfies $A_{u,v} \geq 0$ for all $u, v \in V$. In fact, we can combine the two types of flow constraints (demand/congestion) together, and get the same thing:

Exercise: $\mathsf{mcf}(G) = \max\{D : A_{u,v} \geq 0 \, \forall u, v\}$

Now, the dual of the Sparsest Cut SDP is precisely the same thing, but with a *global* constraint on A, instead of having a constraint per entry:

$$\mathsf{SDP}(G) = \max\{D : L(A) \succeq 0\}.$$

Here, $L(A)$ denotes the Laplacian of A, i.e.

$$L(A)_{i,j} = \begin{cases} \sum_{k \neq i} A_{i,k} & i = j \\ -A_{i,j} & \text{otherwise.} \end{cases}$$

and we write $L(A) \succeq 0$ to denote that $L(A)$ is positive semi-definite.

Now, if we write, for some $\kappa \geq 1$,

$$A^{(\kappa)}_{u,v} = F[u, v] - D + \kappa \cdot \mathbf{1}_{\{(u,v) \in E\}} - C_F(u, v),$$

then clearly

$$\max\{D : A^{(\kappa)}_{u,v} \geq 0 \, \forall u, v\} \geq \max\{D : A_{u,v} \geq 0 \, \forall u, v\}$$

because we have bumped up the edge capacities. On the other hand, given an $A^{(\kappa)}$-feasible flow of value D, we can always get an actual feasible flow with value D/κ by simply scaling down the flow by factor $1/\kappa$, i.e.

$$\max\{D : A^{(\kappa)}_{u,v} \geq 0 \, \forall u, v\} = \kappa \cdot \max\{D : A_{u,v} \geq 0 \, \forall u, v\}.$$

Question 100 *Is the same kind of thing true for "PSD-flows"? In other words, are*

$$\max\{D : L(A) \succeq 0\} \qquad \textit{and} \qquad \max\{D : L(A^{(\kappa)}) \succeq 0\}$$

related by a factor depending only on κ?

If this question has a positive answer, then it makes integrality gaps for the Sparsest Cut SDP much easier to understand, because SDP duality shows that $\mathsf{SDP}(G) = \max\{D : L(A) \succeq 0\}$ while $\mathsf{SDP}_\kappa(G) = \max\{D : L(A^{(\kappa)}) \succeq 0\}$.

The answer to this question is affirmative if we can decouple the $L(A) \succeq 0$ constraint into two constraints, i.e. let $X_{u,v} = F[u,v] - D$ for $u \neq v$ and let $Y_{u,v} = \mathbf{1}_{(u,v)\in E} - C_F(u,v)$.

Question 101 *Can we relate (e.g. within constant factors)* $\max\{D : L(A) \succeq 0\}$ *to* $\max\{D : L(X) \succeq 0$ *and* $L(Y) \succeq 0\}$ *as we can for normal flows? It is easy to check that this would give an affirmative answer to Question 100.*