Optimal Approximabilities beyond CSPs

Euiwoong Lee

CMU-CS-17-115
May 2017

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Venkatesan Guruswami, Chair
Anupam Gupta
R. Ravi
Ryan O'Donnell
Ola Svensson, EPFL

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

This research was sponsored by the National Science Foundation under grant number CCF111525 and CCF1526092, the Simons Foundation under grant 360895, the United States Israeli Binational Science Foundation under grant number 2008293, and the Samsung Scholarship. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government or any other entity.

Keywords: Approximation algorithms, Hardness of approximation, Combinatorial optimization, Constraint satisfaction problems, Coloring problems, Covering problems, Cut problems, Convex relaxations, LP/SDP hierarchies, Unique games conjecture

To Yeasul.

Abstract

The theory of approximation algorithms has seen great progress since the nineties, and the optimal approximation ratio was discovered for many fundamental combinatorial optimization problems. This progress has been most successfully applied to a subclass of optimization problems called maximum constraint satisfaction problems (MAX CSPs), where there is a simple semidefinite programming based algorithm provably optimal for every problem in this class under the Unique Games Conjecture. Such a complete understanding is not known for other basic classes such as coloring, covering, and graph cut problems.

This thesis tries to expand the frontiers of approximation algorithms, with respect to the range of optimization problems as well as mathematical tools for algorithms and hardness. We show tight approximabilities for various fundamental problems in combinatorial optimization beyond MAX CSP. It consists of the following five parts:

1. CSPs: We introduce three variants of MAX CSP, called HARD CSP, BALANCE CSP, and SYMMETRIC CSP. Our results show that current hardness theories for MAX CSP can be extended to its generalizations (HARD CSP, BALANCE CSP) to prove much stronger hardness, or can be significantly simplified for a special case (SYMMETRIC CSP).
2. Applied CSPs: Numerous new optimization problems have emerged since the last decade, as computer science has more actively interacted with other fields. We study three problems called UNIQUE COVERAGE, GRAPH PRICING, and decoding LDPC codes, motivated by networks, economics, and error-correcting codes respectively. Extending the tools for MAX CSP, we show nearly optimal hardness results or integrality gaps for these problems.
3. Coloring: We study the complexity of hypergraph coloring problems when instances are promised to have a structure much stronger than admitting a proper 2-coloring, and prove both algorithmic and hardness results. For both algorithms and hardness, we give unified frameworks that can be used for various settings.

4. H-TRANSVERSAL: We study the problem H-TRANSVERSAL, where given a graph G and a fixed "pattern" graph H, the goal is to remove the minimum number of vertices from G to make sure it does not include H as a subgraph. We show an almost complete characterization of the approximability of H-TRANSVERSAL depending on properties of H. One of our algorithms reveals a new connection between path transversal and graph partitioning.

5. We also study various cut problems on graphs, where the goal is to remove the minimum number of vertices or edges to cut some desired paths or cycles. We present a general tool called *length-control dictatorship tests* to prove strong hardness results under the Unique Games Conjecture, which allow us to prove improved hardness results for multicut, bicut, double cut, interdiction, and firefighter problems.

Acknowledgments

There were two major moments during my undergraduate and masters study that led me to pursue a Ph.D. in theoretical computer science. The first moment happened in 2008 when I found out that my first research project, a cute theory problem given by a networking lab, was actively studied in the mainstream theory community. The paper that introduced this problem in computer science was titled *On Profit-Maximizing Envy-free Pricing*. The other happened at FOCS 2011 where I was fascinated by one paper titled *Lasserre Hierarchy, Higher Eigenvalues, and Approximation Schemes for Quadratic Integer Programming with PSD Objectives,* due to its elegant ideas that apply linear algebraic tools to combinatorial problems.

Naturally I became a Ph.D. student of Venkat Guruswami who wrote these two papers, and that turned out to be one of the best decisions I have made. From giving numerous one-on-one lectures, to immediately but carefully answering my technical (silly) emails at night, to encouraging me to finally resolve the above first research problem, Venkat really has been a great *student-adaptive* advisor as he told me in our first meeting at CMU, who could be hands-off or hands-on as I wanted. For all these wonderful academic experiences and even beyond, thank you Venkat!

I also would like to thank Anupam Gupta, Ryan O'Donnell, R. Ravi, and Ola Svensson for serving on the thesis committee and providing the helpful comments on the thesis. As I mentioned during my defense, I was really honored to present the outcome of my Ph.D. research to the team of the world's leading experts in approximation algorithms and hardness of approximation.

I have had the privilege of working with a number of people during my Ph.D study: Kristóf Bérczi, Vijay Bhattiprolu, Karthekeyan Chandrasekaran, Julia Chuzhoy, Marek Cygan, Badih Ghazi, Mrinalkanti Ghosh, Sivakanth Gopi, Guru Guruganesh, Venkatesan Guruswami, Sungjin Im, Tamás Király, Janardhan Kulkarni, Jason Li, Konstantin Makarychev, Yury Makarychev, Jinoh Oh, Ariel Procaccia, Melanie Schmidt, Kijung Shin, Mohit Singh, Sahil Singla, Madhur Tulsiani, David Witmer, John Wright, and Chao Xu.

While not all projects led to a publication, I can confidently say that every project and every meeting was both fun and inspiring. Most results in this thesis could not have been done without their technical mastery and encouragement.

In my early years as a Ph.D. student, I enjoyed taking or auditing great theory courses that helped me have a broader perspective on theoretical computer science. Many thanks to the professors who taught me in the classroom: Venkat, Ryan, Gary, Mor, Bernhard, Manuel, Steven, and Anupam at CSD, Alan, Po-Shen, and Tom at math, and Ravi at Tepper.

I would like to thank my mentors during my undergraduate and masters study at Caltech. Leonard Schulman, Katrina Ligett, and Steven Low offered not only interesting courses that grew my passion for computer science and theory but also great research opportunities where I could take my first steps as a researcher. I am also greatly indebted to my mentors during my internships at Toyota Technological Institute at Chicago and Microsoft Research Redmond. I truly learned a lot from Yury Makarychev and Julia Chuzhoy at TTIC and Kostya Makarychev at MSR from mathematical knowledge to their passion for research.

I would like to thank the students and postdocs in our theory group for being great friends and coworkers. Jeremiah, Richard, Jakub, Ameya, John, Julian, Sarah, Naama, Vijay, Guru, Nika, Jason, Nic, David (both Witmer and Wajc), Colin, Shen, Goran, Yu, Ankit, and many more. Our current friendly and collaborative culture has crucially helped me both academically and emotionally throughout my time here, and I hope that it lasts for a long time in the GHC building. I also thank Deb Cavlovich for making everything go smoothly, especially about my unique external funding situation.

I would also like to thank my great friends from the CMU Korean community. Sungho, Jungeun, Daegun, Eunyong, Junsung, Minkyung, Jaejun, Joohyung, Heewook, Dohyun, Sejoon, Woong Young, Jay-yoon, Suyoun, Seungwhan and many more. I have learned numerous things from these friends about science, life, and beyond.

It is impossible to express the debt of gratitude that I owe my family. My mom, dad, and Euiin have given endless support and encouragement for my entire life. Though we are living in different continents, my heart is always with them.

Finally, I would like to thank my wife Yeasul for being a source of great joy and inspiration. Marrying this intelligent, beautiful, and courageous person made this thesis the second most significant achievement during the past five years. But this thesis is still for her.

Contents

Chapter 1

Introduction

1.1 Combinatorial Optimization

Combinatorial optimization is a topic that involves searching for an optimal object from a finite set of objects. Every combinatorial optimization problem studied in this thesis has the following general structure.

> Problem P.
>
> Input: An object I that implicitly defines the feasible set X_I and the objective function $f_I : X_I \to \mathbb{R}^+$.
>
> Output: A feasible solution $x \in X_I$.
>
> Goal: Minimize (or maximize) $f_I(x)$.

Let $n := |I|$ denote the length of the string representing I. Every problem in this thesis is an *NP-optimization problem* in the sense that (1) the length of every $x \in X_I$, (2) the time to check whether $x \in X_I$, (3) the time to compute $f_I(x)$ are all bounded by a polynomial in n. The biggest bottleneck is that the feasible set X_I may contain as many as $\exp(\text{poly}(n))$ feasible solutions, which makes the trivial algorithm of trying every $x \in X_I$ prohibitively inefficient. Therefore, one of the biggest goals of combinatorial optimization is to design an algorithm that computes the optimal feasible solution in polynomial time.

Since its establishment as a coherent discipline in the 1950's, it has been actively studied in mathematics, operations research, and computer science. Its inherent interdisciplinarity is based on the three *pillars* that are indispensable for a complete understanding of each problem in the topic — algorithms, hardness, and mathematical programming. Edmonds' seminal paper for MAXIMUM MATCHING [Edm65] pioneered the relationship between efficient algorithms and the properties of associated polyhedra formed by natural linear programming (LP) relaxations. This tight relationship has been proven to hold for many combinatorial optimization problems; in fact, one of the most comprehensive textbooks in the field is devoted to it [Sch03].

On the other hand, the same paper successfully established polynomial time as the major criterion to measure an algorithm's efficiency. This work is followed by the Cook-Levin theorem [Coo71, Lev73] and Karp's 21 NP-complete problems [Kar72] that show NP-completeness of some fundamental combinatorial optimization problems, implying they will not admit an efficient algorithm unless $\mathbf{P} = \mathbf{NP}$. Active subsequent research has successfully identified the complexity of each problem, so already in the early twenty-first century "almost every combinatorial optimization problem has since been either proven to be polynomial time solvable or NP-complete." [Sch03]

1.2 Approximation Algorithms

Unfortunately, numerous combinatorial optimization problems are NP-complete, so polynomial time algorithms are unlikely to exist for those problems. Approximation algorithms are considered one of the most natural ways to circumvent this difficulty.

Definition 1.2.1. *For a maximization problem P, an algorithm $\mathcal{A}$ is called a c-approximation algorithm ($c \leq 1$) for P if for every input I, $\mathcal{A}$ runs in time $\mathrm{poly}(|I|)$ and outputs $\mathcal{A}(I) \in X_I$ such that*

$$f_I(\mathcal{A}(I)) \geq c \cdot \max_{x \in X_I} f_I(x).$$

For a minimization problem P, an algorithm $\mathcal{A}$ is called a c-approximation algorithm ($c \geq 1$) for P if for every input I, $\mathcal{A}$ runs in time $\mathrm{poly}(|I|)$ and outputs $\mathcal{A}(I) \in X_I$ such that

$$f_I(\mathcal{A}(I)) \leq c \cdot \min_{x \in X_I} f_I(x).$$

The number c is called an approximation ratio *or* approximation factor.

Similarly to the study of designing an exact algorithm, the study of approximation algorithms and hardness of approximation have produced beautiful results that highlight the synergy between algorithms, hardness, and mathematical programming. From the algorithms side, the work of Goemans and Williamson [GW95] introduced the first application of semidefinite programming (SDP) to approximation algorithms, strictly improving the previous algorithms based on linear programming (LP). The search for more powerful convex relaxations beyond LPs and SDPs has produced several LP and SDP hierarchies, including Lovász-Schrijver [LS91], Sherali-Admas [SA90], and Sum-of-Squares [Par00, Las01].

From the hardness side, the celebrated PCP theorem [ALM+98, AS98] first proved that there exists a universal constant $c < 1$ such that it is NP-hard to approximate MAX 3-SAT within a factor of c. The parallel repetition theorem [Raz98] and the introduction of long codes [BGS98] created a framework to prove strong hardness results for many problems, culminating in Håstad's optimal inapproximability results for various constraint satisfaction problems [Hås01]. The Unique Games Conjecture [Kho02b] (UGC) suggested even tighter relationships between mathematical programming and hardness of approximation, though its truth seems far from being settled. Raghavendra [Rag08] showed that assuming the UGC, for every problem in the wide class of problems called MAX CSP, no polynomial time algorithm outperforms an algorithm based on a natural SDP relaxation. Conversely, ideas from computational hardness results often led to limitations of convex relaxations [KV05, Sch08, Tul09, BCK15] for some problems by showing *integrality gaps*.

1.2.1 Three Examples

For many decades, numerous combinatorial optimization problems have been studied in terms of their exact solvability and approximability. While precisely classifying them into well-defined categories is hard, there are several classes of problems that share some conceptual/technical properties and techniques to study them. In this thesis, we focus on three classes of problems, namely Constraint Satisfaction Problems (CSPs), coloring problems, and covering problems. These three classes have been a focal point of combinatorial optimization and approximation algorithms research. They include 13 out of 21 of Karp's NP-hard problems [Kar72] and 13 out of 23 problems covered in Vazirani's textbook [Vaz01]. The more recent textbook by Williamson and Shmoys [WS11] also spend 30 out of its 60 technical sections on these three classes. We introduce the following three problems, one from each class, whose approximabilities have been actively studied.

CSPs. From 3-SAT, CSPs have always played a crucial role in both exact and approximate optimization. One of the most well-known CSPs in the approximation algorithms literature is the following simple problem.

MAX CUT

Input: A graph $G = (V, E)$.

Output: A partition (A, B) of V (i.e., $A \cup B = V$, $A \cap B = \emptyset$).

Goal: Maximize the number of edges that have one endpoint in A and one endpoint in B.

It belongs to a wide class of problems known as MAX CSP, where the input consists of a set of variables and a set of constraints whose satisfaction depends on a small number of variables, and the goal is to find an assignment to the variables to maximize the number of satisfied constraints (see Section 2.1 for the precise definition).

For MAX CUT, a simple greedy algorithm achieves a 0.5-approximation. Goemans and Williamson [GW95] designed a 0.878-approximation algorithm that spurred the study of applying SDPs to approximation algorithms. From the hardness side, the PCP theorem [ALM+98, AS98] proved that there exists $c > 0$ such that it is NP-hard to approximate within a $(1-c)$ factor. Hastad's optimal inapproximability results for some other problems in MAX CSP, combined with the gadget of Trevisan et al. [TSSW00], implied that MAX CUT is NP-hard to approximate within a factor of $16/17$. Finally, Khot et al. [KKMO07] showed that the 0.878-approximation algorithm by Goemans and Williamson is tight under the Unique Games Conjecture.

While the approximability of MAX CUT is well understood, there are still important open questions regarding the approximability of CSPs, motivated by both theory and applications. Building on the previous algorithms and hardness techniques for MAX CUT and other CSPs, this thesis answers some of these questions.

Coloring. Given a map of countries, can we always color them with four colors so that no two adjacent countries get the same color? Starting from this natural question in mathematics, graph coloring has a history longer than computer science. Surprisingly, the approximability of the following basic optimization problem is still wide open.

COLORING 3-COLORABLE GRAPHS

Input: A graph $G = (V, E)$ that admits a 3-coloring $c : V \to \{1, 2, 3\}$ such that every edge $(u, v) \in E$ satisfies $c(u) \neq c(v)$.

Output: A coloring $d : V \to \{1, 2, \dots, \chi\}$ such that every edge $(u, v) \in E$ satisfies $d(u) \neq d(v)$.

Goal: Minimize χ.

This problem has been an interesting testbed for both combinatorial ([Wig83, Blu94, KT12]) and SDP-based algorithms ([KMS98a, AC06, Chl07]). Currently the best known approximation algorithm guarantees $\chi = O(n^{0.19996})$ colors [KT17]. On the hardness front, Garey et al. [GJS76] showed that it is NP-hard to color G with three colors, which was later improved to four colors [KLS00, GK04]. Assuming a variant of the Unique Games Conjecture, Dinur et al. [DMR09] proved that it is NP-hard to color with any constant number of colors.

This notion of graph coloring can be extended to hypergraphs where each edge can contain more than two vertices. There are multiple notions of hypergraph coloring that have been actively studied in mathematics and computer science, and this thesis studies three such notions of coloring.

Cut and Covering Problems. Given a graph, *cut problems* ask to remove the minimum number of edges or vertices so that the resulting graph is not well-connected. While MINIMUM CUT is one of the most famous combinatorial optimization problems that admit a polynomial time algorithm, most other natural cut problems are NP-hard and we need to rely on approximation algorithms. The following problem is a natural generalization of MINIMUM CUT.

> UNDIRECTED MULTICUT
>
> Input: A graph $G = (V, E)$ and k pairs of vertices $(s_1, t_1), \ldots, (s_k, t_k)$.
>
> Output: A subset $F \subseteq E$ of edges such that in $(V, E \setminus F)$, every s_i is disconnected from t_i.
>
> Goal: Minimize $|F|$.

This problem initiated the study of cuts and metrics and their relations to LP [LR99, LLR95, GVY96, AR98], and the best approximation ratio is $O(\log k)$ [GVY96, AR98]. It is NP-hard to approximate within a factor $1 + \epsilon$ for some $\epsilon > 0$ [DJP$^+$94], and NP-hard to approximate within any constant factor assuming the Unique Games Conjecture [CKK$^+$06].

Cut problems can be generalized in many ways. They can be generalized to *graph covering problems* where we want to choose the smallest subset of edges or vertices that satisfy certain properties. For example, another important subclass of graph covering problems is *connectivity problems* where we want to choose the smallest subset of edges or vertices such that they induce a well-connected subgraph (e.g., TRAVELING SALESMAN and STEINER TREE).

They all belong to the wide class of covering problems, where the input is a general set system (a universe U and a collection of its subsets $\mathcal{F} = \{S_1, \ldots, S_m\}$), and the goal is to find the smallest subset $T \subseteq U$ that intersects every S_i (for UNDIRECTED MULTICUT, given an instance $G = (V, E)$ and $(s_i, t_i)_{1 \leq i \leq k}$, take $U = E$ and $\mathcal{F}$ to be collection of all paths from s_i to t_i). The approximability of the most general problem in this class, MINIMUM SET COVER, is well understood [Fei98], but the techniques for this problem cannot be easily applied to the above special classes. This thesis provides general frameworks to study cut problems and graph covering problems.

1.3 This Thesis

This thesis tries to continue this line of work to prove the (nearly) optimal approximabilities for combinatorial optimization problems. Our main contributions can be classified into the following two categories: we expand the range of problems with optimal approximation ratios, and the body of techniques and perspectives used to study them.

1.3.1 Expanding the Range of Problems

As alluded to by the three example problems in the previous section, this thesis studies approximabilities of the three classes of combinatorial optimization problems: MAX CSP, coloring, and graph covering problems. Results regarding MAX CSP and related problems appear in Part I and II. Part III studies various notions of hypergraph coloring. Part IV studies a special case of graph covering problems called H-TRANSVERSAL, and Part V proves hardness of various cut problems. The final chapter discusses some future directions naturally emerging from the results of this thesis, as well as other major classes of optimization problems not covered in the thesis.

Recall that among three problems introduced (MAX CUT, COLORING 3-COLORABLE GRAPHS, UNDIRECTED MULTICUT), MAX CUT is the only problem where the best approximation algorithm and the best hardness result match. This is not a coincidence, and most of the aforementioned classic algorithms and hardness results, including the Goemans-Williamson algorithm [GW95], the PCP theorem [ALM+98, AS98], the parallel repetition theorem [Raz98], Håstad's optimal inapproximability results [Hås01], the UGC [Kho02b], and the optimality of SDP-based algorithms [Rag08] focus on MAX CSP.

While it is a wide class that contains many natural problems including MAX 3-SAT and MAX CUT, there are many problems not captured by this class.

Covering and packing problems form a fundamental class not captured by MAX CSP, and include basic graph problems such as MAXIMUM MATCHING, MAXIMUM DISJOINT PATHS, MINIMUM CUT. They also have played an important role in the development of exact combinatorial optimization algorithms. Although the most general problems in the class (e.g., MINIMUM SET COVER and MAXIMUM INDEPENDENT SET), or some simple problems (e.g., VERTEX COVER) are well understood, approximabilities of many fundamental problems in the class are not completely understood yet. Also for coloring, as the large gap for COLORING 3-COLORABLE GRAPHS indicates, there are large gaps between algorithms and hardness results.

In addition to classical combinatorial optimization problems, the success of formulating numerous computational tasks as optimization problems creates new challenges and opportunities for approximation algorithms. Such examples arise from various academic fields including information/coding theory, machine learning/data mining, and computational economics.

One of the biggest goals of this thesis is to prove the (nearly) optimal approximation factors for old and new problems beyond MAX CSP, hence the title *Optimal approximabilities beyond CSPs*. We briefly summarize our contribution categorized by the problem classes. See Chapter 2 for the formal definitions of the problems and our results on them.

Part I: CSPs. We start this thesis by revisiting MAX CSP. The result of Raghavendra [Rag08] states that assuming the Unique Games Conjecture, a simple SDP-based algorithm achieves the best approximation factor. While this seems to completely close the study of MAX CSP, we address the following two questions:

1. The original MAX CSP allows every assignment to be feasible. What happens if we only allow certain assignments to be feasible?

2. There are some special cases of MAX CSP where it is NP-hard to strictly outperform a simple algorithm that assigns an independent random value to each variable. Can we completely characterize these special cases?

Chapter 4 studies the first question. We introduce two variants of MAX CSP called BALANCE CSP and HARD CSP that capture the first question. For BALANCE CSP, an assignment becomes feasible if and only if every value in the domain is assigned to the same number of variables. For HARD CSP, the input consists of *hard constraints* and *soft* *constraints*. An assignment is feasible if and only if it satisfies all hard constraints, and our goal is to maximize the number of satisfied soft constraints. We show that these two ways of restricting the set of feasible assignments make MAX CSP significantly harder. One of the most surprising results of this chapter is the large difference between MAX 2-SAT and MAX CUT. The best approximation factor for MAX CUT and MAX 2-SAT is 0.878 [GW95] and 0.941 [LLZ02] respectively, which seems to suggest that MAX 2-SAT is *easier* than MAX CUT. However, the hard version and the balance version of MAX CUT both admit a *robust algorithm*, but neither of those for MAX 2-SAT does admit such an algorithm (HARD 2-SAT does not even admit a constant factor approximation)!

Chapter 5 studies the second question. A CSP is called *approximation resistant* if the best approximation factor (achieved by Raghavendra's SDP) is also achieved by a simple algorithm where each variable gets an independently sampled value from its domain. There is a large volume of work characterizing when a CSP is approximation resistant [AH13, AK13, KTW14], but the characterization tends to be technically complicated. We study a natural subclass of MAX CSP called SYMMETRIC CSP, and give a simpler characterization for approximation resistance.

Part II: Applied CSPs. This part studies problems that can model real-world scenarios, especially in economics and error-correcting codes. These problems can be thought as variants of CSPs, but additional technical ideas are required in order to understand their approximabilities. Chapter 6 proves nearly optimal NP-hardness of UNIQUE COVERAGE. While this problem can be thought as a special case of MAX CSP, it also has a close connection to MINIMUM SET COVER, which allows it to model situations in wireless/radio networks and pricing.

Chapter 7 settles the tight approximation ratio for GRAPH PRICING that models the following simple scenario with a single seller and multiple customers. Given a graph $G = (V, E)$, each vertex represents a type of item that the seller has, each with unlimited copies. Each edge $e = (u, v)$ corresponds to a customer that has her own budget b_e and is interested in buying one item of type u and one item of type v. The customers are *single-minded* in the sense that each customer $e = (u, v)$ buys both u and v if the sum of the prices does not exceed her budget (i.e. $b_e \geq p(u) + p(v)$, where $p(v)$ indicates the price of item v), in which the seller gets $p(u) + p(v)$ from the customer. Otherwise, the customer does not buy anything and the seller gets no profit from this customer. The goal of the seller is to set a nonnegative price to each item to maximize her profit from m customers. There is a very simple $(1/4)$-approximation algorithm [BB07, LBA$^+$07], and there have been significant efforts to improve this algorithm or prove its optimality [KKMS09, KMR11, CKLN13]. We finally prove the optimality of this simple approximation algorithm, by introducing an intermediate problem called GENERALIZED MAX DICUT. While GRAPH PRICING cannot be formally captured by combinatorial optimization since each vertex can be assigned a positive real number, this intermediate problem bridges the current tools for combinatorial optimization problems and GRAPH PRICING.

Chapter 8 studies the performance of an algorithm that decodes LDPC codes by formulating it as a convex optimization problem. Previous results [Fel03, FWK05, FMS$^+$07] showed that formulating it as a LP achieves some nontrivial guarantee, and also proved its limitations. We prove that even a much stronger *Sum-of-Squares hierarchy* formulation does not significantly improve the performance.

Part III: Coloring. Graph coloring requires that for each edge (u, v), the color of u must be different from the color of v. If we consider hypergraphs where each edge can have more than two vertices, there are multiple ways to generalize this notion. One of the most popular notions in computer science is *weak coloring*, where each hyperedge has to have at least two different colors (i.e., it is acceptable to have all but one vertex have the same color). Under this notion, strong hardness results were proved [GHS02, DRS05, DG13, GHH$^+$14, KS14a, Hua15].

How can we strengthen this notion of weak coloring? In a k-uniform hypergraph and a valid 2-weak coloring that colors each vertex with blue or red, each hyperedge can have 1 blue color and $k - 1$ red colors. What if we require that the number of blue vertices and the number of red vertices are roughly the same in each hyperedge? Will it make the task of weak coloring easier?

This notion is called *low-discrepancy coloring*. Formally, a 2-coloring of a hypergraph is called a discrepancy-ℓ coloring if the number of blue vertices and the number of red vertices differ by at most ℓ in each hyperedge. We consider two more notions called *rainbow coloring* and *strong coloring*. Informally, rainbow coloring requires that every hyperedge has to have all possible colors at least once (so the number of colors is at most the minimum cardinality of a hyperedge), and strong coloring requires that in every hyperedge, no color appears more than once (so the number of colors is at least the maximum cardinality of a hyperedge). All three notions imply weak-colorability. These problems naturally capture a wider class of combinatorial optimization problems (e.g., scheduling), and exhibit a richer connection to discrepancy theory.

In Part III, we study the approximability of weak coloring where the input hypergraph is promised to admit one of these three strong notions of coloring. We present algorithms that exploit such a strong structure, and also prove that weak coloring is still computationally hard even under these strong promises. It was known that for hypergraphs that admit a discrepancy-0 coloring, there are simple random walk and SDP-based algorithms that efficiently find a valid 2-weak coloring. One of our main results proves that even when a hypergraph admits a discrepancy-2 coloring, it is NP-hard to find a weak c-coloring for any constant $c > 0$.

Part IV: Subgraph Transversal and Graph Partitioning. Recall that a graph covering problem refers to a problem where the input is a graph $G = (V, E)$ and the goal is to choose the minimum number of vertices or edges that satisfy some properties. In this part, we are interested in the property that G does not include a specific subgraph H after removing the chosen vertices or edges. There are numerous theorems from extremal graph theory that estimate the size of the optimal solution for fixed G and H, starting from the famous Turán's theorem that studies how many edges we need to remove from an n-clique in order to exclude a k-clique.

Let H-TRANSVERSAL be the optimization problem where given a graph G and a fixed "pattern" graph H, the goal is to remove the minimum number of vertices from G so that it does not include H as a subgraph.[1] Besides its connection to extremal graph theory, this problem captures many fundamental combinatorial optimization problems as special cases.

- When H is a single edge, H-TRANSVERSAL becomes the famous VERTEX COVER.
- When H is a k-Star, a tree with k vertices where each of $k - 1$ leaves has an edge between the root, H-TRANSVERSAL becomes a natural degree reduction problem where we want to remove the minimum number of vertices so that the maximum degree becomes strictly less than $k - 1$. If G is a $(k - 1)$-regular graph, it becomes MINIMUM DOMINATING SET.
- When H is a simple path with k vertices, H-TRANSVERSAL has been known as k-PATH VERTEX COVER or P_k-HITTING SET that has been actively studied for small values of $k = 3, 4$.
- When H is a family of all cycles instead of a single graph, H-TRANSVERSAL becomes FEEDBACK VERTEX SET.

We try to characterize the approximability of H-TRANSVERSAL depending on H. Whenever H has k vertices, there is a simple k-approximation algorithm. We show that for any 2-vertex connected H, this algorithm is likely to be optimal, and complement this hardness

[1] We also study the packing version in Part IV and the edge deletion version in the full version [GL15b].

by presenting $O(\log k)$-approximation algorithms for two 1-connected graphs, k-STAR TRANSVERSAL and k-PATH TRANSVERSAL. Our algorithm for k-PATH TRANSVERSAL uses the algorithm for k-VERTEX SEPARATOR as a subroutine, where given a graph G, the goal is to remove the minimum number of vertices so that each connected component has at most k vertices.

Part V: Cut Problems. Since the Max-Flow Min-Cut theorem, cut problems have played an important role in the development of combinatorial optimization. One of the most important consequences of this theorem is a polynomial time exact algorithm for s-t MIN CUT, where given a directed graph G and two vertices s and t, the goal is to remove the minimum number of edges to ensure that there is no path from s to k. Consider a natural generalization of this problem where we are given four vertices s_1, t_1, s_2, t_2, and we want to remove the minimum number of edges to ensure that there is no path from s_1 to t_1 and s_2 to t_2. A simple 2-approximation algorithm exists because we can separately compute the minimum s_1-t_1 cut and the minimum s_2-t_2 cut and take their union. Is this simple algorithm optimal?

In Part V, we also study various cut problems on graphs, where the goal is to remove the minimum number of vertices or edges to cut *desired* paths, and prove strong hardness results. The problems we study include multicut, bicut, doublecut, length-bounded cut, interdiction, and firefighter problems. In particular, our results imply that the simple 2-approximation above is optimal under the Unique Games Conjecture. Approximation algorithms for cut problems, as well as the exact algorithm for s-t MIN CUT, are closely related to their LP relaxations and metrics formed by the LP solutions. We introduce a *length-control dictatorship test* that can prove hardness of such problems in a unified manner based on their LP gap instances.

1.3.2 Methods and Tools

The results in this thesis are the combination of algorithms, hardness results, and performances of LP/SDP relaxations. Chapter 3 formally introduces mathematical and algorithmic tools commonly used in this thesis, with simple illustrations showing how they are used to design approximation algorithms or prove hardness of approximation results. More advanced tools required for specific parts of the thesis (e.g., advanced discrete Fourier analysis used to prove NP-hardness of approximation in Part III) will appear in respective parts.

Three Pillars. The relationships between the three pillars of approximation algorithms — algorithms, hardness, and convex relaxations — have been further solidified in recent years. Table 1.1 briefly summarizes our main results in this thesis in terms of which angle we took to study each problem. Our results present even stronger pairwise relationships between the three pillars. Some representative connections include:

- Algorithms and convex relaxations: All our algorithms are based on either LP or SDP relaxations. See Section 3.4 for the formal introduction of LP/SDP relaxations and the complete list of our algorithms.
- Convex relaxations and hardness: Integrality gaps, which show the limitation of a certain convex relaxation for an optimization problem, can be interpreted as hardness results under a restricted model of computation. For cut problems, we show that the insights from these integrality gaps can be used to prove NP-hardness of approximation under the Unique Games Conjecture. The results of this type were known for MAX CSP [Rag08] and other variants of CSPs [KMTV11], but not for more structured problems like cut problems.
- Hardness and algorithms: Several problems studied in this thesis *could have gone either way*, which means that there was little consensus on whether each problem admits a good approximation algorithm or it is hard to approximate. Some of the results in this thesis, including hardness of GRAPH PRICING and algorithms for k-PATH TRANSVERSAL, were achieved by the insights obtained in an effort to prove the result from the other side.

In the following, we briefly introduce some common features of our methods and tools used in this thesis: building general frameworks applicable to many problems, and introducing new conceptual viewpoints that bridge seemingly different problems or ideas.

Providing General Frameworks. We extend the techniques developed mainly for MAX CSP and VERTEX COVER. While the current techniques are generally applicable for MAX CSP, many other classes of problems lack such a general framework. While it is impossible to create tools applicable to every combinatorial optimization problem, another goal of this thesis is to provide a toolkit for wide classes of problems. We believe that the tools presented in this thesis will be useful to study the associated classes beyond the problems we study.

This task is based on unifying the techniques of previous work. Even though most of them have implicitly shared a large portion of common ideas and technical work, different characteristics of problems that require specialized ideas make it hard to unify them

	LP Algorithms	SDP Algorithms	Hierarchies	UG Hardness	NP Hardness
HARD CSP/ BALANCE CSP	O			O	O
SYMMETRIC CSP		O			
UNIQUE COVERAGE					O
GRAPH PRICING			O	O	
LDPC Decoding			O		
Coloring Problems		O			O
H-TRANSVERSAL / Graph Partitioning	O				O
Cut Problems				O	

Table 1.1: Summary of our main results according to the *pillars* they belong to. UG-hardness denotes NP-hardness assuming the Unique Games Conjecture.

in a common framework. For each class of problems we study, we present a powerful framework that simultaneously captures those specialized ideas and is applicable for many problems in the class. Unifying the existing tools often led to creating new techniques by reinterpreting them or combining ideas from two independently studied techniques.

1. Part III: For hypergraph coloring, we present a general *recipe* to prove strong hardness under the promise that hypergraphs not only admit a coloring with few colors but also have additional structure. This encompasses numerous previous results on hypergraph coloring in one framework, and yields new results.

2. Part V: Our results for various cut problems are based on the common framework called *length-control dictatorship tests* presented in Part V. They are inspired by the earlier results by Bansal and Khot [BK10] and Svensson [Sve13]. We give a new interpretation of their ideas, which allows us to use the intuition from known LP gaps for cut problems to show our results.

Introducing New Conceptual Viewpoints. Another paradigm frequently used in this thesis is introducing a slightly different conceptual viewpoint for each optimization problem or proof technique. To understand an optimization problem, we often establish a connection to a seemingly different optimization problem that provides both conceptual and technical tools to study the original problem from a new angle. We also present reinterpretations of some existing techniques in a new framework that make them more widely applicable.

1. Chapter 6: UNIQUE COVERAGE is an interesting problem in the sense that it is a special case of MAX CSP by definition, but it exhibits properties closer to MINIMUM SET COVER. We prove nearly optimal NP-hardness of UNIQUE COVERAGE by combining ideas from hardness results for MAX CSP and MINIMUM SET COVER that have been developed separately. More specifically, our proof reinterprets Feige's MINIMUM SET COVER hardness result [Fei98] in a probabilistic proof checking framework mainly used for MAX CSP.

2. Chapter 7: Our hardness for GRAPH PRICING is proved via an intermediate problem called GENERALIZED MAX DICUT. While the direct formulation of GRAPH PRICING as a special case of MAX CSP results in complicated constraints, GENERALIZED MAX DICUT provides the right abstraction of GRAPH PRICING as a MAX CSP that allows us to seamlessly use the tools developed for MAX CSP.

3. Chapter 15: Our algorithm for k-PATH TRANSVERSAL uses an algorithm for a graph partitioning problem called k-VERTEX SEPARATOR. We prove that any connected graph without a long path has a small subset of vertices whose deletion partitions the graph into smaller components. This connection allows us to apply techniques for graph partitioning algorithms for k-PATH TRANSVERSAL. Our approximation algorithm runs in fixed-parameter tractable (FPT) time, and uses tools from both approximation algorithms and FPT algorithms.

1.4 Previous Versions and Credits

Most of the results in this thesis have previously appeared in different forms, and many of them are done with different coauthors.

- In Part I, Chapter 4 for HARD CSP, BALANCE CSP is based on the work with Venkat Guruswami [GL14] and its journal version [GL16a]. Chapter 5 for SYMMETRIC CSP is also based on another work with Venkat Guruswami [GL15a].

- In Part II, Chapter 6 for UNIQUE COVERAGE is based on the work with Venkat Guruswami [GL16b]. Chapter 7 GRAPH PRICING is based on [Lee15]. Chapter 8 for decoding LDPC codes is based on the work with Badih Ghazi [GL15a] and its journal version [GL17].

- Part III is based on the following two papers: The work with Venkat Guruswami [GL15d] proved hardness of MIN COLORING, and the work with Vijay Bhattiprolu and Venkat Guruswami proved hardness of MAX 2-COLORING and all coloring algorithms [BGL15].

- Part IV is based on the following two papers: The work with Venkat Guruswami [GL15c] proved hardness of H-TRANSVERSAL and studied k-STAR TRANSVERSAL. [Lee17b] gave algorithms for k-VERTEX SEPARATOR and k-PATH TRANSVERSAL.

- Part V is based on the following two papers: [Lee17a] introduced the length control dictatorship tests to prove hardness of multicut, length-bounded cut, interdiction, and firefighter problems, and the work with Kristóf Bérczi, Karthekeyan Chandrasekaran, Tamás Király, and Chao Xu [BCK$^+$17] applied it to bicut and double cut problems.

Chapter 2

Problems and Summary of Results

This chapter formally introduces the problems studied in this thesis and states our results. Section 2.1, 2.2, 2.3, 2.4, and 2.5 correspond to Part I, II, III, IV, and V respectively.

2.1 Constraint Satisfaction Problems

Constraint Satisfaction Problems (CSPs) are among the most fundamental and well-studied classes of optimization problems. For a fixed domain $\mathcal{D}$, a CSP is specified by a finite set $\Pi = \{P_1, \dots, P_l\}$ of relations, where each relation P_i is a subset of $\mathcal{D}^{k_i}$ for some $k_i \in \mathbb{N}$. Given such Π, MAX CSP(Π) is defined as follows.

MAX CSP(Π)

Input: A set of variables $X = \{x_1, ..., x_n\}$ and a collection of constraints $\mathcal{C} = \{C_1, ..., C_m\}$. Each constraint C_i is an expression of the form $R(x_{i_1}, \dots, x_{i_k})$ where R is a relation of arity k contained in Π, and x_{i_j}'s are variables.

Output: An assignment $\sigma : X \to \mathcal{D}$.

Goal: Maximize the number of satisfied constraints. A constraint $R(x_{i_1}, \dots, x_{i_k})$ is satisfied when $(\sigma(x_{i_1}), \dots, \sigma(x_{i_k})) \in R$.

2.1.1 Hard / Balanced CSPs

We consider two natural extensions of MAX CSP(Π).

BALANCE CSP(Π)

Input: A set of variables $X = \{x_1, ..., x_n\}$ and a collection of constraints $\mathcal{C} = \{C_1, ..., C_m\}$.

Output: A *balanced* assignment $\sigma : X \to \mathcal{D}$. An assignment is called balanced if for each $q \in \mathcal{D}$, $|\sigma^{-1}(q)| = \frac{n}{|\mathcal{D}|}$.

Goal: Maximize the number of satisfied constraints.

Partitioning a set of objects into equal-sized subsets with desired properties is a basic scheme used in Divide-and-Conquer algorithms. BALANCE CUT, also known as MAXIMUM BISECTION, is one of the most well-known examples of BALANCE CSP. The balance constraint is also one of the simplest *non-local* constraints where the current algorithmic and hardness results on ordinary CSPs do not work.

HARD CSP(Π)

Input: A set of variables $X = \{x_1, ..., x_n\}$, a collection *soft* constraints $\mathcal{S} = \{C_1, ..., C_{m_s}\}$, and a collection *hard* constraints $\mathcal{H} = \{H_1, ..., H_{m_h}\}$.

Output: An assignment $\sigma : X \to \mathcal{D}$ that satisfies every hard constraint.

Goal: Maximize the number of satisfied soft constraints.

HARD CSP contains every MAX CSP by definition, and also several additional fundamental combinatorial optimization problems, such as (HYPERGRAPH) INDEPENDENT SET, MULTICUT, GRAPH k-COLORING, and many other covering/packing problems. While every assignment is *feasible* in ordinary MAX CSP, in HARD CSP only certain assignments that satisfy all the hard constraints are considered as feasible, giving a more general framework to study combinatorial optimization problems.

By the seminal work of Schaefer [Sch78], there are only three nontrivial classes of Boolean CSPs for which satisfiability can be checked in polynomial time: 2-SAT, HORN-SAT, and LIN-MOD-2. Among them, only MAX 2-SAT and MAX HORN-SAT admit a *robust algorithm*, which outputs an assignment satisfying at least $(1 - g(\epsilon))$ fraction of constraints given a $(1 - \epsilon)$-satisfiable instance, where $g(\epsilon) \to 0$ as $\epsilon \to 0$, and $g(0) = 0$. We study how balance and hard constraints affect approximabilities of these two problems. More specifically, we ask whether each variant admits a robust algorithm or a constant factor approximation algorithm and obtain the following results. Given an assignment

$\sigma : X \to \mathcal{D}$, let $\mathsf{Val}(\sigma)$ be the fraction of constraints satisfied by σ, and let Opt indicate the fraction of constraints satisfied by an optimal assignment (satisfied soft constraints for HARD CSP).

Theorem 2.1.1. *There exists an absolute constant $\delta > 0$ such that given an instance $\mathcal{I}$ of* BALANCE HORN 2-SAT *(special case of* BALANCE 2-SAT *and* BALANCE HORN-SAT*), it is NP-hard to distinguish between the following cases.*

- $\mathsf{Opt} = 1$
- $\mathsf{Opt} \leq 1 - \delta$

Theorem 2.1.2. *For any $\epsilon > 0$, there is a randomized algorithm such that given an instace $\mathcal{I}$ of* BALANCE SAT*, in time $\mathrm{poly}(size(\mathcal{I}), \frac{1}{\epsilon})$, outputs σ with $\mathsf{Val}(\sigma) \geq (\frac{3}{4} - \epsilon)\mathsf{Opt}(\mathcal{I})$ with constant probability.*

Theorem 2.1.3. *For any $\epsilon > 0$, given an instance $\mathcal{I}$ of* HARD 2-SAT*, it is UG-hard to distinguish the following cases.*

- $\mathsf{Opt} \geq 1 - \epsilon$
- $\mathsf{Opt} \leq \epsilon$

Theorem 2.1.4. *For any $\epsilon > 0$, given an instance $\mathcal{I}$ of* HARD HORN 3-SAT*, it is NP-hard to distinguish between the following cases.*

- $\mathsf{Opt} \geq 1 - \epsilon$
- $\mathsf{Opt} \leq \epsilon$

The above theorems imply that for both MAX 2-SAT and MAX HORN-SAT, balance constraints rule out robust algorithms but still allow constant-factor approximation algorithms, while hard constraints rule out both robust algorithms and constant-factor approximation algorithms. Table 2.1 summarizes our results.

2.1.2 Symmetric CSPs

In this section we assume that the domain is Boolean (i.e., $\mathcal{D} = \{0, 1\}$). Recent works on approximability of CSPs focus on characterizing every CSP according to its approximation resistance. We define *random assignments* to be the class of algorithms that assign $x_i \leftarrow 1$ with probability α independently for some $\alpha \in [0, 1]$. A CSP is called *approximation*

		Ordinary	Balance	Hard
MAX 2-SAT	Ratio	0.9401	0.9401	**N/A**
	Robust	$(1-\epsilon, 1-O(\sqrt{\epsilon}))$	**N/A**	**N/A**
	Hardness	UG: $(1-\epsilon, 1-\Omega(\sqrt{\epsilon}))$	**NP:** $(1, 1-\delta)$	**UG:**$(1-\epsilon, \epsilon)$
MAX HORN-SAT	Ratio	0.7968	**0.75**	**N/A**
	Robust	$(1-\epsilon, 1-O(\frac{\log\log(1/\epsilon)}{\log(1/\epsilon)}))$	**N/A**	**N/A**
	Hardness	UG: $(1-\epsilon, 1-\Omega(\frac{1}{\log(1/\epsilon)}))$	**NP:** $(1, 1-\delta)$	**NP:**$(1-\epsilon, \epsilon)$

Table 2.1: Summary of our results. $\epsilon > 0$ indicates an arbitrary positive constant, while $\delta > 0$ is a fixed absolute constant. Ratio indicates the best approximation ratio. $(1-\epsilon, 1-f(\epsilon))$ in a row labeled Robust indicates that there is a robust algorithm that find an assignment satisfying $(1-f(\epsilon))$ fraction of constraints given an $(1-\epsilon)$-satisfiable instance. $(1-\epsilon, 1-f(\epsilon))$ in a row labeled Hardness indicates that it is NP-hard to find an assignment satisfying $(1-f(\epsilon))$ fraction of constraints given an $(1-\epsilon)$-satisfiable instance. NP indicates that it is an NP-hardness result; UG indicates that it is based on the Unique Games Conjecture. N/A means that a robust or constant-factor approximation algorithm are ruled out by the hardness results. The results of this thesis are in boldface.

resistant, if for any $\epsilon > 0$, it is NP-hard to have a $(\rho^* + \epsilon)$-approximation algorithm, where ρ^* is the approximation ratio achieved by the best random assignment. Even assuming the UGC, the complete characterization of approximation resistance has not been found, and previous works either change the notion of approximation resistance or study a subclass of CSPs to find a characterization, and more general results tend to suggest more complex characterizations.

We study a natural subclass of CSPs where the domain is Boolean and the constraint language Π has one predicate Q which is *symmetric* — for any permutation $\pi : [k] \to [k]$, $(x_1, \ldots, x_k) \in Q$ if and only if $(x_{\pi(1)}, \ldots, x_{\pi(k)}) \in Q$. Equivalently, for every such Q, there exists $S \subseteq [k] \cup \{0\}$ such that $(x_1, \ldots, x_k) \in Q$ if and only if $(x_1 + \cdots + x_k) \in S$. Let SYMMETRIC CSP($S$) WITHOUT NEGATION denote such a symmetric CSP. We also study SYMMETRIC CSP(S) WITH NEGATION where each constraint C_j is speficied by a tuple $(x_{j_1}, \ldots, x_{j,k})$ as well as $b_{j,1}, \ldots, b_{j,k}$ and satisfied if $((x_{j,1} \oplus b_{j,1}) + \cdots + (x_{j,k} \oplus b_{j,k})) \in S$ where $\oplus$ denotes the addition in $\mathbb{F}_2$ and $+$ denotes the addition in $\mathbb{Z}$.

While this is a significant restriction, it is a natural one that still captures the following fundamental problems, such as MAX SAT, MAX NOT-ALL-EQUAL-SAT, MAX t-OUT-OF-k-SAT (with negation), and MAX CUT, MAX-SET-SPLITTING, DISCREPANCY MINIMIZATION (without negation).

There is a simple sufficient condition to be approximation resistant due to Austrin and

Mossel [AM09] with negation, and due to Austrin and Håstad [AH13] without negation. For $s \in [k] \cup \{0\}$, let $P(s) \in \mathbb{R}^2$ be the point defined by $P(s) := (\frac{s}{k}, \frac{s(s-1)}{k(k-1)})$. For any s, $P(s)$ lies on the curve $y = \frac{k}{k-1}x^2 - \frac{x}{k-1}$, which is slightly below the curve $y = x^2$ for $x \in [0,1]$. Given a subset $S \subseteq [k] \cup \{0\}$, let $P_S := \{P(s) : s \in S\}$ and $\mathsf{conv}(P_S)$ be the convex hull of P_S. For symmetric CSPs, the conditions of [AH13] and [AM09] depend on whether this convex hull intersects a certain curve or a point.

For SYMMETRIC CSP(S) WITHOUT NEGATION, the condition simply becomes whether $\mathsf{conv}(P_S)$ intersects the curve $y = x^2$. If we let s_{min} and s_{max} be the minimum and maximum number in S respectively, by convexity of $y = \frac{k}{k-1}x^2 - \frac{x}{k-1}$, it is equivalent to that the line passing through $P(s_{min})$ and $P(s_{max})$ and $y = x^2$ intersect, which is again equivalent to

$$\frac{(s_{max} + s_{min} - 1)^2}{k-1} \geq \frac{4s_{max}s_{min}}{k}. \tag{2.1}$$

For SYMMETRIC CSP(S) WITH NEGATION, the condition of Austrin and Mossel [AM09] is simplified to that $\mathsf{conv}(P_S)$ contains the point $(\frac{1}{2}, \frac{1}{4})$. We suggest that these simplified sufficient conditions might also be necessary and thus precisely characterize approximation resistance. We prove it for two natural special cases (which capture all problems mentioned in the last paragraph) for both symmetric CSPs with/without negation, and provide reasons that we believe this is true at least for symmetric CSPs without negation.

Conjecture 2.1.5. *For $S \subseteq [k-1]$,* SYMMETRIC CSP*(S)* WITHOUT NEGATION *is approximation resistant if and only if* (2.1) *holds.*

Theorem 2.1.6. *If $S \subseteq [k-1]$ and S is either an interval or even,* SYMMETRIC CSP*(S)* WITHOUT NEGATION *is approximation resistant if and only if* (2.1) *holds (the hardness claim, i.e., the "if" part, is under the Unique Games Conjecture).*

Theorem 2.1.7. *If $S \subset [k] \cup \{0\}$ and S is either an interval or even,* SYMMETRIC CSP*(S)* WITH NEGATION *is approximation resistant if and only if* $\mathsf{conv}(P_S)$ *contains* $(\frac{1}{2}, \frac{1}{4})$ *(the hardness claim, i.e., the "if" part, is under the Unique Games Conjecture).*

2.2 Applied CSPs

Part II studies *applied CSPs*. These problems are motivated by the intersection of combinatorial optimization and other fields such as economics and error-correcting codes. They can be thought as a variant of CSPs, but also require new insights to study their approximabilities.

2.2.1 UNIQUE COVERAGE

We study the following natural problem that models numerous practical situations arising from wireless networks, radio broadcast, and envy-free pricing.

> UNIQUE COVERAGE
>
> Input: A universe V of n elements and a collection E of m subsets of V.
>
> Output: $S \subseteq V$.
>
> Goal: Maximize the number of $e \in E$ that intersects S in exactly one element.

When each $e \in E$ has size at most k, this problem is also known as 1-IN-k HITTING SET. While this problem can be captured as a MAX CSP, this problem differs from other famous MAX CSP in the sense that arities of constraints (sizes of $e \in E$) can be different and grow with n so that the traditional results for MAX CSP are not applicable. It admits a simple $\Omega(\frac{1}{\log k})$-approximation algorithm.

For constant k, we prove that 1-IN-k HITTING SET is NP-hard to approximate within a factor $O(\frac{1}{\log k})$. This improves the result of Guruswami and Zhou [GZ12], who proved the same result assuming the Unique Games Conjecture.

Theorem 2.2.1. *Assuming* $\mathbf{P} \neq \mathbf{NP}$*, for large enough constant* k*, there is no polynomial time algorithm that approximates* 1*-in-*k *HS within a factor better than* $O(\frac{1}{\log k})$.

For UNIQUE COVERAGE, we prove that it is hard to approximate within a factor $O(\frac{1}{\log^{1-\epsilon} n})$ for any $\epsilon > 0$, unless NP admits quasipolynomial time algorithms. This improves the results of Demaine et al. [DFHS08], including their $\approx 1/\log^{1/3} n$ inapproximability factor which was proven under the Random 3SAT Hypothesis.

Theorem 2.2.2. *Assuming* $\mathbf{NP} \not\subseteq \mathbf{QP}$*, for any* $\epsilon > 0$*, there is no polynomial time algorithm that approximates* UNIQUE COVERAGE *within a factor better than* $\frac{1}{\log^{1-\epsilon} n}$.

Our simple proof combines ideas from two classical inapproximability results for MINIMUM SET COVER and MAX CSP, made efficient by various derandomization methods based on bounded independence.

2.2.2 Decoding LDPC Codes

Low-density parity-check (LDPC) codes are a class of linear error correcting codes originally introduced by Gallager [Gal62] and that have been extensively studied in the last decades. A (d_v, d_c)-LDPC code of block length n is described by a parity-check matrix $H \in \mathbb{F}_2^{m \times n}$ (with $m \leq n$) having d_v ones in each column and d_c ones in each row. In many studies of LDPC codes, *random LDPC codes* have been considered. For instance, Gallager studied in his thesis the distance and decoding-error probability of an ensemble of random (d_v, d_c)-LDPC codes. Random (d_v, d_c)-LDPC codes were further studied in several works (e.g., [SS94, Mac99, RU01, MB01, DPT$^+$02, LS02, KRU12]). The reasons why random (d_v, d_c)-LDPC codes have been of significant interest are their nice properties, their tendency to simplify the analysis of the decoding algorithms and the potential lack of known explicit constructions for properties satisfied by random codes.

Sipser and Spielman [SS94] gave a linear-time decoding algorithm correcting a constant fraction of errors (for $d_v, d_c = O(1)$). More precisely, the linear-time decoding algorithm of Sipser-Spileman corrects $\Omega(1/d_c)$-errors on a random (d_v, d_c)-LDPC code. A few years after, Feldman, Karger and Wainwright [FWK05, Fel03] introduced a decoding algorithm that is based on a simple linear programming (LP) relaxation that corrects $\Omega(1/d_c)$-errors on a random (d_v, d_c)-LDPC code.

However, the fraction of errors that is corrected by the Sipser-Spielman algorithm and the LP relaxation of [FWK05] (which is $O(1/d_c)$) can be much smaller than the best possible: in fact, [Gal62] (as well as [MB01]) showed that for a random (d_v, d_c)-LDPC code, the exponential-time nearest-neighbor Maximum Likelihood (ML) algorithm corrects close to $H_b^{-1}(d_v/d_c)$ probabilistic errors, which by Shannon's channel coding theorem is the best possible.

Inspired by the Sherali-Adams hierarchy, Arora, Daskalakis and Steurer [ADS12] improved the best known fraction of correctable probabilistic errors by the LP decoder (which was previously achieved by Daskalakis et al. [DDKW08]) for some range of values of d_v and d_c. Both Arora et al. [ADS12] and the original work of Feldman et al. [FWK05, Fel03] asked whether tightening the base LP using linear or semidefinite hierarchies can improve its performance, potentially approaching the information-theoretic limit. More precisely, in all previous work on LP decoding of error-correcting codes, the base LP decoder of Feldman et al. succeeds in the decoding task if and only if the transmitted codeword is the unique optimum of the relaxed polytope with the objective function being the (normalized) l_1 distance between the received vector and a point in the polytope. On the other hand, the decoder is considered to fail whenever there is an optimal non-integral vector.

In this paper, we prove the first lower bounds on the performance of the Sherali-Adams and Sum-of-Squares hierarchies when applied to the problem of decoding random (d_v, d_c)-LDPC codes.

Theorem 2.2.3 (Lower bounds in the Sherali-Adams hierarchy)**.** *For any d_v and $d_c \geq 5$, there exists $\eta > 0$ (depending on d_c) such that a random (d_v, d_c)-LDPC code satisfies the following with high probability: for any received vector, there is a fractional solution to the ηn rounds of the Sherali-Adams hierarchy of value $1/(d_c - 3)$ (for odd d_c) or $1/(d_c - 4)$ (for even d_c). Consequently, ηn rounds cannot decode more than $a \approx 1/d_c$ fraction of errors.*

Theorem 2.2.4 (Lower bounds in the Sum-of-Squares hierarchy)**.** *For any d_v and $d_c = 3 \cdot 2^i + 3$ with $i \geq 1$, there exists $\eta > 0$ (depending on d_c) such that a random (d_v, d_c)-LDPC code satisfies the following with high probability: for any received vector, there is a fractional solution to the ηn rounds of the Sum-of-Squares hierarchy of value $3/(d_c - 3)$. Consequently, ηn rounds cannot decode more than $a \approx 3/d_c$ fraction of errors.*

2.2.3 GRAPH PRICING

We consider the following natural problem for a seller with a profit-maximization objective. Let $\mathbb{I}$ denote the indicator function.

> GRAPH PRICING
>
> Input: A graph $G = (V, E)$. For each edge e, its *budget* $b_e \in \mathbb{R}^+$.
>
> Output: Pricing $p : V \to (\mathbb{R}^+ \cup \{0\})$.
>
> Goal: Maximize $\sum_{(u,v)\in E} \mathbb{I}[p(u) + p(v) \leq b_{(u,v)}](p(u) + p(v))$.

This problem was proposed by Guruswami et al. [GHK$^+$05], and has received much attention. The best known approximation algorithm for a general instance, which guarantees $\frac{1}{4}$ of the optimal solution, is given by Balcan and Blum [BB07] and Lee et al. [LBA$^+$07]. The algorithm is simple enough to state here. First, assign 0 to each vertex with probability half independently. For each remaining vertex v, assign the price which maximizes the profit between v and its neighbors already assigned 0. This simple algorithm has been neither improved nor proved to be optimal. GRAPH PRICING is APX-hard [GHK$^+$05], but the only strong hardness of approximation result rules out an approximation algorithm with a guarantee better than $\frac{1}{2}$ [KKMS09] under the Unique Games Conjecture (via reduction from MAXIMUM ACYCLIC SUBGRAPH).

The $\frac{1}{4}$-approximation algorithm is surprisingly simple and does not even rely on the power of a linear programming or semidefinite programming relaxation. The efforts to exploit the power of LP relaxations to find a better approximation algorithm have produced positive results for special classes of graphs. Krauthgamer et al. [KMR11] studied the case where all budgets are the same (but the graph might have a self-loop), and proposed a $\frac{5+\sqrt{2}}{6+\sqrt{2}} \approx 0.86$-approximation algorithm based on a LP relaxation. In general case, the standard LP is shown to have an integrality gap close to $\frac{1}{4}$ [KKMS09]. Therefore, it is natural to consider hierarchies of LP relaxations such as the Sherali-Adams hierarchy [SA90] (see [CT12] for a general survey and [GTW13, YZ14] for recent algorithmic results using the Sherali-Adams hierarchy). Especially, Chalermsook et al. [CKLN13] recently showed that there is a FPTAS when the graph has bounded treewidth, based on the Sherali-Adams hierarchy. However, the power of the Sherali-Adams hierarchy and SDP, as well as the inherent hardness of the problem, was not well-understood in general case.

We show that any polynomial time algorithm that guarantees a ratio better than $\frac{1}{4}$ must be powerful enough to refute the Unique Games Conjecture.

Theorem 2.2.5. *Under the Unique Games Conjecture, for any $\epsilon > 0$, it is NP-hard to approximate* GRAPH PRICING *within a factor of $\frac{1}{4} + \epsilon$.*

By the results of Khot and Vishnoi [KV05] and Raghavendra and Steurer [RS09] that convert a hardness under the UGC to a SDP gap instance, our result unconditionally shows that even a SDP-based algorithm will not improve the performance of a simple algorithm. For the Sherali-Adams hierarchy, we prove that even polynomial rounds of the Sherali-Adams hierarchy has an integrality gap close to $\frac{1}{4}$.

Theorem 2.2.6. *Fix $\epsilon > 0$. There exists $\delta > 0$ such that the integrality gap of n^{δ}-rounds of the Sherali-Adams hierarchy for* GRAPH PRICING *is at most $\frac{1}{4} + \epsilon$.*

2.3 Hypergraph Coloring

Coloring (hyper)graphs is one of the most important and well-studied tasks in discrete mathematics and theoretical computer science. A K-uniform hypergraph $G = (V, E)$ is said to be χ-colorable if there exists a coloring $c : V \to \{1, \ldots, \chi\}$ such that no hyperedge is monochromatic, and such a coloring c is referred to as a proper χ-coloring. Coloring has been the focus of active research in both fields, and has served as the benchmark for new research paradigms such as the probabilistic method (Lovász local lemma [EL75]) and semidefinite programming (Lovász theta function [Lov79]).

Given a general χ-colorable K-uniform hypergraph, the problem of reconstructing a χ-coloring is known to be a hard task. Even assuming 2-colorability, reconstructing a proper 2-coloring is a classic NP-hard problem for $K \geq 3$. Given the intractability of proper 2-coloring, two notions of *approximate coloring* of 2-colorable hypergraphs have been studied in the literature of approximation algorithms. The first notion, called MIN COLORING, is to minimize the number of colors while still requiring that every hyperedge be non-monochromatic. The second notion, called MAX 2-COLORING allows only 2 colors, but the objective is to maximize the number of non-monochromatic hyperedges.[1]

Even with these relaxed objectives, the promise that the input hypergraph is 2-colorable seems grossly inadequate for polynomial time algorithms to exploit in a significant way. For MIN COLORING, given a 2-colorable K-uniform hypergraph, the best known algorithm uses $O(n^{1-\frac{1}{K}})$ colors [CF96, AKMH96], which tends to the trivial upper bound n as K increases. On the other hand, [KS14b] shows quasi-NP-hardness of $2^{(\log n)^{\Omega(1)}}$-coloring a 2-colorable hypergraph (very recently the exponent was shown to approach $1/10$ in [Hua15]).

The hardness results for MAX 2-COLORING show an even more pessimistic picture, wherein the naive *random assignment* (randomly give one of two colors to each vertex independently to leave a $(\frac{1}{2})^{K-1}$ fraction of hyperedges monochromatic in expectation), is shown to have the best guarantee for a polynomial time algorithm when $K \geq 4$ (see [Hås01]).

Given these strong intractability results, it is natural to consider what further relaxations of the objectives could lead to efficient algorithms. This motivates our main question *"how strong a promise on the input hypergraph is required for polynomial time algorithms to perform significantly better than naive algorithms for* MIN COLORING *and* MAX 2-COLORING*?"*

There is a very strong promise on K-uniform hypergraphs which makes the task of proper 2-coloring easy. If a hypergraph is K-partite (i.e., there is a K-coloring such that each hyperedge has each color exactly once), then one can properly 2-color the hypergraph in polynomial time [Alo14, McD93]. The same algorithm can be generalized to hypergraphs which admit a c-balanced coloring (i.e., c divides K and there is a K-coloring such that each hyperedge has each color exactly $\frac{K}{c}$ times).

The promises on structured colorings that we consider in this thesis are natural relaxations of the above strong promise of a perfectly balanced coloring.

[1] The maximization version is also known as MAX-SET-SPLITTING, or more specifically MAX K-SET-SPLITTING when considering K-uniform hypegraphs, in the literature.

- A hypergraph is said to have *discrepancy* ℓ when there is a 2-coloring such that in each hyperedge, the difference between the number of vertices of each color is at most ℓ.
- A χ-coloring $(\chi \leq K)$ is called *rainbow* if every hyperedge contains each color at least once.
- A χ-coloring $(\chi \geq K)$ is called *strong* if every hyperedge contains K different colors.

These three notions are interesting in their own right, and have been independently studied. It is easy to see that ℓ-discrepancy ($\ell < K$), χ-rainbow colorability ($2 \leq \chi \leq K$), and χ-strong colorability $(K \leq \chi \leq 2K - 2)$ all imply 2-colorability. For odd K, both $(K + 1)$-strong colorability and $(K - 1)$-rainbow colorability imply discrepancy-1, so strong colorability and rainbow colorability seem stronger than low discrepancy.

2.3.1 MIN COLORING

We prove the following strong hardness result.

Theorem 2.3.1. *For any $\epsilon > 0$ and $Q, k \geq 2$, given a Qk-uniform hypergraph $H = (V, E)$, it is NP-hard to distinguish between the following cases.*

- *Completeness: There is a k-coloring $c : V \to [k]$ such that for every hyperedge $e \in E$ and color $i \in [k]$, e has at least $Q - 1$ vertices of color i.*
- *Soundness: Every $I \subseteq V$ of measure ϵ induces at least a fraction $\epsilon^{O_{Q,k}(1)}$ of hyperedges. In particular, there is no independent set of measure ϵ, and every $\lfloor \frac{1}{\epsilon} \rfloor$-coloring of H induces a monochromatic hyperedge.*

Fixing $Q = 2$ gives a hardness of rainbow coloring with K optimized to be $2k$.

Corollary 2.3.2. *For all integers $c, k \geq 2$, given a $2k$-uniform hypergraph H, it is NP-hard to distinguish whether H is rainbow k-colorable or is not even c-colorable.*

On the other hand, fixing $k = 2$ gives a strong hardness result of discrepancy minimization (with 2 colors).

Corollary 2.3.3. *For any $c, Q \geq 2$, given a $2Q$-uniform hypergraph $H = (V, E)$, it is NP-hard to distinguish whether H is 2-colorable with discrepancy 2 or is not even c-colorable.*

The above result strengthens the result of Austrin et al [AGH14] that shows hardness of 2-coloring in the soundness case. However, their result also holds in $(2Q+1)$-uniform hypergraphs with discrepancy 1, which is not covered by the results in this thesis. For algorithms, we prove that all three promises lead to an $\tilde{O}(n^{\frac{1}{k}})$-coloring that is decreasing in k.

Theorem 2.3.4. *Consider any k-uniform hypergraph $H = (V, E)$ with n vertices and m edges. For any $\ell < O(\sqrt{k})$, If H has discrepancy-ℓ, $(k-\ell)$-rainbow colorable, or $(k+\ell)$-strong colorable, one can color H with $\tilde{O}((\frac{m}{n})^{\frac{\ell^2}{k^2}}) \leq \tilde{O}(n^{\frac{\ell^2}{k}})$ colors.*

These results significantly improve the current best algorithm that assumes only 2-colorability and uses $\tilde{O}(n^{1-\frac{1}{k}})$ colors. Table 2.2 summarizes our results.

Promised Coloring Structure	Algorithm	Hardness
K-partite	2-colorable	Not rainbow K-colorable (Almost/UG) Not weak $O(1)$-colorable
Rainbow $(K-1)$-colorable	$\tilde{O}(n^{1/K})$**-colorable**	**(Almost) Not weak** $O(1)$**-colorable**
Rainbow $\frac{K}{2}$-colorable		**Not weak** $O(1)$**-colorable**
2-colorable with perfect balance	2-colorable	
2-colorable with discrepancy 1	$\tilde{O}(n^{1/K})$**-colorable**	Not 2-colorable
2-colorable with discrepancy 2	$\tilde{O}(n^{4/K})$**-colorable**	**Not weak** $O(1)$**-colorable**

Table 2.2: Summary of algorithmic and hardness results for MIN COLORING a highly structured K-uniform hypergraph. Almost means that $\epsilon > 0$ fraction of vertices and incident hyperedges must be deleted to have the structure. UG indicates that the result is based on the Unique Games Conjecture. The results of this thesis are in boldface.

2.3.2 MAX 2-COLORING

For MAX 2-COLORING, we prove that our three promises, unlike mere 2-colorability, give enough structure for polynomial time algorithms to perform significantly better than naive algorithms. We also study these promises from a hardness perspective to understand the asymptotic threshold at which beating naive algorithms goes from easy to UG/NP-Hard. In particular assuming the UGC, for MAX 2-COLORING under ℓ-discrepancy or $(K-\ell)$-rainbow colorability, this threshold is $\ell = \Theta(\sqrt{K})$.

Theorem 2.3.5. *There is a randomized polynomial time algorithm that produces a 2-coloring of a K-uniform hypergraph H with the following guarantee. For any $0 < \epsilon < \frac{1}{2}$*

(let $\ell = K^{\epsilon}$), there exists a constant $\eta > 0$ such that if H is $(K-\ell)$-rainbow colorable or $(K+\ell)$-strong colorable, the fraction of monochromatic edges in the produced 2-coloring is $O((\frac{1}{K})^{\eta K})$ in expectation.

For the ℓ-discrepancy case, we observe that when $\ell < \sqrt{K}$, our work for SYMMETRIC CSP yields an approximation algorithm that marginally (by an additive factor much less than 2^{-K}) outperforms the random assignment.

The following hardness results suggest that this gap between low-discrepancy and rainbow/strong colorability might be intrinsic. Table 2.3 summarizes our results.

Theorem 2.3.6. *For sufficiently large odd K, given a K-uniform hypergraph which admits a 2-coloring with at most a $(\frac{1}{2})^{6K}$ fraction of edges of discrepancy larger than 1, it is UG-hard to find a 2-coloring with a $(\frac{1}{2})^{5K}$ fraction of monochromatic edges.*

Theorem 2.3.7. *For even $K \geq 4$, given a K-uniform hypergraph which admits a 2-coloring with no edge of discrepancy larger than 2, it is NP-hard to find a 2-coloring with a $K^{-O(K)}$ fraction of monochromatic edges.*

Theorem 2.3.8. *For K sufficiently large, given a K-uniform hypergraph which admits a 2-coloring with no edge of discrepancy larger than $O(\log K)$, it is NP-hard to find a 2-coloring with a $2^{-O(K)}$ fraction of monochromatic edges.*

Theorem 2.3.9. *For K such that $\chi := K - \sqrt{K}$ is an integer greater than 1, and any $\epsilon > 0$, given a K-uniform hypergraph which admits a χ-coloring with at most ϵ fraction of non-rainbow edges, it is UG-hard to find a 2-coloring with a $(\frac{1}{2})^{K-1}$ fraction of monochromatic edges.*

Promises	Algorithm	Hardness
ℓ-Discrepancy	$1-(1/2)^{K-1}+\delta, \quad \ell < \sqrt{K}$	UG: $1-(1/2)^{5K}, \quad \ell = 1$ NP: $1-(1/K)^{O(K)}, \quad \ell = 2$ NP: $1-(1/2)^{O(K)}, \quad \ell = \Omega(\log K)$ UG: $1-(1/2)^{K-1}, \quad \ell \geq \sqrt{K}$
$(K-\ell)$-Rainbow	$1-(1/K)^{\Omega(K)}, \quad \ell = o(K)$	UG: $1-(1/2)^{K-1}, \quad \ell \geq \sqrt{K}$
$(K+\ell)$-Strong	$1-(1/K)^{\Omega(K)}, \quad \ell = o(K).$	

Table 2.3: Summary of our algorithmic and hardness results for MAX 2-COLORING with valid ranges of ℓ.

2.4 Subgraph Transversal and Graph Partitioning

Consider the following two classic problems.

> MINIMUM SET TRANSVERSAL
>
> Input: A universe U and a collection of subsets $S_1, \ldots, S_m$.
>
> Output: $F \subseteq U$ such that F intersects every S_i.
>
> Goal: Minimize $|F|$.

This problem is equivalent to MINIMUM SET COVER by taking the dual set system.

> MAXIMUM SET PACKING
>
> Input: A universe U and a collection of subsets $S_1, \ldots, S_m$.
>
> Output: A subcollection $S_{i_1}, \ldots, S_{i_{m'}}$ which are pairwise disjoint.
>
> Goal: Maximize m'.

Given the same input, it is clear that the optimum of the former is always at least that of the latter (i.e. weak duality holds). Studying the (approximate) reverse direction of the inequality (i.e. strong duality) as well as the complexity of both problems for many interesting classes of set systems is arguably the most studied paradigm in combinatorial optimization.

We focus on set systems where the size of each set is bounded by a constant k. With this restriction, MINIMUM SET TRANSVERSAL and MAXIMUM SET PACKING are known as k-HYPERGRAPH VERTEX COVER (k-HVC) and k-SET PACKING (k-SP), respectively. This assumption significantly simplifies the problem since there are at most n^k sets. While there is a simple factor k-approximation algorithm for both problems, it is NP-hard to approximate k-HVC and k-SP within a factor less than $k-1$ [DGKR05] and $O(\frac{k}{\log k})$ [HSS06] respectively.

2.4.1 H-Transversal / Packing

We study the following special cases of k-HYPERGRAPH VERTEX COVER and k-SET PACKING. Let H be a fixed graph with k vertices.

H-TRANSVERSAL

Input: A graph $G = (V, E)$

Output: $F \subseteq V$ such that the induced subgraph $G|_{V \setminus F}$ does not have H as a subgraph.

Goal: Minimize $|F|$.

H-PACKING

Input: A graph $G = (V, E)$

Output: Disjoint subsets $S_1, \dots, S_m \subseteq V$ where for each i, $|S_i| = k$ and $G|_{S_i}$ has H as a subgraph.

Goal: Maximize m.

These problems capture VERTEX COVER and MAXIMUM MATCHING as special cases when H is a single edge. Other special cases where H is a clique or a cycle have been also actively studied. In this thesis, we study approximabilities of H-TRANSVERSAL and H-PACKING for every fixed graph H. They admit a simple k-approximation algorithm as special cases of k-HYPERGRAPH VERTEX COVER and k-SET PACKING. We study whether significantly better approximation algorithms (i.e., k^δ-approximation for some $\delta < 1$) exist. Our main hardness result is the following.

Theorem 2.4.1. *If H is a 2-vertex connected with k vertices, unless* $\mathbf{NP} \subseteq \mathbf{BPP}$, *no polynomial time algorithm approximates H-*TRANSVERSAL *within a factor better than $k - 1$, and H-*PACKING *within a factor better than $\Omega(\frac{k}{\log^7 k})$.*

This result leaves us to study 1-connected graphs. In particular, we focus on k-Star and k-Path, where k-Star denote $K_{1,k-1}$, the complete bipartite graph with 1 and $k-1$ vertices on each side, and k-Path is a simple path with k vertices. It is easy to see that k-STAR PACKING is as hard to approximate as MAXIMUM INDEPENDENT SET on $(k-1)$-regular graphs, which is NP-hard to approximate within a factor $\Omega(\frac{k}{\log^4 k})$ [Cha13]. We show that both k-STAR TRANSVERSAL and k-PATH TRANSVERSAL admit a good approximation algorithm.

Theorem 2.4.2. k-STAR TRANSVERSAL *can be approximated within a factor of $O(\log k)$ in polynomial time.*

Theorem 2.4.3. *There is an $O(\log k)$-approximation algorithm for k-*PATH TRANSVERSAL *that runs in time $2^{O(k^3 \log k)} n^{O(1)}$.*

Note that the exponential dependence of the running time on k is necessary since finding a k-Path for general k is NP-hard. Table 2.4 summarizes our results.

Promises	H-TRANSVERSAL	H-PACKING
2-connected	Hard to approximate within $k-1$	Hard to approximate within $\Omega(\frac{k}{\text{polylog}(k)})$
k-Star	Admits $O(\log k)$-approximation	Hard to approximate within $\Omega(\frac{k}{\text{polylog}(k)})$
k-Path	Admits $O(\log k)$-approximation	?

Table 2.4: Summary of our algorithmic and hardness results for H-TRANSVERSAL and H-PACKING for different H.

2.4.2 Partitioning a Graph into Small Pieces

On the way to our algorithm for k-PATH TRANSVERSAL, we study the following natural graph partitioning problems.

k-VERTEX SEPARATOR

Input: An undirected graph $G = (V, E)$ and $k \in \mathbb{N}$.

Output: A subset $S \subseteq V$ such that in the subgraph induced by $V \setminus S$ (denoted by $G|_{V \setminus S}$), each connected component has at most k vertices.

Goal: Minimize $|S|$.

k-VERTEX SEPARATOR is a special case of $(k+1)$-HVC and similar to H-TRANSVERSAL in the sense that the goal is to remove the minimum number of vertices such that G has no connected graph with $k+1$ vertices as a subgraph (H is replaced by a family of connected graphs with $k+1$ vertices). The edge version can be defined similarly.

k-EDGE SEPARATOR

Input: An undirected graph $G = (V, E)$ and $k \in \mathbb{N}$.

Output: A subset $S \subseteq E$ such that in the subgraph $(V, E \setminus S)$, each connected component has at most k vertices.

Goal: Minimize $|S|$.

Our primary focus is on the case where k is either a constant or a slowly growing function of n (e.g. $O(\log n)$ or $n^{o(1)}$). Our problems can be interpreted as a special case of three general classes of problems that have been studied separately (balanced graph partitioning, k-HYPERGRAPH VERTEX COVER, and fixed parameter tractability (FPT)).

Our main result is the following algorithm for k-VERTEX SEPARATOR. For fixed constants $b, c > 1$, an algorithm for k-VERTEX SEPARATOR is called an (b, c)-bicriteria approximation algorithm if given an instance $G = (V, E)$ and $k \in \mathbb{N}$, it outputs $S \subseteq V$ such that (1) each connected component of $G|_{S \setminus V}$ has at most bk vertices and (2) $|S|$ is at most c times the optimum of k-VERTEX SEPARATOR.

Theorem 2.4.4. *For any $\epsilon \in (0, 1/2)$, there is a polynomial time $(\frac{1}{1-2\epsilon}, O(\frac{\log k}{\epsilon}))$-bicriteria approximation algorithm for k-*VERTEX SEPARATOR.

Setting $\epsilon = \frac{1}{4}$ and running the algorithm yields $S \subseteq V$ with $|S| \leq O(\log k) \cdot \mathsf{Opt}$ such that each component in $G|_{V \setminus S}$ has at most $2k$ vertices. Performing an exhaustive search in each connected component yields the following true approximation algorithm whose running time depends exponentially only on k.

Corollary 2.4.5. *There is an $O(\log k)$-approximation algorithm for k-*VERTEX SEPARATOR *that runs in time $n^{O(1)} + 2^{O(k)}n$.*

This gives an FPT approximation algorithm when parameterized by only k, and its approximation ratio $O(\log k)$ improves the simple $(k+1)$-approximation from k-HYPERGRAPH VERTEX COVER. When $\mathsf{Opt} \gg k$, it runs even faster than the time lower bound $k^{\Omega(\mathsf{Opt})} n^{\Omega(1)}$ for the exact algorithm assuming the Exponential Time Hypothesis [DDvH14].

The natural question is whether superpolynomial dependence on k is necessary to achieve *true* $O(\log k)$-approximation. The following theorem proves hardness of k-VERTEX SEPARATOR based on DENSEST k-SUBGRAPH. In particular, a polynomial time $O(\log k)$-approximation algorithm for k-VERTEX SEPARATOR will imply $O(\log^2 n)$-approximation algorithm for DENSEST k-SUBGRAPH. Given that the best approximation algorithm achieves $\approx O(n^{1/4})$-approximation [BCC+10] and $n^{\Omega(1)}$-rounds of the Sum-of-Squares hierarchy have a gap at least $n^{\Omega(1)}$ [BCV+12], such a result seems unlikely or will be considered as a breakthrough.

Theorem 2.4.6. *If there is a polynomial time f-approximation algorithm for k-*VERTEX SEPARATOR*, then there is a polynomial time $2f^2$-approximation algorithm for* DENSEST k-SUBGRAPH.

For k-EDGE SEPARATOR, we prove that the true $O(\log k)$-approximation can be achieved in polynomial time. This shows a stark difference between the vertex version and the edge version.

Theorem 2.4.7. *There is an $O(\log k)$-approximation algorithm for k-EDGE SEPARATOR that runs in time $n^{O(1)}$.*

When $k = n^{o(1)}$, our algorithm outperforms the previous best approximation algorithm [KNS09, ENRS99].

2.5 Cut Problems

Part V proves the improved hardness results for many cut problems. Almost all results are based on the common framework called the length-control dictatorship test.

2.5.1 DIRECTED MULTICUT

Given a directed graph and two vertices s and t, one of the most natural variants of s-t MIN CUT is to remove the fewest edges to ensure that there is no directed path from s to t and no directed path from t to s. This problem is known as s-t BICUT and admits the trivial 2-approximation algorithm by computing the minimum s-t cut and t-s cut.

DIRECTED MULTIWAY CUT is a generalization of s-t BICUT that has been actively studied. Given a directed graph with k terminals $s_1, \dots, s_k$, the goal is to remove the fewest number of edges such that there is no path from s_i to s_j for any $i \neq j$. DIRECTED MULTIWAY CUT also admits 2-approximation [NZ01, CM16]. If k is allowed to increase polynomially with n, there is a simple reduction from VERTEX COVER that shows $(2-\epsilon)$-approximation is hard under the UGC [GVY94, KR08].

DIRECTED MULTIWAY CUT can be further generalized to DIRECTED MULTICUT. Given a directed graph with k source-sink pairs $(s_1, t_1), \dots, (s_k, t_k)$, the goal is to remove the fewest number of edges such that there is no path from s_i to t_i for any i. Computing the minimum s_i-t_i cut for all i separately gives the trivial k-approximation algorithm. Chuzhoy and Khanna [CK09] showed DIRECTED MULTICUT is hard to approximate within a factor $2^{\Omega(\log^{1-\epsilon} n)} = 2^{\Omega(\log^{1-\epsilon} k)}$ when k is polynomially growing with n. Agarwal et al. [AAC07] showed $\tilde{O}(n^{\frac{11}{23}})$-approximation algorithm, which improves the trivial k-approximation when k is large.

Very recently, Chekuri and Madan [CM16] showed a simple approximation-preserving reduction from DIRECTED MULTICUT with $k = 2$ to s-t BICUT (the other direction is trivially true), and (UNDIRECTED) NODE-WEIGHTED MULTIWAY CUT with $k = 4$ to s-t BICUT. Since NODE-WEIGHTED MULTIWAY CUT with $k = 4$ is hard to approximate within a factor $1.5 - \epsilon$ under the UGC [EVW13] (matching the algorithm of Garg et al. [GVY94]), the same hardness holds for s-t BICUT, DIRECTED MULTIWAY CUT, and DIRECTED MULTICUT for constant k. To the best of our knowledge, $1.5 - \epsilon$ is the best hardness factor for constant k even assuming the UGC. In the same paper, Chekuri and Madan [CM16] asked whether a factor $2-\epsilon$ hardness holds for s-t BICUT under the UGC.

We prove that for any constant $k \geq 2$, the trivial k-approximation for DIRECTED MULTICUTmight be optimal. Our result for $k = 2$ gives the optimal hardness result for s-t BICUT, answering the question of Chekuri and Madan.

Theorem 2.5.1. *Assuming the Unique Games Conjecture, for every constant $k \geq 2$ and $\epsilon > 0$,* DIRECTED MULTICUT *with k source-sink pairs is NP-hard to approximate within a factor $k - \epsilon$.*

Corollary 2.5.2. *Assuming the Unique Games Conjecture, for any $\epsilon > 0$, s-t* BICUT *is hard to approximate within a factor $2 - \epsilon$.*

2.5.2 Bicuts

The hardness of s-t BICUT suggests that it may be hard to outperform a simple approximation algorithm that outputs the union of the min s-t cut and the min t-s cut. This strong hardness result also motivates the following question: *Can an algorithm do better if it can choose s and t?* Formally, in the global version of bicut, denoted EDGE BICUT, the goal is to find the smallest number of edges whose deletion ensures that there exist two distinct nodes s and t such that s cannot reach t and t cannot reach s in the resulting digraph.

The dichotomy between global cut problems and fixed-terminal cut problems in undirected graphs is well-known. For concreteness, recall EDGE 3-CUT and EDGE 3-WAY CUT. In EDGE 3-CUT, the input is an undirected graph and the goal is to find the smallest number of edges to delete so that the resulting graph has at least 3 connected components. In EDGE 3-WAY CUT, the input is an undirected graph with 3 specified nodes and the goal is to find the smallest number of edges to delete so that the resulting graph has at least 3 connected components with at most one of the 3 specified nodes in each component. While EDGE 3-WAY CUT is NP-hard [DJP+94], EDGE 3-CUT is solvable efficiently [GH94]. Similarly, while s-t EDGE BICUT is inapproximable to a factor better than 2 assuming UGC, EDGE BICUT is approximable within a factor of $2 - 1/448$ [BCK+17].

We also consider the problem between s-t BICUT and EDGE BICUT, denoted s-$*$ EDGE BICUT: Given a directed graph with a specified node s, find the smallest number of edges to delete so that there exists a node t such that s cannot reach t and t cannot reach s in the resulting graph. s-$*$ EDGE BICUT admits a 2-approximation by guessing the terminal t and then using the 2-approximation for s-t EDGE BICUT. We show the following inapproximability results for s-$*$ EDGE BICUT:

Theorem 2.5.1. s-$*$ EDGE BICUT *has no efficient* $(4/3 - \epsilon)$*-approximation for any* $\epsilon > 0$ *assuming the Unique Games Conjecture.*

Furthermore, we consider the node-weighted variant of bicut, denoted NODE BICUT: Given a directed graph, find the smallest number of nodes whose deletion ensures that there exist nodes s and t such that s cannot reach t and t cannot reach s in the resulting graph. Every directed graph that is not a tournament has a feasible solution to NODE BICUT. NODE BICUT admits a 2-approximation by a simple reduction to s-t EDGE BICUT. We show the following inapproximability results.

Theorem 2.5.2. NODE BICUT *has no efficient* $(3/2 - \epsilon)$*-approximation for any* $\epsilon > 0$ *assuming the Unique Games Conjecture.*

2.5.3 Double Cuts

Recall that an arborescence in a directed graph $D = (V, E)$ is a minimal subset $F \subseteq E$ of arcs such that there exists a node $r \in V$ with every node $u \in V$ having a unique path from r to u in the subgraph (V, F) (e.g., see [Sch03]).

The input to the NODE DOUBLE CUT problem is a directed graph and the goal is to find the smallest number of nodes whose deletion ensures that the remaining graph has no arborescence. This problem is key to understanding fault tolerant consensus in networks [TV15].

A directed graph $D = (V, E)$ has no arborescence if and only if [2] there exist two distinct nodes $s, t \in V$ such that every node $u \in V$ can reach at most one node in $\{s, t\}$ [BP13]. By this characterization, every directed graph that is not a tournament has a feasible solution to NODE DOUBLE CUT. This characterization motivates the following fixed-terminal version, denoted s-t NODE DOUBLE CUT: Given a directed graph with two specified nodes s and t, find the smallest number of nodes whose deletion ensures that

[2]We believe that this characterization led earlier authors [BP13] to coin the term *double cut* to refer to the edge deletion variant of the problem and we are following this naming convention.

every remaining node u can reach at most one node in $\{s, t\}$ in the resulting graph. An instance of s-t NODE DOUBLE CUT has a feasible solution provided that the instance has no edge between s and t. An efficient algorithm to solve/approximate s-t NODE DOUBLE CUT immediately gives an efficient algorithm to solve/approximate NODE DOUBLE CUT.

In the edge-weighted variation of two-terminal double cut, namely s-t EDGE DOUBLE CUT, the goal is to delete the smallest number of edges to ensure that every node in the graph can reach at most one node in $\{s, t\}$. Similarly, in the global variant, denoted EDGE DOUBLE CUT, the goal is to delete the smallest number of edges to ensure that there exist nodes s, t such that every node u can reach at most one node in $\{s, t\}$. Thus, EDGE DOUBLE CUT is equivalent to deleting the smallest number of edges to ensure that the graph has no arborescence. The fixed-terminal variant s-t EDGE DOUBLE CUT is solvable in polynomial time using maximum flow and, consequently, EDGE DOUBLE CUT is also solvable in polynomial time [BP13].

We show the following inapproximability results for s-t NODE DOUBLE CUT.

Theorem 2.5.3. s-t NODE DOUBLE CUT *has no efficient* $(2 - \epsilon)$*-approximation for any* $\epsilon > 0$ *assuming the Unique Games Conjecture.*

This matches a 2-approximation algorithm for s-t NODE DOUBLE CUT [BCK+17], which also leads to a 2-approximation for NODE DOUBLE CUT. Note that the inapproximability results for s-t NODE DOUBLE CUT do not imply the hardness of NODE DOUBLE CUT. We also have the following inapproximability of NODE DOUBLE CUT.

Theorem 2.5.4. NODE DOUBLE CUT *has no efficient* $(3/2 - \epsilon)$*-approximation for any* $\epsilon > 0$ *assuming the Unique Games Conjecture.*

2.5.4 NODE k-CUT and VERTEX COVER ON k-PARTITE GRAPHS

Another way to show hardness of NODE DOUBLE CUT is a reduction from the node-weighted 3-cut problem in undirected graphs, though Theorem 16.2.4 shows a better hardness using length-control dictatorship tests and we do not show this reduction in this thesis ([BCK+17] presents this reduction to show an inapproximability result only assuming $\mathbf{P} \neq \mathbf{NP}$).

In the node weighted k-cut problem, denoted NODE 3-CUT, the input is an undirected graph and the goal is to find the smallest subset of nodes whose deletion leads to at least k connected components in the remaining graph. A classic result of Goldschmidt and Hochbaum [GH94] showed that the edge-weighted variant, denoted EDGE k-CUT(more commonly known as k-CUT)—namely find a smallest subset of edges of a given undirected graph whose deletion leads to at least k connected components—is solvable in polynomial time when k is a constant. Surprisingly, the complexity of NODE k-CUT for $k = 3$ is open. NODE k-CUT admits a $2(k-1)/k$-approximation algorithm [GVY04], and there is a simple approximation preserving reduction from VERTEX COVER ON k-PARTITE GRAPHS to NODE k-CUT. We prove that VERTEX COVER ON k-PARTITE GRAPHS is hard to approximate within a factor $2(k-1)/k$ assuming the UGC, so that $2(k-1)/k$ may be the optimal approximation factor for both VERTEX COVER ON k-PARTITE GRAPHS and NODE k-CUT.

Theorem 2.5.5. VERTEX COVER ON k-PARTITE GRAPHS *has no efficient* $(2(k-1)/k - \epsilon)$*-approximation algorithm for any* $\epsilon > 0$ *assuming the Unique Games Conjecture.*

Theorem 16.2.1 for s-$*$ EDGE BICUT and Theorem 16.2.2 for NODE BICUT follow from the above theorem as they are as hard to approximate as VERTEX COVER ON k-PARTITE GRAPHS for $k = 3$ and $k = 4$ respectively (see Section 16.5). We finally note that Theorem 16.2.5 is the only UG-hardness result in this part that does not require a length-control dictatorship test.

2.5.5 LENGTH-BOUNDED CUT/ SHORTEST PATH INTERDICTION

Another natural variant of s-t MIN CUT is the LENGTH-BOUNDED CUT problem, where given an integer l, we only want to cut s-t paths of length strictly less than l.[3] Its practical motivation is based on the fact that in most communication/transportation networks, short paths are preferred to be used to long paths [MM10].

Lovász et al. [LNLP78] gave an exact algorithm for LENGTH-BOUNDED VERTEX CUT ($l \leq 5$) in undirected graphs. Mahjoub and McCormick [MM10] proved that LENGTH-BOUNDED VERTEX CUT admits an exact polynomial time algorithm for $l \leq 4$ in undirected graphs. Baier et al. [BEH$^+$10] showed that both LENGTH-BOUNDED VERTEX CUT ($l > 5$) and LENGTH-BOUNDED EDGE CUT($l > 4$) are NP-hard to approximate within a factor 1.1377. They presented $O(\min(l, \frac{n}{l})) = O(\sqrt{n})$-approximation algorithm for LENGTH-BOUNDED VERTEX CUT and $O(\min(l, \frac{n^2}{l^2}, \sqrt{m})) = O(n^{2/3})$-approximation algorithm for LENGTH-BOUNDED EDGE CUT, with matching LP gaps. LENGTH-BOUNDED CUT problems have been also actively studied in terms of their fixed parameter tractability [GT11, DK15, BNN15, FHNN15].

[3]It is more conventional to cut s-t paths of length *at most* l. We use this slightly unconventional way to be more consistent with SHORTEST PATH INTERDICTION.

If we exchange the roles of the objective k and the length bound l, the problem becomes SHORTEST PATH INTERDICTION, where we want to maximize the length of the shortest s-t path after removing at most k vertices or edges. It is also one of the central problems in a broader class of *interdiction* problems. The study of SHORTEST PATH INTERDICTION started in 1980's when the problem was called as the *k-most-vital-arcs* problem [CD82, MMG89, BGV89] and proved to be NP-hard [BGV89]. Khachiyan et al. [KBB$^+$07] proved that it is NP-hard to approximate within a factor less than 2. While many heuristic algorithms were proposed [IW02, BB08, Mor11] and hardness in planar graphs [PS13] was shown, whether the general version admits a constant factor approximation was still unknown.

Given a graph $G = (V, E)$ and $s, t \in V$, let $\mathsf{dist}(G)$ be the length of the shortest s-t path. For $V' \subseteq V$, let $G \setminus V'$ be the subgraph induced by $V \setminus V'$. For $E' \subseteq E$, we use the same notation $G \setminus E'$ to denote the subgraph $(V, E \setminus E')$. We primarily study undirected graphs. We first present our results for the vertex version of both problems (collectively called as SHORT PATH VERTEX CUT onwards).

Theorem 2.5.3. *Assume the Unique Games Conjecture. For infinitely many values of the constant $l \in \mathbb{N}$, given an undirected graph $G = (V, E)$ and $s, t \in V$ where there exists $C^* \subseteq V \setminus \{s, t\}$ such that $\mathsf{dist}(G \setminus C^*) \geq l$, it is NP-hard to perform any of the following tasks.*

1. *Find $C \subseteq V \setminus \{s, t\}$ such that $|C| \leq \Omega(l) \cdot |C^*|$ and $\mathsf{dist}(G \setminus C) \geq l$.*
2. *Find $C \subseteq V \setminus \{s, t\}$ such that $|C| \leq |C^*|$ and $\mathsf{dist}(G \setminus C) \geq O(\sqrt{l})$.*
3. *Find $C \subseteq V \setminus \{s, t\}$ such that $|C| \leq \Omega(l^{\frac{\epsilon}{2}}) \cdot |C^*|$ and $\mathsf{dist}(G \setminus C) \geq O(l^{\frac{1+\epsilon}{2}})$ for some $0 < \epsilon < 1$.*

The first result shows that LENGTH-BOUNDED VERTEX CUT is hard to approximate within a factor $\Omega(l)$. This matches the best $\frac{l}{2}$-approximation up to a constant. [BEH$^+$10]. The second result shows that SHORTEST PATH VERTEX INTERDICTION is hard to approximate with in a factor $\Omega(\sqrt{\mathsf{Opt}})$, and the third result rules out *bicriteria approximation* — for any constant c, it is hard to approximate both l and $|C^*|$ within a factor of c.

The above results hold for directed graphs by definition. Our hard instances will have a natural *layered* structure, so it can be easily checked that the same results (up to a constant) hold for directed acyclic graphs. Since one vertex can be split as one directed edge, the same results hold for the edge version in directed acyclic graphs.

For LENGTH-BOUNDED EDGE CUT and SHORTEST PATH EDGE INTERDICTION in undirected graphs (collectively called SHORTEST PATH EDGE CUT onwards), we prove the following theorems.

Theorem 2.5.4. *Assume the Unique Games Conjecture. For infinitely many values of the constant $l \in \mathbb{N}$, given an undirected graph $G = (V, E)$ and $s, t \in V$ where there exists $C^* \subseteq E$ such that $\mathsf{dist}(V \setminus C^*) \geq l$, it is NP-hard to perform any of the following tasks.*

1. *Find $C \subseteq E$ such that $|C| \leq \Omega(\sqrt{l}) \cdot |C^*|$ and $\mathsf{dist}(G \setminus C) \geq l$.*
2. *Find $C \subseteq E$ such that $|C| \leq |C^*|$ and $\mathsf{dist}(G \setminus C) \geq l^{\frac{2}{3}}$.*
3. *Find $C \subseteq E$ such that $|C| \leq \Omega(l^{\frac{2\epsilon}{3}}) \cdot |C^*|$ and $\mathsf{dist}(G \setminus C) \geq O(l^{\frac{2+2\epsilon}{3}})$ for some $0 < \epsilon < \frac{1}{2}$.*

Our hardness factors for the edge versions, $\Omega(\sqrt{l})$ for LENGTH-BOUNDED EDGE CUT and $\Omega(\sqrt[3]{\mathsf{Opt}})$ for SHORTEST PATH EDGE INTERDICTION, are slightly weaker than those for their vertex counterparts, but we are not aware of any approximation algorithm specialized for the edge versions. It is an interesting open problem whether there exist better approximation algorithms for the edge versions.

2.5.6 RMFC

RESOURCE MINIMIZATION FOR FIRE CONTAINMENT (RMFC) is a problem closely related to LENGTH-BOUNDED CUT with the additional notion of time. Given a graph G, a vertex s, and a subset T of vertices, consider the situation where fire starts at s on Day 0. For each Day i ($i \geq 1$), we can *save* at most k vertices, and the fire spreads from currently burning vertices to its unsaved neighbors. Once a vertex is burning or saved, it remains so from then onwards. The process is terminated when the fire cannot spread anymore. RMFC asks to find a strategy to save k vertices each day with the minimum k so that no vertex in T is burnt. These problems model the spread of epidemics or ideas through a social network, and have been actively studied recently [CC10, ACHS12, ABZ16, CV16].

RMFC, along with other variants, is first introduced by Hartnell [Har95]. Another well-studied variant is called the FIREFIGHTER problem, where we are only given $s \in V$ and want to maximize the number of vertices that are not burnt at the end. It is known to be NP-hard to approximate within a factor $n^{1-\epsilon}$ for any $\epsilon > 0$ [ACHS12]. King and MacGillivray [KM10] proved that RMFC is hard to approximate within a factor less than 2. Anshelevich et al. [ACHS12] presented an $O(\sqrt{n})$-approximation algorithm for general graphs, and Chalermsook and Chuzhoy [CC10] showed that RMFC admits $O(\log^* n)$-approximation in trees. Very recently, the approximation ratio in trees has been improved to $O(1)$ [ABZ16]. Both Anshelevich et al. [ACHS12] and Chalermsook and Chuzhoy [CC10] independently studied directed layer graphs with b layers, showing $O(\log b)$-approximation.

Our final result on RMFC assumes a variant of the Unique Games Conjecture which is not known to be equivalent to the original UGC. See Conjecture 3.2.4 for the exact statement.

Theorem 2.5.5. *Assuming Conjecture 3.2.4, it is NP-hard to approximate* RMFC *in undirected graphs within any constant factor.*

Again, our reduction has a natural layered structure and the result holds for directed layered graphs. With b layers, we prove that it is hard to approximate with in a factor $\Omega(\log b)$, matching the best approximation algorithms [CC10, ACHS12]. Table 2.5 summarizes our results.

Promises	Algorithm	Hardness
DIRECTED MULTICUT	k	$k-\epsilon$
s-t BICUT	2	$2-\epsilon$
s-$*$ EDGE BICUT	2	$4/3-\epsilon$
NODE BICUT	2	$3/2-\epsilon$
s-t NODE DOUBLE CUT	2	$2-\epsilon$
NODE DOUBLE CUT	2	$3/2-\epsilon$
LENGTH-BOUNDED VERTEX CUT	$O(l)$	$\Omega(l)$
SHORTEST PATH VERTEX INTERDICTION		$\Omega(\sqrt{\mathsf{Opt}})$
LENGTH-BOUNDED EDGE CUT	$O(l)$	$\Omega(\sqrt{l})$
SHORTEST PATH EDGE INTERDICTION		$\Omega(\sqrt[3]{\mathsf{Opt}})$
NODE k-CUT/ VERTEX COVER ON k-PARTITE GRAPHS	$2(k-1)/k$	$2(k-1)/k-\epsilon$
RMFC	$\sqrt{n}$	$\omega(1)$

Table 2.5: Summary of our hardness results for various cut problems.

Chapter 3

Preliminaries

This chapter introduces the common tools used to design an efficient approximation algorithms or prove hardness of approximation, including LABEL COVER, UNIQUE GAMES, Fourier analysis, LP/SDP relaxations and their hierarchies.

3.1 LABEL COVER

The celebrated PCP theorem [ALM+98, AS98] and the parallel repetition theorem [Raz98] proved a strong hardness result for LABEL COVER. It has been used to prove optimal hardness results for MAXIMUM INDEPENDENT SET [Hås96], MAX CSP [Cha13], CHROMATIC NUMBER [FK98], and MINIMUM SET COVER [Fei98]. See Trevisan's survey [Tre05] and the textbook of Ausiello et al. [ACG+99] for overview of these results. Many textbooks on approximation algorithms [Hoc96, Vaz01, WS11] also have a chapter devoted to introduce hardness of LABEL COVER and its consequences. In this thesis, it is used in Chapter 6 for UNIQUE COVERAGE, and Part III for hypergraph coloring.

An instance of LABEL COVER consists of a biregular bipartite graph $G = (U_G \cup V_G, E_G)$ where each edge $e = (u, v)$ is associated with a projection $\pi_e : [R] \mapsto [L]$ for some positive integers R and L. For $u \in U_G$, let $N(u)$ denote its neighbors and $D := |N(u)|$ be the left degree. We additionally require that every projection π_e is d-regular, i.e., $R = dL$ and for every $j \in [L]$, $|\pi_e^{-1}(j)| = d$. A labeling $l : U_G \cup V_G \mapsto [R]$ satisfies $e = (u, v)$ when $\pi_e(l(v)) = l(u)$. The standard application of PCP Theorem, Parallel Repetition Theorem, and the trick of Wenner [Wen13] to make each projection

d-regular[1], implies the following theorem.

Theorem 3.1.1. *There exists an absolute constant $\tau < 1$ such that the following is true. For any positive integer $r > 0$, there is a reduction that given an instance ϕ of 3SAT with n variables, outputs an instance of* LABEL COVER $(G, \{\pi_e\}_e)$ *with* $|U_G|, |V_G| = n^{O(r)}$, $R = 10^r, L = 2^r, d = D = 5^r$ *in time $n^{O(r)}$, and satisfies the following.*

- *Completeness: If ϕ is satisfiable, there exists a labeling that satisfies every projection.*
- *Soundness: If ϕ is not satisfiable, every labeling satisfies at most τ^r fraction of projections.*

The above basic LABEL COVER is used to prove a nearly optimal hardness result for UNIQUE COVERAGE in Chapter 6. Several variants of LABEL COVER have been previously developed to prove hardness of various problems (e.g., SMOOTH LABEL COVER to prove hardness for hypergraph coloring [Kho02a], MULTILAYERED LABEL COVER for K-HYPERGRAPH VERTEX COVER[DGKR05] HYPERGRAPH LABEL COVER to prove hardness for polynomial reconstruction [GKS10]). For hardness of hypergraph coloring under strong promises in Part III, we introduce the following variants of LABEL COVER. We introduce them and their hardness results in Section 10.2.

3.2 UNIQUE GAMES

The Unique Games Conjecture is first introduced by Khot [Kho02b] to prove hardness of MIN 2CNF DELETION and other problems. Since then, it has been used to prove various hardness results that LABEL COVER-based techniques could not. It includes VERTEX COVER [KR08], MAX CUT [KKMO07], and every MAX CSP [Rag08]. We refer the reader to the survey of Khot [Kho10] and Trevisan [Tre12] for overview of these results. We formally introduce the UNIQUE GAMES and Unique Games Conjecture.

Definition 3.2.1. *An instance $\mathcal{L}(B(U \cup W, E), [R], \{\pi(u,w)\}_{(u,w) \in E})$ of* UNIQUE GAMES *consists of a biregular bipartite graph $B(U \cup W, E)$ and a set $[R]$ of labels. For each edge $(u, w) \in E$ there is a constraint specified by a permutation $\pi(u, w) : [R] \to [R]$.*

[1]The basic 2-prover game based on 3SAT does not make the projections d-regular, but a simple trick allows us to assume this without loss of generality. See Theorem 1.17 of Wenner [Wen13] for the formal proof.

Given a labeling $l : U \cup W \to [R]$, an edge $e = (u, w)$ is said to be satisfied if $l(u) = \pi(u, w)(l(w))$. Let

$$\mathsf{Val}_{\mathsf{UG}}(l) := \frac{1}{|E|} \cdot |\{e \in E : l \text{ satisfies } e\}|,$$

and

$$\mathsf{Opt}_{\mathsf{UG}}(\mathcal{L}) := \max_{l:U\cup W\to[R]} \mathsf{Val}_{\mathsf{UG}}(l).$$

Conjecture 3.2.2 (The Unique Games Conjecture [Kho02b])**.** *For any constants $\eta > 0$, there is $R = R(\eta)$ such that, for a* UNIQUE GAMES *instance $\mathcal{L}$ with label set $[R]$, it is NP-hard to distinguish between*

- $\mathsf{Opt}_{\mathsf{UG}}(\mathcal{L}) \geq 1 - \eta$.
- $\mathsf{Opt}_{\mathsf{UG}}(\mathcal{L}) \leq \eta$.

We call a computational task *UG-hard* if it is NP-hard assuming the Unique Games Conjecture. To show the optimal hardness result for VERTEX COVER, Khot and Regev [KR08] introduced the following seemingly stronger conjecture, and proved that it is in fact equivalent to the original Unique Games Conjecture.

Conjecture 3.2.3 (Khot and Regev [KR08])**.** *For any constants $\eta > 0$, there is $R = R(\eta)$ such that, for a* UNIQUE GAMES *instance $\mathcal{L}$ with label set $[R]$, it is NP-hard to distinguish between*

- *There is a set $W' \subseteq W$ such that $|W'| \geq (1-\eta)|W|$ and a labeling $l : U \cup W \to [R]$ that satisfies every edge (u, w) for $u \in U$ and $w \in W'$.*
- $\mathsf{Opt}_{\mathsf{UG}}(\mathcal{L}) \leq \eta$.

For RMFC, we use the following variant of UNIQUE GAMES, which is not known to be equivalent to the original conjecture.

Conjecture 3.2.4. *For any constants $\eta > 0$, there is $R = R(\eta)$ such that, for a* UNIQUE GAMES *instance $\mathcal{L}$ with label set $[R]$, it is NP-hard to distinguish between*

- *There is a set $W' \subseteq W$ such that $|W'| \geq (1-\eta)|W|$ and a labeling $l : U \cup W \to [R]$ that satisfies every edge (u, w) for $u \in U$ and $w \in W'$.*

- $\mathsf{Opt}_{\mathsf{UG}}(\mathcal{L}) \leq \eta$. *Moreover, the instance satisfies the following expansion property: For every set* $S \subseteq W$, $|S| = \frac{|W|}{10}$, *we have* $|N(S)| \geq \frac{9}{10}|U|$, *where* $N(S) := \{u \in U : \exists w \in S, (u, w) \in E\}$.

Conjecture 3.2.4 is similar to that of Bansal and Khot [BK09], under which the optimal hardness of *Minimizing Weighted Completion Time with Precedence Constraints* is proved. Their conjecture requires that in the soundness case, $\forall S \subseteq W$ with $|S| = \delta|W|$, we must have $|N(S)| \geq (1 - \delta)|U|$ for arbitrarily small δ. Our conjecture is a weaker (so more likely to hold) since we require this condition for only one value $\delta = \frac{1}{10}$.

3.3 Fourier Analysis

Whether Unique Games Conjecture is true or not, one of the greatest influences of Unique Games Conjecture is to spur extensive use of discrete Fourier analysis in hardness of approximation. As Khot [Kho10] wrote in his survey, applications to hardness of approximation actually led to some new Fourier analytic theorems. Furthermore, some of these advanced tools developed primarily for UNIQUE GAMES were used with LABEL COVER to prove NP-hardness results. We refer the reader to the textbook of O'Donnell [O'D14] for general introduction to discrete Fourier analysis and its applications on hardness of approximation. Throughout this thesis, tools from discrete Fourier analysis are used extensively in the following places.

- UG hardness: Chapter 4 for HARD CSP, Chapter 7 for GRAPH PRICING, Part V for cut problems.
- NP hardness: Part III for coloring problems.
- SDP-based algorithm: Chapter 5 for SYMMETRIC CSP.

In this section, we present some of basic tools of discrete Fourier analysis. More advanced techniques used in a specific part of this thesis, mostly those for NP-hardness of coloring in Part III, will be presented in the respective sections.

First we introduce standard tools on correlated probability spaces from Mossel [Mos10]. Fix a finite set Ω. Given a probability space (Ω, μ) (we always consider finite probability spaces), let $\mathcal{L}(\Omega)$ be the set of functions $\{f : \Omega \to \mathbb{R}\}$ and for an interval $I \subseteq \mathbb{R}$, $\mathcal{L}_I(\Omega)$ be the set of functions $\{f : \Omega \to I\}$. For a subset $S \subseteq \Omega$, define *measure* of S to be $\mu(S) := \sum_{\omega \in S} \mu(\omega)$. A collection of probability spaces are said to be correlated if there is a joint probability distribution on them. We will denote k correlated spaces $\Omega_1, \ldots, \Omega_k$ with a joint distribution μ as $(\Omega_1 \times \cdots \times \Omega_k, \mu)$.

Definition 3.3.1. *Given two correlated spaces $(\Omega_1 \times \Omega_2, \mu)$, we define the correlation between Ω_1 and Ω_2 by*

$$\rho(\Omega_1, \Omega_2; \mu) := \sup \left\{ \mathsf{Cov}[f, g] : f \in \mathcal{L}(\Omega_1), g \in \mathcal{L}(\Omega_2), \mathsf{Var}[f] = \mathsf{Var}[g] = 1 \right\}.$$

Definition 3.3.2. *Given a probability space (Ω, μ) and a function $f \in \mathcal{L}(\Omega)$ and $p \in \mathbb{R}^+$, let $\|f\|_p := \mathbb{E}_{x \sim \mu}[|f(x)|^p]^{1/p}$.*

We use the following lemma to bound the correlation $\rho(\Omega_1, \Omega_2; \nu)$. While the quantitative guarantee of this lemma is often not optimal, the fact that $\rho < 1$ suffices for our purpose for various covering and cut problems. In Section 7.4 for GRAPH PRICING and Section 17.6 for SHORT PATH EDGE CUT, we use more direct methods to prove better bounds on ρ.

Lemma 3.3.3 (Lemma 2.9 of [Mos10])**.** *Let $(\Omega_1 \times \Omega_2, \mu)$ be two correlated spaces such that the probability of the smallest atom in $\Omega_1 \times \Omega_2$ is at least $\alpha > 0$. Define a bipartite graph $G = (\Omega_1 \cup \Omega_2, E)$ where $(a, b) \in \Omega_1 \times \Omega_2$ satisfies $(a, b) \in E$ if $\mu(a, b) > 0$. If G is connected, then $\rho(\Omega_1, \Omega_2; \mu) \leq 1 - \frac{\alpha^2}{2}$.*

At the heart of discrete Fourier analysis is the following decomposition of any function on a product space.

Definition 3.3.4. *Consider a product space $(\Omega^R, \mu^{\otimes R})$ and $f \in \mathcal{L}(\Omega^R)$. The* Efron-Stein decomposition *of f is given by*

$$f(x_1, \ldots, x_R) = \sum_{S \subseteq [R]} f_S(x_S)$$

where (1) f_S depends only on x_S and (2) for all $S \not\subseteq S'$ and all $x_{S'}$, $\mathbb{E}_{x' \sim \mu^{\otimes R}}[f_S(x') | x'_{S'} = x_{S'}] = 0$.

This decomposition allows us to understand many natural combinatorial quantities using tools from analysis. The *influence* is a such quantity.

Definition 3.3.5. *The* influence *of the ith coordinate on f is defined by*

$$\mathsf{Inf}_i[f] := \mathbb{E}_{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_R}[\mathsf{Var}_{x_i}[f(x_1, \ldots, x_R)]].$$

The influence has a convenient expression in terms of the Efron-Stein decomposition.

$$\mathsf{Inf}_i[f] = \| \sum_{S : i \in S} f_S \|_2^2 = \sum_{S : i \in S} \|f_S\|_2^2.$$

As the name suggests, the influence measures how much value of f depends on the ith coordinate. Let $\Omega = \{+1, -1\}$ and $\mu(+1) = \mu(-1) = 1/2$, and consider the following three functions from $\Omega^R \to \{-1, 1\}$ for some odd number R.

1. Dictator: Fix $i \in [R]$, and let $f_D(x_1, \dots, x_R) = x_i$.
   - $\mathsf{Inf}_i[f_D] = 1$ and $\mathsf{Inf}_j[f_D] = 0$ for all $j \neq i$.
2. Majority: $f_M(x_1, \dots, x_R) = \mathsf{sign}(x_1 + \cdots + x_R)$.
   - For every $i \in [R]$, $\mathsf{Inf}_i[f_M]$ is the probability that $R - 1$ independently flipped coins show exactly $\frac{R-1}{2}$ heads and tails respectively, so $\mathsf{Inf}_i[f_M] = \Theta(1/\sqrt{R})$.
3. XOR: $f_X(x_1, \dots, x_R) = \prod_{i=1}^{R} x_i$.
   - $\mathsf{Inf}_i[f_X] = 1$ for every $i \in [R]$.

Among these functions, we can clearly say that i is the only influential coordinate in f_D, and f_M has no single influential coordinate. The situation becomes unclear for f_X since every coordinate has the highest possible influence. It turns out that having too many coordinates of high influence causes a technical problem for applications in hardness of approximation. To fix it, we introduce two variants of influence.

Definition 3.3.6. *Let $\rho \in [-1, 1]$. The* noisy operator *T_ρ acts on f by*

$$T_\rho f(x) = \mathbb{E}_y[f(y)] \,,$$

where each coordinate y_i of y is independently sampled such that it is equal to x_i with probability ρ and newly sampled from μ with probability $(1 - \rho)$. We call $\mathsf{Inf}_i[T_\rho f]$ the noisy influence *of the ith coordinate. Equivalently it can be defined as*

$$\mathsf{Inf}_i[T_\rho f] = \sum_{S: i \in S} \rho^{|S|} \|f_S\|_2^2.$$

Definition 3.3.7. *We also define the* low-degree influence *of the ith coordinate.*

$$\mathsf{Inf}_i^{\leq d}[f] := \sum_{S: i \in S, |S| \leq d} \|f_S\|_2^2.$$

Since the Efron-Stein decomposition of f_X is $f_X = (f_X)_{[R]}$, we can see that $\mathsf{Inf}_i[T_\rho f_X] = \rho^R$ and $\mathsf{Inf}_i^{\leq d}[f_X] = 0$ for $d < R$. The following lemma formally shows that both notions significantly reduce the number of influential coordinates.

Lemma 3.3.8 ([O'D14]). $\sum_i \mathsf{Inf}_i^{\leq d}[f] \leq d \cdot \mathsf{Var}[f]$ *and* $\sum_i \mathsf{Inf}_i[T_{1-\delta} f] \leq (1/\delta) \cdot \mathsf{Var}[f]$.

The final crucial tool of discrete Fourier analysis for hardness of approximation is the following *invariance principle*. It states that if two functions $f, g : \Omega^R \to \mathbb{R}$ do not share an (low-degree) influential coordinate, we can treat f and g as functions on Gaussian spaces with the same expected values and deduce upper and lower bounds on $\mathbb{E}[fg]$.

Definition 3.3.9. *For* $a, b \in [0, 1]$ *and* $\rho \in (0, 1)$*, let*

$$\underline{\Gamma}_\rho(a, b) := \Pr[X \leq \Phi^{-1}(a), Y \geq \Phi^{-1}(1 - b)],$$
$$\overline{\Gamma}_\rho(a, b) := \Pr[X \leq \Phi^{-1}(a), Y \leq \Phi^{-1}(b)],$$

where X *and* Y *are* ρ*-correlated standard Gaussian variables and* Φ *denotes the cumulative distribution function of a standard Gaussian.*

Theorem 3.3.10 (Lemma 6.8 of [Mos10]). *Let* $(\Omega_1 \times \Omega_2, \mu)$ *be correlated spaces such that the minimum nonzero probability of any atom in* $\Omega_1 \times \Omega_2$ *is at least* α *and such that* $\rho(\Omega_1, \Omega_2; \mu) \leq \rho < 1$. *Then for every* $\epsilon > 0$ *there exist* τ, d *depending only on* ρ, ϵ, α *such that if* $f : \Omega_1^R \to [0, 1], g : \Omega_2^R \to [0, 1]$ *satisfy* $\min(\mathsf{Inf}_i^{\leq d}[f], \mathsf{Inf}_i^{\leq d}[g]) \leq \tau$ *for all* i*, then*

$$\underline{\Gamma}_\rho(\mathbb{E}_x[f], \mathbb{E}_y[g]) - \epsilon \leq \mathbb{E}_{(x,y)\in\mu^{\otimes R}}[f(x)g(y)] \leq \overline{\Gamma}_\rho(\mathbb{E}_x[f], \mathbb{E}_y[g]) + \epsilon.$$

Some of our results only use the fact that $\underline{\Gamma}_\rho(\mathbb{E}_x[f], \mathbb{E}_y[g]) > 0$ and take $\epsilon = \underline{\Gamma}_\rho(\mathbb{E}_x[f], \mathbb{E}_y[g])/2$ to show that the expected value is strictly positive, while others use additional quantitative bounds in Gaussian spaces. Section 5.2.2 for SYMMETRIC CSP, Section 7.7 for GRAPH PRICING, Section 17.6 for cut problems contain our technical results on Gaussian space.

The tools introduced in this section suffice for most UG-hardness results in this thesis (except more direct calculations for bounding the correlation ρ). To adapt them for LABEL COVER to prove NP-hardness, we need more advanced tools since the set of R coordinates have an additional structure — there is an equivalence relation on $[R]$ and coordinates in the same equivalence class (we call *block*) are more correlated than those outside. See Section 10.2 for more Fourier analysis background for this purpose.

3.3.1 Dictatorship Tests for VERTEX COVER

As an illustration of how these tools from discrete Fourier analysis are used for hardness of approximation, we present a *dictatorship test* for VERTEX COVER. A dictatorship test for a combinatorial problem P is an instance of the problem where its set of vertices (if P is a graph or hypergraph problem) or its set of variables (if P is a CSP) is Ω^R for some R. Then any solution to this problem can be represented as a function $f : \Omega^R \to Q$. For CSPs, Q denotes the domain from which each variable can take a value, and for covering or cut problems, $Q = \{0, 1\}$ where 1 indicates that the vertex will be removed. It is required to have two properties.

- Completeness: For each $i \in [R]$, there is a dictator function $f_D(x_1, \ldots, x_R)$ that only depends on x_i and corresponds to a good solution P.

- Soundness: Any $f : \Omega^R \to Q$ that does not reveal an (low-degree or noisy) influential coordinate will not be a good solution to P. Equivalently, any f that corresponds to a good solution will reveal a corodinate with large (low-degree or noisy) influential coordinate.

It is also known as the *long code* and was first introduced by Bellare, Goldreich, and Sudan [BGS98], and has played a crucial role in proving both NP-hardness and UG-hardness of approximation.

Now we present the dictatorship test for VERTEX COVER constructed by Bansal and Khot [BK10]. We do not present the full reduction from UNIQUE GAMES here, but similar reductions for our new results of this thesis will be presented in the respective sections. This dictatorship test can be used to show that VERTEX COVER is UG-hard to approximate within a factor $2 - \epsilon$ for any $\epsilon > 0$. Khot and Regev [KR08] first proved this result, and Austrin et al. [AKS09] and Bansal and Khot [BK09, BK10] proved stronger versions of this results using more advanced tools from Fourier analysis.

Let $\Omega := \{*, 0, 1\}$, and fix $\epsilon > 0$ and $R \in \mathbb{N}$. Our dictatorship test $\mathcal{D}^{\mathsf{vc}}_{R,\epsilon} = (\Omega^R, E)$ is defined as follows. Each vertex is represented by v_x where $x \in \Omega^R$ is a R-dimensional vector. Consider the probability space (Ω, μ) where $\Omega := \{0, 1, *\}$, and $\mu : \Omega \mapsto [0, 1]$ with $\mu(*) = \epsilon$ and $\mu(x) = (1-\epsilon)/2$ for $x \neq *$. We define the weight $\mathsf{wt}(v_x) := \mu^{\otimes R}(x) = \prod_{i=1}^{R} \mu(x_i)$. The sum of weights is 1. Two vertices v_x and v_y have an edge if and only if for any $1 \leq l \leq R$, $[x_l \neq y_l]$ or $[y_l = *]$ or $[x_l = *]$. We now prove the two desired properties.

Completeness. Fix $q \in [R]$ and let $U_q := \{v_x : x_q = 0 \text{ or } *\}$. The weight of U_q is $\mathsf{wt}(U_q) = (1 + \epsilon)/2$. Note that U_q only depends on x_q.

Lemma 3.3.11. *U_q is a vertex cover.*

Proof. Let $\{v_x, v_y\}$ be an edge of $\mathcal{D}^{\mathsf{vc}}_{R,\epsilon}$. If both endpoints do not belong to U_q, it implies $x_q = y_q = 1$. It contradicts our construction. $\square$

Soundness. To analyze soundness, we define a correlated probability space $(\Omega_1 \times \Omega_2, \nu)$ where both Ω_1, Ω_2 are copies of Ω. It is defined by the following process to sample $(x, y) \in \Omega^2$.

1. Sample $x \in \{0, 1\}$ uniformly at random. Let $y = 1 - x$.

2. Change x to $*$ with probability ϵ. Do the same for y independently.

We note that the marginal distribution of both x and y is equal to μ. Assuming $\epsilon < 1/3$, the minimum probability of any atom in $\Omega_1 \times \Omega_2$ is ϵ^2. Consider the bipartite graph on $\Omega_1 \cup \Omega_2$ such that $x \in \Omega_1$ and $y \in \Omega_2$ are connected if and only if (x, y) is sampled with nonzero probability. For any $x \in \Omega_1$, note that $(x, *)$ is a valid edge, and so is $(*, *)$, so x is connected to the edge $(*, *)$. The same argument shows that any $y \in \Omega_2$ is connected to $(*, *)$, so this bipartite graph is indeed connected. Therefore, we can apply Lemma 3.3.3 to have $\rho(\Omega_1, \Omega_2; \nu) \leq \rho := 1 - \epsilon^4/2$.

Fix an arbitrary vertex cover $U \subseteq V$ of $\mathsf{wt}(U) \leq 1 - \epsilon$. Let $f : \Omega^R \mapsto \{0, 1\}$ be the indicator function of U so that $\mathbb{E}_{x \in \mu^{\otimes R}}[f(x)] = \mu^{\otimes R}(U) \leq 1 - \epsilon$. Theorem 3.3.10 ($\rho \leftarrow \rho, \alpha \leftarrow \epsilon^2, \epsilon \leftarrow \underline{\Gamma}_\rho(\epsilon, \epsilon)/2$) states that there exist τ and d such that if $\mathsf{Inf}_i^{\leq d}[f] \leq \tau$ for all $i \in [R]$,

$$\mathbb{E}_{(x,y) \sim \nu^{\otimes R}}[(1-f)(x) \cdot (1-f)(y)] \geq \underline{\Gamma}_\rho(1-\mathbb{E}[f], 1-\mathbb{E}[f]) - \underline{\Gamma}_\rho(\epsilon, \epsilon)/2 \geq \underline{\Gamma}_\rho(\epsilon, \epsilon)/2 > 0.$$

This implies that there exists x, y such that there is an edge between v_x and v_y but neither v_x nor v_y is contained in U. This contradicts that U is a vertex cover.

Therefore, if $\mathsf{Inf}_i^{\leq d}[U] \leq \tau$ for all i (i.e., U does not reveal any influential coordinate), then $\mathsf{wt}(U) \geq 1 - \epsilon$. Note that by Lemma 3.3.8, the number of coordinates i with $\mathsf{Inf}_i^{\leq d}[f] \geq \tau$ is at most $\frac{d}{\tau}$, and that d and τ only depend on ϵ, but not R. This fact is crucial for the reduction from UNIQUE GAMES since we would like to keep the number of coordinates with large low-degree influence small regardless of the total number of coordinates R. See the respective reductions for details.

In summary, in the completeness case, there is a dictator function corresponding to a vertex cover of weight $(1 + \epsilon)/2$. Indeed, after deleting vertices of total weight at most ϵ, the graph becomes a bipartite graph. In the soundness case, unless we reveal an (low-degree) influential coordinate, every vertex cover has weight at least $(1 - \epsilon)$. The gap between the two cases approaches to 2 as $\epsilon \to 0$.

3.4 LP/SDP and Integrality Gaps

Linear programming (LP) concerns the problem of maximizing or minimizing a linear function over a polyhedron. It is expressed as

$$\begin{array}{ll} \text{maximize} & c^T x \\ \text{subject to} & Ax \leq b \\ & x \geq 0, \end{array}$$

where $x \in \mathbb{R}^n$ is a variable, and $c \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ are fixed constants.

Semidefinite programming (SDP) concerns the problem of maximizing or minimizing a linear function over the intersection of a positive semidefinite cone and an affine space. It is expressed as

$$\begin{array}{ll} \text{maximize} & c^T x \\ \text{subject to} & \sum_{i=1}^{n} x_i A_i \preceq B \\ & x \geq 0, \end{array}$$

where $x \in \mathbb{R}^n$ is a variable, and $c \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{m \times m}, C \in \mathbb{R}^{m \times m}$ are fixed constants. Note that $A \preceq B$ if and only if $y^T A y \leq y^T B y$ for all $y \in \mathbb{R}^m$.

Given a maximization problem P, we often consider a natural LP or SDP relaxation R. The optimum of such relaxation Opt_R is at least as big as the integral optimum Opt_P. The *integrality gap* of the relaxation R is defined to be the infimum of $\frac{\mathsf{Opt}_P}{\mathsf{Opt}_R}$ over every instance of P. For the minimization problem, it is defined to be the supremum of $\frac{\mathsf{Opt}_P}{\mathsf{Opt}_R}$.

All algorithmic results in this thesis rely on LP or SDP relaxations. Table 3.1 summarizes them. Among them, our algorithms for BALANCE SAT, GENERALIZED MAX DICUT, k-STAR TRANSVERSAL, and k-VERTEX SEPARATOR also prove that the integrality gaps for these problems are small. The algorithms for SYMMETRIC CSP, MAX 2-COLORING, MIN COLORING are based special relaxations designed for certain *promises* on input instances, so they do not necessarily prove that the integrality gaps are small. For k-PATH TRANSVERSAL, the algorithm starts from solving a LP relaxation that is known to have a large integrality gap, but *bypasses* this integrality gap to achieve an approximation ratio much better than the integrality gap while still using the optimal LP solution.

Problem	Type	Relaxation	Location
BALANCE SAT	CSP	LP	Section 4.3.2
SYMMETRIC CSP	CSP	SDP	Chapter 5
GENERALIZED MAX DICUT	CSP	LP	Chapter 7.6
MAX 2-COLORING	Coloring	SDP	Chapter 11
MIN COLORING	Coloring	SDP	Chapter 11
k-STAR TRANSVERSAL	Covering	LP	Chapter 14
k-VERTEX SEPARATOR	Covering	LP	Chapter 15
k-PATH TRANSVERSAL	Covering	LP	Chapter 15

Table 3.1: Summary of our algorithmic results in this thesis.

3.4.1 Relaxations for CSP and Hierarchies

One of the most natural LP hierarchies is the following *Sherali-Adams* hierarchy [SA90]. Let $(V, \mathcal{C})$ be an instance of CSP where each variable $v \in V$ can take a value in finite set Q and each constraint $C \in \mathcal{C}$ depends on at most k variables.

Let r be a positive integer. A r-local distribution is defined as the collection of variables $\{x_S(\alpha)\}$ where S ranges over a subset of V of size at most r, and $\alpha : S \to Q$ is a partial assignment to variables in S. As the name suggests, it is required to satisfy the following three constraints. For $\alpha : S \to Q$ and $\beta : T \setminus S \to Q$, let $\alpha \circ \beta : T \to Q$ be the assignment consistent with α and β.

- $x_\emptyset = 1$.
- $x_S(\alpha) \geq 0$ for all S, α.
- $\sum_{\beta : T \setminus S \to Q} x_T(\alpha \circ \beta) = x_S(\alpha)$ for all $S \subseteq T$, $\alpha : S \to Q$.

If $r \geq t$, for each constraint $C \in \mathcal{C}$ that depends on variables in S, the probability that C is satisfied by this local distribution is

$$y_C := \sum_{\alpha : S \to Q \text{ and } \alpha \text{ satisfies } C} x_S(\alpha).$$

Therefore, the r-rounds of Sherali-Adams relaxation for this CSP is

$$\begin{array}{lll} \text{maximize} & \sum_{C\in\mathcal{C}} y_C & \\ \text{subject to} & y_C = \sum_{\alpha:S\to Q \text{ and } \alpha \text{ satisfies } C} x_S(\alpha) & \forall C \in \mathcal{C}. \\ & \{x_{S,\alpha}\} \text{ is a } r\text{-local distribution.} & \end{array}$$

If $r < t$, y_C cannot be defined naturally, but for BALANCE SAT in Section 4.3.2, we have a natural way to define y_C to have a nontrivial approximation algorithm when $r = 1$ and t is not bounded.

We can further strengthen this Sherali-Adams hierarchy to *Lasserre* or *Sum-of-Squares* hierarchies introduced in [Las01, Par00]. The r-rounds of Sum-of-Squares hierarchy involve a $2r$-local distribution $\{x_S(\alpha)\}$ as well as a set of vectors $\{v_S(\alpha)\}$ for every $|S| \le r$ and $\alpha : S \to Q$ with the following property: for any $S, T \subseteq V$ with $|S|, |T| \le t$ and any $\alpha : S \to Q$ and $\beta : T \to Q$, we have

$$\langle v_S(\alpha), v_T(\beta)\rangle = x_{S\cup T}(\alpha \circ \beta)$$

if α and β are consistent on $S \cap T$, and $\langle v_S(\alpha), v_T(\beta)\rangle = 0$ otherwise.

Sherali-Adams and Sum-of-Squares hierarchies have been actively studied recently. See the survey of Chlamtac and Tulsiani [CT12] and Barak and Steurer [BS14]. For GRAPH PRICING (and GENERALIZED MAX DICUT) and decoding LDPC codes, we present strong lower bounds on the integrality gaps of these hierarchies. Table 3.2 summarizes them.

Problem	Relaxation	Location
GRAPH PRICING and GENERALIZED MAX DICUT	Sherali-Adams	Section 7.5
Decoding LDPC codes	Sherali-Adams and Sum-of-Squares	Chapter 8

Table 3.2: Summary of our integrality gap results in this thesis.

3.4.2 Relaxations for Covering

Let $(V, \mathcal{S})$ be a set system where $\mathcal{S} = \{S_1, \dots, S_m\}$ and each S_i is a subset of V. Consider the covering problem where we want to find the smallest set $U \subseteq V$ such that $U \cap S_i \ne \emptyset$

for every i. This problem can be viewed as a variant of CSP where each element $v \in V$ becomes a variable that can have either 1 (meaning $v \in U$) or 0 (meaning $v \notin U$), and each set S_i becomes a constraint that at least one variable $v \in S_i$ has to be assigned 1. Let $k := \max_i |S_i|$. Whenever $r \geq k$, the following Sherali-Adams hierarchy gives a relaxation.

$$\begin{array}{lll} \text{minimize} & \sum_{v \in V} x_v(1) & \\ \text{subject to} & x_{S_i}(0, \ldots, 0) = 0 & \forall S_i \in \mathcal{S}. \\ & \{x_{S,\alpha}\} \text{ is a } r\text{-local distribution.} & \end{array}$$

Our algorithms for k-Star Transversal and k-Vertex Separator are inspired by the above relaxation (r = maximum degree for k-Star Transversal and $r = O(k)$ for k-Vertex Separator), but the final relaxations we use have at most $O(n^2)$ variables in total (the one for k-Vertex Separator has an exponential number of constraints and we construct a good separation oracle). When $r = 1$, we have the following basic relaxation where variables are $\{x_v\}_{v \in V}$.

$$\begin{array}{lll} \text{minimize} & \sum_{v \in V} x_v & \\ \text{subject to} & \sum_{v \in S_i} x_v \geq 1 & \forall S_i \in \mathcal{S}. \\ & x \geq 0. & \end{array}$$

If every S_i has at least k elements, setting $x_v = \frac{1}{k}$ for all $v \in V$ is a feasible LP solution. This LP has a large integrality gap $(k - o(1))$ for k-Path Transversal, but our $O(\log k)$-approximation algorithm for k-Path Transversal starts solving the above LP, bypassing its integrality gap while still using its solution.

3.4.3 Relaxations for Coloring

The study of (hyper)graph coloring is one of the first places where connections between SDP and combinatorial optimization were made. Given a graph $G = (V, E)$, Lovász [Lov79] defined the *Lovász theta function* $\theta(G)$, which is lower bounded by the size of the maximum independent set of G and upper bounded by the chromatic number of tis complement graph $\overline{G} = (V, \overline{E})$ where $\overline{E} = \binom{V}{2} \setminus E$. As a relaxation of Maximum Independent Set, $\theta(G)$ can be defined as the optimal value of the following SDP. The only variable is an $n \times n$ matrix X. Let J be the matrix of all ones.

$$\begin{array}{ll} \text{maximize} & \langle J, X \rangle \\ \text{subject to} & \mathsf{Tr}(X) = 1 \\ & X(i,j) = 0 \qquad \forall (i,j) \in E \\ & X \succeq 0 \end{array}$$

This is a relaxation of MAXIMUM INDEPENDENT SET, since if $S \subseteq V$ is an independent set, letting $\mathbf{1}_S$ be the indicator vector of S and $X = \mathbf{1}_S \mathbf{1}_S^T / |S|$ certifies that $\theta(G) \geq |S|$. Its dual is the following SDP.

$$\begin{array}{ll} \text{minimize} & \alpha \\ \text{subject to} & \alpha I + B \succeq J \\ & B(i,j) = 0 \qquad \forall (i,j) \notin E \end{array}$$

which is equivalent to the following vector program by considering vectors $b_1, \ldots, b_n$ such that $(\alpha I + B - J)(i,j) = \langle b_i, b_j \rangle$.

$$\begin{array}{ll} \text{minimize} & \alpha \\ \text{subject to} & ||b_i||^2 = \alpha - 1 \\ & \langle b_i, b_j \rangle = -1 \qquad \forall i \neq j \text{ and } (i,j) \neq E \end{array}$$

Note that the optimal value of this program is always at least 1. Let $c_i := b_i/(\alpha - 1)$, and let the new objective function be $t := -1/\sqrt{\alpha - 1}$. We were minimizing $\alpha \geq 1$, so we are also minimizing t. Since $\theta(G)$ is at most the chromatic number of $\overline{G}$, we can conclude that the following program is a relaxation of the chromatic number of $\overline{G}$ in the sense that if $\overline{G}$ can be colored by χ colors, the optimal value of the program is at most $-\frac{1}{\chi-1}$.

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & ||c_i||^2 = 1 \\ & \langle c_i, c_j \rangle = t \qquad \forall (i,j) \in \overline{E} \end{array}$$

Consider a regular $(\chi - 1)$-simplex centered at the origin where each vertex is at distance 1 from the center. Let $u_1, \ldots, u_\chi$ be the vertices. It is simple to check that $\langle u_i, u_j \rangle = -1/(\chi - 1)$. The geometric intuition of this program is that if $\overline{G}$ is χ-colorable, assigning one of u_i to each vertex according to its color certifies that $t \leq -1/(\chi - 1)$. There are other natural equivalent formulations of the θ function. We refer the reader to the lecture notes of Todd [Tod12] and the survey of Knuth [Knu94] for those formulations and their equivalences.

In this thesis, we study the problem of hypergraph coloring under strong proimses. They allow us to write strong and intuitive SDPs and guarantee its feasibility. These SDPs will still assign a unit vector to each vertex. Given a hypergraph $H = (V, E)$, the three promises we study are the following.

- A hypergraph is said to have *discrepancy* ℓ when there is a 2-coloring such that in each hyperedge, the difference between the number of vertices of each color is at most ℓ.
- A χ-coloring ($\chi \le K$) is called *rainbow* if every hyperedge contains each color at least once.
- A χ-coloring ($\chi \ge K$) is called *strong* if every hyperedge contains K different colors.

Our SDP relaxations for low-discrepancy, rainbow-colorability, and strong-colorability are the following.

Discrepancy ℓ. This program is feasible by assigning a unit vector w or $-w$ to each vertex according to the 2-coloring minimizing the discrepancy.

$$\begin{aligned} &\left|\left|\sum_{i\in e} u_i\right|\right|_2 \le \ell && \forall\, e \in E \\ &||u_i||_2 = 1 && \forall\, i \in [n] \\ &u_i \in \mathbb{R}^n && \forall\, i \in [n] \end{aligned}$$

$(K - \ell)$**-Rainbow Colorability.** This program is feasible by assigning a vertex of a regular $(K - \ell - 1)$ simplex to each vertex according to a rainbow coloring.

$$\begin{aligned} &\left|\left|\sum_{i\in e} u_i\right|\right|_2 \le \ell && \forall\, e \in E \\ &\langle u_i, u_j\rangle \ge \frac{-1}{K-\ell-1} && \forall\, e \in E,\ \forall\, i < j \in e \\ &||u_i||_2 = 1 && \forall\, i \in [n] \\ &u_i \in \mathbb{R}^n && \forall\, i \in [n] \end{aligned}$$

$(K+\ell)$**-Strong Colorability.** This program is feasible by assigning a vertex of a regular $(K+\ell-1)$ simplex to each vertex according to a strong coloring.

$$\begin{aligned}
&\langle u_i, u_j \rangle = \frac{-1}{K+\ell-1} && \forall\, e \in E,\ \forall\, i < j \in e \\
&||u_i||_2 = 1 && \forall\, i \in [n] \\
&u_i \in \mathbb{R}^n && \forall\, i \in [n]
\end{aligned}$$

Part I

Constraint Satisfaction Problems

Chapter 4

Balance / Hard CSP

4.1 Introduction

The study of the complexity of Constraint Satisfaction Problems (CSPs) has seen much progress, with beautiful well-developed theories explaining when they admit efficient satisfiability and approximation algorithms. A CSP is specified by a finite set Π of relations (relations can have different arities) over some finite domain Q. An instance of a MAX CSP(Π) consists of a set of variables $X = \{x_1, ..., x_n\}$ and a collection of constraints $C = \{C_1, ..., C_m\}$ each of which is a relation from Π applied to some tuple of variables from X. Constraints are weighted and we assume that $\sum_i \mathsf{wt}(C_i) = 1$. For any assignment $\sigma : X \to Q$, $\mathsf{Val}(\sigma)$ is the total weight of satisfied constraints by σ, and our goal is to find σ that maximizes $\mathsf{Val}(\sigma)$. We consider two natural extensions of MAX CSP(Π).

Definition 4.1.1 (CSP with balance constraints)**.** *An instance of* BALANCE CSP*(Π) over domain Q, $\mathcal{I} = (X, C)$ consists of set of variables X and a collection of constraints C, as in* MAX CSP*(Π). An assignment $\sigma : X \to Q$ is called balanced if for each $q \in Q$, $|\sigma^{-1}(q)| = \frac{n}{|Q|}$. Define* $\mathsf{Val}_B(\sigma) = \mathsf{Val}(\sigma)$ *if σ is balanced, and* $\mathsf{Val}_B(\sigma) = 0$ *otherwise. Let* $\mathsf{Opt}_B(\mathcal{I}) = \max_\sigma \mathsf{Val}_B(\sigma)$*. Our goal is to find σ that maximizes* $\mathsf{Val}_B(\sigma)$.

The notion of BALANCE CSP is interesting both practically and theoretically. Partitioning a set of objects into equal-sized subsets with desired properties is a basic scheme used in Divide-and-Conquer algorithms. BALANCE CUT, also known as MAXIMUM BISECTION, is one of the most well-known examples of BALANCE CSP. Theoretically, the balance constraint is one of the simplest *non-local* constraints where the current algorithmic and hardness results on ordinary CSPs do not work.

Definition 4.1.2 (CSP with hard constraints)**.** *An instance of* HARD CSP*(*Π*)* $\mathcal{I} = (X, S, H)$ *consists of set of variables* X *and a two collections of constraints* $S = \{S_1, ..., S_{m_S}\}$ *and* $H = \{H_1, ..., H_{m_H}\}$ *(*S *stands for* soft *constraints and* H *stands for* hard *constraints, and only soft constraints are weighted). An assignment* $\sigma : X \to Q$ *is* feasible *if it satisfies all constraints in* H*. Let* $\mathsf{Val}_H(\sigma)$ *be the total weight of satisfied constraints in* S *if* σ *is feasible, and* 0 *otherwise. Let* $\mathsf{Opt}_H(\mathcal{I}) = \max_\sigma \mathsf{Val}_H(\sigma)$*. Our goal is to find* σ *that maximizes* $\mathsf{Val}_H(\sigma)$*.*

HARD CSP contains every MAX CSP by definition, and also several additional fundamental combinatorial optimization problems, such as (Hypergraph) Independent Set, Multicut, Graph-k-Coloring, and many other covering/packing problems. While every assignment is *feasible* in ordinary MAX CSP, in HARD CSP only certain assignments which satisfy all the hard constraints are considered as feasible, giving a more general framework to study combinatorial optimization problems.

By the seminal work of Schaefer [Sch78], there are only three nontrivial classes of Boolean CSPs for which satisfiability can be checked in polynomial time: 2-SAT, Horn-SAT, and LIN-mod-2.[1] Even among them, there is a stark difference in terms of approximability. Håstad [Hås01] showed that for MAX LIN-MOD-2, it is NP-hard to find an assignment that satisfies $(\frac{1}{2} + \epsilon)$ fraction of constraints even when there is an assignment that satisfies $(1 - \epsilon)$ fraction of constraints, for any $\epsilon > 0$. Even for the special case where all linear equations are homogeneous (so that $x_1 = \cdots = x_n = 0$ trivially satisfies all equations), Holmerin and Khot [HK04] showed the same result for its balance version. On the other hand, a series of works [GW94, Zwi98b, CMM07] showed that MAX 2-SAT and MAX HORN-SATadmit a *robust* algorithm, which outputs an assignment satisfying at least $(1 - g(\epsilon))$ fraction of constraints given a $(1 - \epsilon)$-satisfiable instance, where $g(\epsilon) \to 0$ as $\epsilon \to 0$, and $g(0) = 0$. The exact behavior of the function $g(\cdot)$ for MAX LIN-MOD-2 and MAX HORN-3-SAT has been pinned down under the Unique Games conjecture [KKMO07, GZ12]. Generalizing this, all Boolean CSPs were classified with respect to how well they can be robustly approximated (i.e., the behavior of the function $g(\cdot)$) in [DK13]. From our perspective, it is natural to investigate the effects of balance and hard constraints applied to the *most tractable* classes of Boolean CSPs (MAX LIN-MOD-2, MAX HORN-SAT, MAX CUT) and study how hard each variant becomes. This is the task we undertake in this chapter.

Between balance and hard constraints, which one makes the original problem harder?

[1] An instance of Horn-SAT is a set of Horn clauses, each with at most one unnegated literal. An instance of LIN-mod-2 is linear equations mod 2. Dual-Horn-SAT in which clauses have at most one negated literal also admits an efficient satisfiability algorithm, but as it obviously has the same properties as Horn-SAT, we focus on Horn-SAT in this chapter.

They are not directly comparable since each variant inherits different characteristics of the original problem. For BALANCE SAT which includes both BALANCE 2-SAT and BALANCE HORN-SAT, it is easy to find σ with $\mathsf{Val}_B(\sigma) \geq 0.5$ – choose an arbitrary assignment with the same number of 0's and 1's, and try it and its complement. Therefore, BALANCE SAT admits a constant-factor approximation algorithm, but given an instance that admits a balanced satisfying assignment, even for 2-SAT and Horn-SAT, it is not clear how to find such an assignment.

The situation is exactly opposite in the case of HARD 2-SAT or HARD HORN SAT. Given an instance $\mathcal{I}$ that admits a satisfying (both hard and soft constraints) assignment (i.e., $\mathsf{Opt}_H(\mathcal{I}) = 1$), well-known algorithms for 2-SAT or Horn-SAT will find such an assignment. However, for any $\epsilon > 0$, when $\mathsf{Opt}_H(\mathcal{I}) = 1 - \epsilon$, it is not clear how to find σ such that $\mathsf{Val}_H(\sigma) \geq c$ for some absolute constant c, or even $\mathsf{Val}_H(\sigma) \geq \epsilon$. Therefore, adding hard constraints preserves the fact that satisfiability can be checked in polynomial time, but does not preserve a simple constant-factor approximation algorithm.

Though each of balance and hard constraints does not preserve one of the important characteristics of the original CSP in a simple way, there might be a hope that a more sophisticated algorithmic idea gives an efficient algorithm deciding satisfiability for BALANCE CSP, a constant-factor approximation algorithm for HARD CSP, or a robust algorithm for both of them.

4.1.1 Our Results

In this chapter, we prove strong hardness results for these problems, proving APX-hardness of deciding satisfiability of balanced versions and Unique Games-hardness of constant factor approximation algorithms on $(1-\epsilon)$-satisfiable instances of hard versions, of both MAX LIN-MOD-2 and MAX HORN-SAT. The results are formally stated below.

Balanced CSP. For BALANCE 2-SAT and BALANCE HORN-SAT, the Schaefer-like dichotomy due to [CSS10] for the (decision version of) Boolean BALANCE CSP, implies that we cannot efficiently decide whether the given instance is satisfiable or not. This dichotomy was extended to all domains in [BM10], which is somewhat surprising given the status of the dichotomy conjecture for CSPs without any balance/cardinality constraint.

We show the following stronger statement that rules out even a robust satisfiability algorithm for the common special case BALANCE HORN 2-SAT.

Theorem 4.1.1. *There exists an absolute constant $\delta > 0$ such that given an instance $\mathcal{I}$ of* BALANCE HORN 2-SAT *(special case of* BALANCE 2-SAT *and* BALANCE HORN-

SAT*), it is NP-hard to distinguish the following cases.*

- $\mathsf{Opt}_B(\mathcal{I}) = 1$
- $\mathsf{Opt}_B(\mathcal{I}) \leq 1 - \delta$

This result should be contrasted with the fact that a special case of BALANCE 2-SAT, namely BALANCE CUT (MAXIMUM BISECTION), does admit a robust algorithm [GMR+11, RT12, ABG13]. The work [ABG13] also shows that the guaranteed approximation ratio for BALANCE 2-SAT (which is the worst ratio $\frac{\mathsf{Val}_B(\sigma)}{\mathsf{Opt}_B(\mathcal{I})}$ over every instance $\mathcal{I}$ and σ found by the algorithm) is indeed equal to the best known approximation ratio for MAX LIN-MOD-2 [LLZ02], i.e., $\alpha_{LLZ} \approx 0.9401$, so adding the balance constraint does not make the problem harder in this regard.

While several nontrivial approximation algorithms for BALANCE 2-SAT have been studied, as far as we know, no algorithm for BALANCE HORN-SAT or even BALANCE SAT has been suggested in the literature. Other than the trivial 0.5-approximation algorithm given above, we also show that a slight modification to $\frac{3}{4}$-approximation algorithm due to Goemans and Williamson [GW94] gives the algorithm with the same ratio.

Theorem 4.1.2. *For any $\epsilon > 0$, there is a randomized algorithm such that given an instace $\mathcal{I}$ of* BALANCE SAT*, in time $\mathrm{poly}(size(\mathcal{I}), \frac{1}{\epsilon})$, outputs σ with $\mathsf{Val}_B(\sigma) \geq (\frac{3}{4} - \epsilon)\mathsf{Opt}_B(\mathcal{I})$ with constant probability.*

Hard CSP. For HARD 2-SAT, robust algorithms are ruled out in a more radical way, assuming the Unique Games Conjecture.

Theorem 4.1.3. *For any $\epsilon > 0$, given an instance $\mathcal{I}$ of* HARD 2-SAT*, it is UG-hard to distinguish the following cases.*

- $\mathsf{Opt}_H(\mathcal{I}) \geq 1 - \epsilon$
- $\mathsf{Opt}_H(\mathcal{I}) \leq \epsilon$

This result again shows a stark difference between MAX LIN-MOD-2 and MAX CUT since the famous algorithm of Goemans and Williamson [GW95] works well with hard constraints; if vertices u and v must be separated, we require the vectors corresponding to them to be placed in antipodal positions, and any hyperplane rounding separates them. It shows that HARD CUT admits *both* a constant-factor approximation algorithm and a robust algorithm while HARD 2-SAT admits *neither* of them.

For HARD HORN 2-SAT, simple algorithmic and hardness tricks show that finding σ with $\mathsf{Val}_H(\sigma) \geq 1 - 2\epsilon$ given $\mathsf{Opt}_H(\mathcal{I}) = 1 - \epsilon$ is the best possible, assuming the Unique Games Conjecture.

Observation 4.1.3. *There is a polynomial time algorithm that for any $\epsilon > 0$, if the given instance $\mathcal{I}$ of* MAX HORN-2-SAT *satisfies $\mathsf{Opt}_H(\mathcal{I}) = 1 - \epsilon$, finds σ with $\mathsf{Val}_H(\sigma) \geq 1 - 2\epsilon$. Furthermore, for any $\epsilon, \delta > 0$, it is UG-hard to distinguish the following cases.*

- $\mathsf{Opt}_H(\mathcal{I}) \geq 1 - \epsilon - \delta$.
- $\mathsf{Opt}_H(\mathcal{I}) \leq 1 - 2\epsilon + \delta$.

The above algorithmic trick does not work for other robust algorithms, and for HARD HORN 3-SAT (or higher arities), we have the following NP-hardness result.

Theorem 4.1.4. *For any $\epsilon > 0$, given an instance $\mathcal{I}$ of* HARD HORN 3-SAT*, it is NP-hard to distinguish the following cases.*

- $\mathsf{Opt}_H(\mathcal{I}) \geq 1 - \epsilon$
- $\mathsf{Opt}_H(\mathcal{I}) \leq \epsilon$

The complete algorithmic and hardness results are summarized in Table 4.1.

Ordering constraints over larger domain. If we go beyond the Boolean domain, the situation is not as clear; indeed even satisfiabilities of ordinary CSPs are not completely classified yet. We consider a simple and natural CSP of arity two over larger domains, namely MAX CSP(<) over domain $[q] = \{1, 2, ..., q\}$ − every constraint says that one variable must be less than another. This problem can also be understood as a graph-theoretic problem; given a directed graph $G = (V, E)$, delete the minimum number of edges so that the remaining graph has no *walk* of length q (for the formal definition, see Section 4.4). If the domain is unbounded (indeed, $[n]$ is enough), this is exactly the well-known MAXIMUM ACYCLIC SUBGRAPH problem.

For all of these problems, a random assignment will give a constant-factor approximation algorithm. We observe that for fixed q, MAX CSP(<) over $[q]$ also admits a robust algorithm, and the same holds also for HARD CSP(<) over $[q]$. However, we show that adding hard constraints rules out the possibility of a constant-factor approximation algorithm, regardless of whether $[q]$ is fixed or unbounded.

Theorem 4.1.5. *For any $k \geq 2$ and $\epsilon > 0$, given an instance $\mathcal{I}$ of* HARD CSP*(<) over $[2k + 1]$, it is UG-hard to distinguish the following cases.*

- $\mathsf{Opt}_H(\mathcal{I}) \geq (1 - \epsilon)\frac{k-1}{k}$
- $\mathsf{Opt}_H(\mathcal{I}) \leq \epsilon$

	Ordinary	Balance	Hard
MAX 2-SAT	0.9401 $(1-\epsilon, 1-O(\sqrt{\epsilon}))$ UG: $(1-\epsilon, 1-\Omega(\sqrt{\epsilon}))$	0.9401 **N/A** **NP:** $(1, 1-\delta)$	**N/A** **N/A** **UG:**$(1-\epsilon, \epsilon)$
MAX CUT	0.8786 $(1-\epsilon, 1-O(\sqrt{\epsilon}))$ UG: $(1-\epsilon, 1-\Omega(\sqrt{\epsilon}))$	0.8776 Same as left Same as left	0.8786 Same as left Same as left
MAX HORN-2-SAT	0.9401 $(1-\epsilon, 1-2\epsilon)$	0.9401 **N/A** **NP:** $(1, 1-\delta)$	**N/A** $(1-\epsilon, 1-2\epsilon)$ **UG:**$(1-\epsilon, 1-2\epsilon)$
MAX HORN-SAT	0.7968 $(1-\epsilon, 1-O(\frac{\log\log(1/\epsilon)}{\log(1/\epsilon)}))$ UG: $(1-\epsilon, 1-\Omega(\frac{1}{\log(1/\epsilon)}))$	**0.75** **N/A** **NP:** $(1, 1-\delta)$	**N/A** **N/A** **NP:**$(1-\epsilon, \epsilon)$

Table 4.1: Summary of several BALANCE CSP and HARD CSP. $\epsilon > 0$ indicates an arbitrary positive constant, while $\delta > 0$ is a fixed absolute constant. In each cell, the first row contains the best approximation ratio. $(1-\epsilon, 1-f(\epsilon))$ in the second row indicates that there is a robust algorithm that find an assignment satisfying $(1-f(\epsilon))$ fraction of constraints given an $(1-\epsilon)$-satisfiable instance. $(1-\epsilon, 1-f(\epsilon))$ in the third row indicates that it is NP-hard to find an assignment satisfying $(1-f(\epsilon))$ fraction of constraints given an $(1-\epsilon)$-satisfiable instance. NP indicates that it is an NP-hardness result; UG indicates that it is based on the Unique Games Conjecture. N/A means that a robust or constant-factor approximation algorithm are ruled out by the hardness results. The results of this thesis are in boldface.

Over $\{1, 2, 3, 4\}$, it is also UG-hard to distinguish whether $\mathsf{Opt}_H(\mathcal{I}) \geq \frac{1-\epsilon}{2}$ *or* $\mathsf{Opt}_H(\mathcal{I}) \leq \epsilon$.

Also, for any $\epsilon > 0$, given an instance $\mathcal{I}$ of HARD CSP*($<$) over the unbounded domain (an instance of* MAXIMUM ACYCLIC SUBGRAPH *with several edges that must be included), it is UG-hard to distinguish the following cases.*

- $\mathsf{Opt}_H(\mathcal{I}) \geq 1 - \epsilon$
- $\mathsf{Opt}_H(\mathcal{I}) \leq \epsilon$

We also provide a constant-factor approximation algorithm for HARD CSP($<$) over $\{1, 2, 3\}$ (albeit with a worse ratio than its ordinary counterpart), showing a difference between $\{1, 2, 3\}$ and $\{1, 2, 3, 4\}$, and complete the picture of the approximability of these Ordering CSPs with hard constraints.

4.1.2 Organization

Section 4.2 proves the results about HARD CSP: HARD 2-SAT in Section 4.2.1 and HARD HORN SAT in Section 4.2.2. Section 4.3 proves the results about BALANCE CSP: hardness results in Section 4.3.1 and algorithmic results in Section 4.3.2. Section 4.4 proves the results about Ordering CSP. Section 4.5 discusses some open problems in this direction.

4.2 CSP with hard constraints

4.2.1 HARD 2-SAT

Fix $p = 1/2 - \epsilon$ for some $\epsilon > 0$. We let $\{0, 1\}^R_{(p)}$ be the R-dimensional Boolean hypercube with p-biased distribution; each coordinate of $x \in \{0, 1\}^R_{(p)}$ is independently set to 1 with probability p and 0 with probability $1 - p$. In the notation of Section 3.3, the p-biased distribution is (Ω, μ) where $\Omega = \{0, 1\}$, $\mu(1) = p$ and $\mu(0) = 1 - p$. The product distribution $\{0, 1\}^R_{(p)}$ is simply equal to $(\Omega^R, \mu^{\otimes R})$. Let $f : \{0, 1\}^n_{(p)} \to \mathbb{R}$.

At the heart of every hardness result based on the Unique Games Conjecture is an appropriate dictatorship test. The dictatorship test for VERTEX COVER (and its maximization version MAXIMUM INDEPENDENT SET) is a graph $G = (V, E)$ where $V = \{0, 1\}^R_{(p)}$ (vertices are weighted according to p-biased distribution). This graph must ensure that

- Completeness: For each $1 \leq i \leq R$, ith dictator function $f_i(x_1, ..., x_R) = x_i$ is the indicator function of a large independent set.
- Soundness: For every moderate-sized independent set D, its indicator function has a coordinate with large noisy influence (as defined in Definition 3.3.6).

The edges are constructed such that there is an edge between $(x_1, ..., x_R)$ and $(y_1, ..., y_R)$ if and only if there is no coordinate i such that $x_i = y_i = 1$. The first condition can be easily checked since $f_i(x) = f_i(y) = 1$ for some i means $x_i = y_i = 1$ and there is no edge between x and y. This construction was used in Dinur and Safra [DS05], and Khot and Regev [KR08] to prove hardness of VERTEX COVER. Later Austrin et al. [AKS09] gave a different (and arguably simpler) analysis of the same test, relying on the invariance principle of Mossel et al. [MOO10].

Theorem 4.2.1 (Implicit in [AKS09])**.** *Let $G = (V, E)$ be the graph constructed as above and $\mathsf{Inf}_i(T_\rho f)$ be the noisy influence defined in Definition 3.3.6. The following properties hold.*

- *Completeness: For each $1 \leq i \leq R$, ith dictator function $f_i(x_1, ..., x_R) = x_i$ is the indicator function of a large independent set of weight p.*
- *Soundness: For every $\epsilon > 0$, there exist $\tau > 0$ and $\delta > 0$ such that the following holds. If D is an independent set of weight at least ϵ and $f : \{0, 1\}^R_{(p)} \to \{0, 1\}$ is the indicator function of D, then there exists i such that $\mathsf{Inf}_i(T_{1-\delta} f) \geq \tau$.*

Combined with the standard technique converting a dictatorship test to a hardness result based on the Unique Games Conjecture, it is shown that it is UG-hard to distinguish whether the maximum independent set has weight at least $1/2 - \epsilon$ or at most ϵ.

There is a simple approximation-preserving reduction from MAXIMUM INDEPENDENT SET to HARD HORN 2-SAT. Given G constructed as above,

- $X = V$; each variable corresponds to one vertex.
- For each edge $(x, y) \in E$, add a hard constraint $(\neg x \vee \neg y)$.
- For each vertex $x \in V$, add a soft constraint x.

Hard constraints ensure that two variables corresponding to neighboring vertices cannot be set to True simultaneously, and maximizing the total weight of satisfied soft constraints is equivalent to maximizing the total weight of the vertices set to True. Therefore, the same hardness result also holds for HARD 2-SAT. This hardness result rules out any constant-factor approximation algorithm, but does not apply to instances that are almost satisfiable. In fact, HARD HORN 2-SAT inherits the same hardness result (as the above reduction only uses Horn clauses), but it has a robust algorithm (See Section 4.2.2).

Our crucial observation to *amplify* this gap to $1 - \epsilon$ and ϵ is that dictatorship functions always give different values to a pair of antipodal points. Since our predicate is a disjunction of two variables (vertices), if we change soft constraints so that

$$S = \Big\{ (u \vee v) : \{u, v\} \text{ is an antipodal pair in } \{0,1\}^R_{(p)} \Big\}$$

Then, every dictator function f_i will satisfy all the hard and soft constraints. To ensure the soundness, we have to *perturb* the distribution a little bit, but still almost all the weight will be concentrated around antipodal points. Given the dictatorship test $G = (V, E)$ for VERTEX COVER, the dictatorship test for HARD 2-SAT is $\mathcal{I} = (X, S, H)$ such that

- $X = V$; each variable corresponds to one vertex.
- For each edge $(x, y) \in E$, add a hard constraint $(\neg x \vee \neg y)$.
- Sample an ordered pair (x, y) such that for each i,

  $$\mathbb{P}[x_i = y_i = 0] = 2\epsilon \qquad \text{and} \qquad \mathbb{P}[x_i = 1, y_i = 0] = \mathbb{P}[x_i = 0, y_i = 1] = 1/2 - \epsilon\ .$$

  Add a soft constraint $(x \vee y)$ with the weight equal to the probability of (x, y).

With this trick, we can obtain a dictatorship test for HARD 2-SAT with a much larger gap.

Theorem 4.2.2. *Let $\mathcal{I} = (X, S, H)$ be the instance of* HARD 2-SAT *constructed as above.*

- *Completeness: For each $1 \le i \le R$, ith dictator function $f_i(x_1, ..., x_R) = x_i$ satisfies $\mathsf{Val}_H(f_i) = 1 - 2\epsilon$.*
- *Soundness: For every $\epsilon > 0$, there exist $\tau > 0$ and $\delta > 0$ such that the following holds. If $f : \{0,1\}^R_{(p)} \to \{0,1\}$ satisfies $\mathsf{Val}_H(f) \ge 2\epsilon$, then there exists i such that $\mathsf{Inf}_i(T_{1-\delta} f) \ge \tau$.*

Proof. COMPLETENESS: It is already shown that ith dictator function f_i satisfies all the hard constraints. The only soft constraints that f_i fail to satisfy is $(x \vee y)$ where $x_i = y_i = 0$. Since the probability of picking $x_i = y_i = 0$ is 2ϵ, $\mathsf{Val}_H(f_i) = 1 - 2\epsilon$.

SOUNDNESS: If f is a function that satisfies all the hard constraints, it is also the indicator function of some independent set D. We claim that the weight of D in $\{0,1\}^R_{(p)}$ with $p = 1/2 - \epsilon$ is at least $\mathsf{Val}_H(f_i)/2$. For $z \in \{0,1\}^R_{(p)}$, let $\mathsf{deg}(z)$ be the sum of the weights of soft constraints $(z \vee x)$ or $(x \vee z)$ for some x ($(z \vee z)$ contributes twice). $(x,y) \sim S$ indicates that it is sampled according to its weight defined above. Note that its marginal distribution of x (and y) is exactly the p-biased distribution on $\{0,1\}^R_{(p)}$. Therefore,

$$\begin{aligned} \mathsf{deg}(z) &= \mathbb{P}_{(x,y)\sim S}[z = x] + \mathbb{P}_{(x,y)\sim S}[z = y] \\ &= 2\,\mathbb{P}_{(x,y)\sim S}[z = x] \\ &= 2\,\mathbb{P}_{x\sim\{0,1\}^R_{(p)}}[z = x] \\ &= 2\,\mathsf{wt}(z)\ . \end{aligned}$$

It means that for any z, switching $f(z)$ from 0 to 1 increases $\mathsf{Val}_H(f)$ by at most $2\mathsf{wt}(z)$. Therefore, $\mathsf{Val}_H(f)$ is at most two times the weight of D. Thus $\mathsf{Val}_H(f) \geq 2\epsilon$ indicates that the weight of D is at least ϵ. We can now use Theorem 4.2.1 to finish the argument. □

Combined with the standard technique converting a dictatorship test to a hardness result based on the Unique Games Conjecture, the main theorem of this section is proved.

Theorem 4.2.3 (Restatement of Theorem 4.1.3)**.** *For any $\epsilon > 0$, given an instance $\mathcal{I}$ of* HARD 2-SAT*, it is UG-hard to distinguish the following cases.*

- $\mathsf{Opt}_H(\mathcal{I}) \geq 1 - \epsilon$
- $\mathsf{Opt}_H(\mathcal{I}) \leq \epsilon$

Proof. Given an instance of $\mathcal{L}(G(U \cup W, E), [R], \{\pi(v,w)\}_{(v,w)\in E})$ of Unique Games, we construct an instance $\mathcal{I} = (X, S, H)$ of HARD 2-SAT.

- $X = W \times \{0,1\}^R$. The weight of (w,x) is the weight of x in $\{0,1\}^R_{(1/2-\epsilon)}$, divided by $|W|$ (weights are only used in the analysis).
- For every pair of edges $(u,w_1),(u,w_2)$ with the same endpoint $u \in U$, and every $x,y \in \{0,1\}^R$, such that there is no i such that $x_{\pi^{-1}(u,w_1)(i)} = y_{\pi^{-1}(u,w_2)(i)} = 1$, add a hard constraint $(\neg(w_1,x) \vee \neg(w_2,y))$.
- For each w, sample an ordered pair (x,y) such that for each i, $\mathbb{P}[x_i = y_i = 0] = 2\epsilon$ and $\mathbb{P}[x_i = 1, y_i = 0] = \mathbb{P}[x_i = 0, y_i = 1] = 1/2 - \epsilon$. Add a soft constraint $((w,x) \vee (w,y))$ with the weight equal to the probability of (x,y) divided by $|W|$.

Note that the variables and hard constraints of this construction are identical to that of VERTEX COVER (MAXIMUM INDEPENDENT SET) of [AKS09]. We use the variant of the Unique Games Conjecture (see Conjecture 3.2.3), and prove the following:

Lemma 4.2.1. *Given an instance of Unique Games $\mathcal{L}$ and the instance $\mathcal{I}$ produced as above,*

- *If there is a set $W' \subseteq W$ such that $|W'| \geq (1-\epsilon)|W|$ and a labeling $l : U \cup W \to [R]$ that satisfies every edge (u, w) for $u \in U$ and $w \in W'$, then $\mathsf{Opt}_H(\mathcal{I}) \geq 1 - 3\epsilon$.*
- *There is a function $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that if $\mathsf{Opt}_{\mathsf{UG}}(\mathcal{L}) \leq f(\epsilon)$, then $\mathsf{Opt}_H(\mathcal{I}) \leq 2\epsilon$.*

Proof. We establish the completeness and soundness in turn.

COMPLETENESS: Let $W' \subseteq W$ and $l : U \cup W \to [R]$ be a subset and a labeling satisfying the above condition. For any $w \in W', x \in \{0,1\}$, set $(w, x) = x_{l(w)}$. If $w \notin W'$, $(w, x) = 0$ for all x. Note that this satisfies all the hard constraints; if there is a violated hard constraint $(\neg(w_1, x) \vee \neg(w_2, y))$, it means that there exist $u \in U$, $w_1, w_2 \in W'$ such that $l(w_1) = \pi^{-1}(u, w_1)(l(u))$, $l(w_2) = \pi^{-1}(u, w_2)(l(u))$, and $x_{l(w_1)} = y_{l(w_2)} = 1$. It contradicts the above contradiction. For each $w \in W'$, the only soft constraints $((w, x) \vee (w, y))$ that are not satisfied by this assignment have $x_{l(u)} = y_{l(u)} = 0$, which happens with probability 2ϵ. Therefore, the total weight of soft constraints satisfied is at least $(1-\epsilon)(1-2\epsilon) \geq 1 - 3\epsilon$.

SOUNDNESS: Suppose there is an assignment $\sigma : U \times \{0,1\}^R \to \{0,1\}$ such that it satisfies all the hard constraints and $\mathsf{Val}_H(\sigma) \geq 2\epsilon$. Let $D \subseteq U \times \{0,1\}^R$ be the support of σ. Since σ satisfies all the hard constraints, D is an independent set of the graph whose vertex set is $U \times \{0,1\}^R$ and each pair of variables in a same hard constraint forms an edge. Note that this is the same graph used in the hardness of VERTEX COVER in [AKS09].

Since the soft constraints are defined within each hypercube $u \times \{0,1\}$ for each u, we can use the same analysis from Theorem 4.2.2 (which says that the sum of weights of the soft constraints containing a variable in the HARD 2-SAT instance is exactly two times the weight of the corresponding vertex in the MAXIMUM INDEPENDENT SET instance) to conclude that the weight of D is at least ϵ. We can invoke Theorem 3.1 of [AKS09] to argue that $\mathsf{Opt}(\mathcal{L}) \geq f(\epsilon)$ for some fixed function $f : \mathbb{R}^+ \to \mathbb{R}^+$. ☐

Therefore, if we can distinguish whether $\mathsf{Opt}_H(\mathcal{L}) \geq 1 - 3\epsilon$ or $\mathsf{Opt}_H(\mathcal{L}) \leq 2\epsilon$ for some $\epsilon > 0$, then we can refute Conjecture 3.2.3, which is equivalent to the original Unique Games Conjecture. This proves Theorem 4.1.3. ☐

4.2.2 HARD HORN SAT

Suppose there is a very strong robust algorithm for MAX CSP(Π) in the sense that if the algorithm can find σ such that $\mathsf{Val}(\sigma) \geq 1 - c\epsilon$ whenever $\mathsf{Opt}(\mathcal{I}) \geq 1 - \epsilon$ for some absolute constant c (or in other words, we have a constant-factor approximation algorithm for the minimization version that seeks to minimize the weight of unsatisfied constraints). The performance of this algorithm is preserved for HARD CSP(Π) by the following trick: convert all hard constraints to soft constraints, but give them very large weight so that they will never be violated in the optimal solution. This will result in $\mathcal{I}'$ such that $\mathsf{Opt}(\mathcal{I}') \geq 1 - \epsilon'$ where $\epsilon' \ll \epsilon$, but the algorithm still finds σ such that $\mathsf{Val}_H(\sigma) \geq 1 - c\epsilon'$ as a solution to $\mathcal{I}'$, which also satisfies $\mathsf{Val}_H(\sigma) \geq 1 - c\epsilon$ as a solution to $\mathcal{I}$.

Now, MAX HORN-2-SAT is a problem that admits such a robust algorithm with $c = 2$ [GZ12] ($c = 3$ was earlier shown in [KSTW01]), and therefore we can conclude that HARD HORN 2-SAT admits a similar robust algorithm.

Furthermore, the reduction from MAXIMUM INDEPENDENT SET to HARD 2-SAT introduced in Section 4.2.2 is indeed a reduction to HARD HORN 2-SAT. Therefore, by previous results about MAXIMUM INDEPENDENT SET [BK10, KR08], for any $\epsilon > 0$, it is UG-hard to find σ with $\mathsf{Val}_H(\sigma) \geq \epsilon$ even when $\mathsf{Opt}_H(\mathcal{I}) \geq 1/2 - \epsilon$. By adding many dummy constraints that are always satisfiable, for any $\epsilon, \delta > 0$, it is UG-hard to find σ with $\mathsf{Val}_H(\sigma) \geq 1 - 2\epsilon + \delta$ even when $\mathsf{Opt}_H(\mathcal{I}) \geq 1 - \epsilon - \delta$. These facts justfy the following observation made about HARD HORN 2-SAT.

Observation 4.2.2 (Restatement of Observation 4.1.3)**.** *There is a polynomial time algorithm that for any $\epsilon \geq 0$, if the given instance $\mathcal{I}$ of* MAX HORN-2-SAT *satisfies* $\mathsf{Opt}_H(\mathcal{I}) = 1 - \epsilon$*, finds σ with* $\mathsf{Val}_H(\sigma) \geq 1 - 2\epsilon$*.*

Furthermore, for any $\epsilon, \delta > 0$, it is UG-hard to distinguish the following cases.

- $\mathsf{Opt}(\mathcal{I}) \geq 1 - \epsilon - \delta$.
- $\mathsf{Opt}(\mathcal{I}) \leq 1 - 2\epsilon + \delta$.

The above algorithmic result for MAX HORN-2-SAT does not hold for MAX HORN-SAT in general since the robust algorithm is only guaranteed to find an assignment satisfying $1 - O\left(\frac{\log\log(1/\epsilon)}{\log(1/\epsilon)}\right)$ fraction of clauses [Zwi98b], and this exponential loss is inherent under the Unique Games conjecture [GZ12]. In fact, even Horn-3-SAT is powerful enough to encode constraints of other hard problems with unbounded arity, which results in a very strong hardness result, stated below.[2]

[2] We learned from Andrei Krokhin that this result, with essentially the same proof, was also shown by

Theorem 4.2.4 (Restatement of Theorem 4.1.4)**.** *For any $\epsilon > 0$, given an instance $\mathcal{I}$ of* HARD HORN 3-SAT*, it is NP-hard to distinguish the following cases.*

- $\mathsf{Opt}_H(\mathcal{I}) \geq 1 - \epsilon$
- $\mathsf{Opt}_H(\mathcal{I}) \leq \epsilon$

Proof. We reduce Ek-HYPERGRAPH INDEPENDENT SET to MAX HORN-3-SAT. An instance of Ek-HYPERGRAPH INDEPENDENT SET is a hypergraph $G = (V, E)$ where each hyperedge $e \in E$ contains exactly k vertices. Our goal is to find a set $D \subseteq V$ with the maximum weight such that no hyperedge e is a subset of D.

Given a graph $G = (V, E)$, we construct the instance $\mathcal{I} = (X, S, H)$ of MAX HORN-3-SAT as follows.

- $X = V \cup \{y_{e,j} : e \in E, 1 \leq j \leq k\}$; there is one variable for each vertex, and k variables for each hyperedge. Each variable corresponding to a vertex indicates that the vertex is picked or not.
- For each hyperedge $e = (v_1, ..., v_k) \in E$, add hard constraints
  - $(v_1 \to \neg y_{e,1}) \equiv (\neg v_1 \vee \neg y_{e,1})$
  - $2 \leq j \leq k$: $(\neg y_{e,i-1} \wedge v_i \to \neg y_{e,i}) \equiv (y_{e,i-1} \vee \neg v_i \vee \neg y_{e,i})$
  - $y_{e,k}$
- For each vertex $v \in V$, add a soft constraint v with the same weight as in G.

The above construction ensures that there is at most one unnegated literal per each clause, so this is indeed an instance of MAX HORN-3-SAT. Once $v_1, ..., v_n$ are fixed, a quick check of the second set of constraints corresponding to hyperedge e ensures that there exists $y_{e,1}, ..., y_{e,k}$ that satisfy all the constraints if and only if at least one of $v_1, ..., v_k$ is set to False. Since the weight of satisfied soft constraints is equal to the weight of vertices picked, this is an approximation-preserving reduction from Ek-HYPERGRAPH INDEPENDENT SET to MAX HORN-3-SAT. Dinur et al [DGKR05] showed that for the former, it is NP-hard to distinguish

- There is an independent set of weight $(1 - \frac{1}{k-1} - \epsilon)$.
- Every independent set is of weight at most ϵ.

By taking k large and ϵ small, we get the desired result for MAX HORN-3-SAT. □

Siavosh Bennabas. But as we are not aware of a published reference, we include the simple proof.

4.3 Balance Constraints

4.3.1 Hardness Results

Theorem 4.3.1 (Restatement of Theorem 4.1.1)**.** *There exists an absolute constant $\delta > 0$ such that given an instance $\mathcal{I}$ of* BALANCE HORN 2-SAT *(special case of* BALANCE 2-SAT *and* BALANCE HORN-SAT*), it is NP-hard to distinguish the following cases.*

- $\mathsf{Opt}_B(\mathcal{I}) = 1$
- $\mathsf{Opt}_B(\mathcal{I}) \leq 1 - \delta$

Proof. We reduce from MAX 3-SAT(B) to MAXIMUM INDEPENDENT SET to BALANCE HORN 2-SAT, where in MAX 3-SAT(B) each variable occurs at most B times. The following description of the reduction from MAX 3-SAT(B) to MAXIMUM INDEPENDENT SET is from Papadimitriou and Yannakakis [PY91].

> *Construct a graph with one node for every occurrence of every literal. There is an edge connecting any two occurrences of complementary literals, and also, an edge connecting literal occurrences from the same clause (thus, there is a triangle for every clause with 3 literals, and an edge for a clause with 2 literals). The size of the maximum independent set in the graph is equal to the maximum number of clauses that can be satisfied. If every variable occurs at most B times in the clauses, then the degree is at most $B + 1$.*

Note that in the above reduction, there is an independent set of size l if and only if there is an assignment that satisfies l clauses. For some constants B and $\delta_0 > 0$, it is NP-hard to find an assignment satisfying $(1 - \delta_0)$ fraction of clauses in a satisfiable instance of MAX 3-SAT(B). Let m be the number of clauses, and n be the number of variables of the given MAX 3-SAT(B) instance, so we have $3m$ vertices, and at most $1.5(B + 1)m \leq 2Bm$ edges in the graph. Our BALANCE HORN 2-SAT instance $\mathcal{I}$ consists of $4m$ variables; $3m$ of them correspond to the vertices of the graph, and m of them do not participate in any constraint. For each edge (u, v) we add the constraint $(\neg u \vee \neg v)$. Finally, we have the balance constraint that exactly $2m$ of them should be 1 (1 means True in 2-SAT, and that the vertex is picked in the Independent Set problem).

If the MAX 3-SAT instance is satisfiable, we have an independent set of size $2m$ consisting of m vertices from the graph and m of the dummy vertices, so the BALANCE HORN 2-SAT instance is also satisfiable.

Now suppose that $\mathsf{Opt}_B(\mathcal{I}) > 1 - \delta$. This means that there is a balanced assignment (at least m 1's amongst the non-dummy variables) such that at most a fraction δ of the edges have both endpoints set to 1. By switching $2\delta Bm$ vertices from 1 to 0, all clauses of BALANCE HORN 2-SAT will be satisfied. This means that there is an independent set of size at least $(1 - 2\delta B)m$, and therefore also an assignment that satisfies $(1 - 2\delta B)m$ clauses of the MAX 3-SAT(B) instance. It follows that we must have $\delta \geq \frac{\delta_0}{2B}$. □

Unlike the reduction from MAXIMUM INDEPENDENT SET to HARD HORN 2-SAT introduced in Section 4.2.1 (hard clauses for independence constraints, and soft clauses for the objective), we use clauses to enforce independence constraints, and the balance constraint to maximize the objective. For the soundness analysis, given a good assignment to BALANCE HORN 2-SAT, when viewed as a (slightly infeasible) solution to MAXIMUM INDEPENDENT SET, the balance constraint ensures that the objective is good, but there might be some adjacent vertices picked. The bounded degrees allow us to fix this solution to an independent set while still retaining many vertices.

4.3.2 Algorithmic Results

One of the most well-studied BALANCE CSP is BALANCE CUT (MAXIMUM BISECTION). Over a long line of work [FJ95, HZ01, Ye01, FL01, RT12, ABG13], the approximation ratio for MAXIMUM BISECTION has become 0.8776 which is very close to the optimal (under the Unique Games Conjecture) approximation ratio for MAX CUT which is about 0.8786. MAXIMUM BISECTION admits a robust algorithm as well [GMR$^+$11, RT12]. Note that BALANCE CSP(Π) is no easier to approximate than MAX CSP(Π), since any instance of MAX CSP(Π) can be reduced to an instance of BALANCE CSP(Π) by adding dummy variables that do not participate in any constraint.

For BALANCE 2-SAT, the best known approximation ratio is 0.9401, matching the best known approximation ratio of MAX 2-SAT [LLZ02]. This result indicates that in the approximation ratio perspective, adding the balance constraint does not make the problem harder. However, Theorem 4.3.1 rules out any robust algorithm for BALANCE 2-SAT, which shows a stark difference between the balanced versions of MAX 2-SAT and MAX CUT.

For BALANCE HORN-SAT, the same hardness result shows that we cannot hope for any robust algorithm. Therefore, the only remaining quetion is whether there is an algorithm whose approximation ratio nearly matches that of the best algorithm for ordinary MAX HORN-SAT. The best approximation ratio for MAX HORN-SAT is achieved by an algorithm for more general MAX SAT, which achieves 0.7968-approximation [ABZ06].[3]

The most recent results on BALANCE CUT and BALANCE 2-SAT [RT12, ABG13] rely on Lasserre SDP hierarchies. The purpose of using a more sophisticated SDP rather than the basic SDP of Goemans and Williamson [GW95], is that during the rounding scheme, the vertices are not rounded independently; even if we add the balance constraint in the SDP relaxation, the final solution is not guaranteed to be approximately balanced when we do not have a guarantee about their correlations. The goal of Lasserre hierarchies is to produce an SDP solution with low global correlation so that each vertex can be rounded almost independently.

However, in the $3/4$-approximation algorithm for MAX SAT by Goemans and Williamson [GW94], which is based on a LP-relaxation, each variable is rounded independently. Therefore, adding the balance constraint ensures that the final solution is almost balanced by a simple application of Chernoff bound. After a simple correction phase to get perfect balance (slightly more sophisticated than [RT12] since the arity is not bounded), we obtain an algorithm which is not far from the best algorithm for MAX SAT in terms of approximation ratio.

Theorem 4.3.2 (Restatement of Theorem 4.1.2)**.** *For any $\epsilon > 0$, there is a randomized algorithm such that given an instace $\mathcal{I}$ of* BALANCE SAT*, in time* $\mathrm{poly}(size(\mathcal{I}), \frac{1}{\epsilon})$*, outputs σ with $\mathsf{Val}_B(\sigma) \geq (\frac{3}{4} - \epsilon)\mathsf{Opt}_B(\mathcal{I})$ with constant probability.*

Proof. Let $\mathcal{I} = (X, C)$ be an instance BALANCE SAT, where $X = \{x_1, ..., x_n\}$ is the set of variables and $C = \{C_1, ..., C_m\}$ is the set of clauses. For each $1 \leq j \leq m$, let C_j^+ (resp. C_j^-) be the set of variables which appear in C_j *unnegated* (resp. *negated*). The following is a natural LP relaxation of BALANCE SAT.

$$\begin{aligned} \text{maximize} \quad & \sum_{j=1}^{m} \mathsf{wt}(C_j) z_j \\ \text{subject to} \quad & \forall 1 \leq j \leq m : \sum_{i \in C_j^+} y_i + \sum_{i \in C_j^-} (1 - y_i) \geq z_j \\ & \forall 1 \leq j \leq m : 0 \leq z_j \leq 1 \\ & \forall 1 \leq i \leq n : 0 \leq y_i \leq 1 \\ & \sum_{i=1}^{n} y_i = \frac{n}{2} \end{aligned}$$

Given the solution (y_i, z_j) to the above LP, the rounding algorithm is simple:

[3]The same paper gives another algorithm whose approximation ratio is 0.8434 under some conjecture.

1. Choose $a \in \{0, 1\}$ uniformly at random.
2. If $a = 0$, set $\sigma(x_i) = 1$ with probability y_i independently; if $a = 1$, set $\sigma(x_i) = 1$ with probability 0.5 independently.
3. If the final solution is unbalanced (there exists $b \in \{0, 1\}$ such that $|\sigma^{-1}(b)| = (0.5 + \eta)n$ for some $\eta > 0$), pick exactly ηn variables from $\sigma^{-1}(b)$ uniformly at random, set them to $1 - b$.
4. Iterate it $O(1/\epsilon)$ times.

Claim 4.3.1. *At the end of step 2, $\mathbb{E}[\mathsf{Val}(\sigma)] \geq 3/4$ (note that it is $\mathsf{Val}(\sigma)$, but not $\mathsf{Val}_B(\sigma)$).*

Proof. For each C_j, let k be the number of variables in C_j and without loss of generality assume $C_j = x_1 \vee ... \vee x_k$ by renaming and negating variables. If $a = 0$, the probability that C_j is satisfied is

$$1 - \Pi_{i=1}^{k}(1 - y_i) \geq 1 - (\frac{\sum_{i=1}^{k}(1 - y_i)}{k})^k \geq 1 - (1 - \frac{z_i}{k})^k \geq \beta_k z_i$$

where $\beta_k = 1-(1-\frac{1}{k})^k$. If $a = 1$, the probability that C_j is satisfied is at least $\alpha_k = 1-2^{-k}$. Overall, the probability that C_j is satisfied is $\frac{\alpha_k+\beta_k}{2}$. For $k = 1, 2$, $\frac{\alpha_k+\beta_k}{2} = \frac{3}{4}$, and for $k \geq 3$, $\frac{\alpha_k+\beta_k}{2} \geq \frac{7/8+(1-1/e)}{2} \geq \frac{3}{4}$. □

By a simple averaging argument, with probability at least ϵ, $\mathsf{Val}(\sigma) \geq 3/4 - \epsilon$. Since each variable is rounded independently once a is fixed, by Chernoff bound, $\mathbb{P}[|\sigma^{-1}(1) - \frac{n}{2}| > \epsilon n] \leq 2exp(-\frac{2\epsilon^2}{n}) = o_n(1)$. Therefore, at the end of step 2, the probability that $\mathsf{Val}(\sigma) \geq 3/4 - \epsilon$ and $|\sigma^{-1}(1) - \frac{n}{2}| \leq \epsilon n$ is at least $\epsilon/2$.

Let b be such that $|\sigma^{-1}(b)| = (1/2 + \eta)n$ for some $0 < \eta \leq \epsilon$. In step 3, we randomly choose ηn variables from $\sigma^{-1}(b)$, and set them to $1 - b$. Fix a constraint C_j that is satisfied by σ. Since C_j is a disjunction, there is only one assignment to the variables appearing C_j that makes C_j unsatisfied. Since C_j is already satisfied, to reach the only unsatisfying assignment, at least one variable must be switched. If there is a variable that must be switched from $1 - b$ to b, then C_j is guaranteed to be satisfied even after step 3. Otherwise, there are k variables ($k \geq 1$) that must be swithced from b to $1 - b$, and the probability of switching all of them is at most the probability of switching one of them, which is $\frac{\eta n}{(1/2+\eta)n} \leq 2\eta$. Therefore, in expectation, at most 2η fraction of already satisfied clauses can be unsatisfied, and with probability half, at most $4\eta \leq 4\epsilon$ fraction of satisfied clauses can be unsatisfied. Combining step 2 and 3, in each iteration, we get a balanced σ with $\mathsf{Val}_B(\sigma) \leq 3/4 - 5\epsilon$ with probability $\epsilon/4$. This probability can be made to a constant by repeating $O(1/\epsilon)$ times. □

4.4 Ordering with Hard Constraints

Another natural class of 2-ary CSPs over non-Boolean domain is MAX CSP($<$). MAX CSP($<$) over the unbounded domain (indeed $[n]$ is enough) is the famous MAXIMUM ACYCLIC SUBGRAPH, which admits a simple $\frac{1}{2}$-approximation algorithm that is optimal under the Unique Games Conjecture [GMR08]. For fixed q, MAX CSP($<$) over $[q]$ also admits a simple $\frac{(q-1)}{2q}$ algorithm by taking a random assignment.

These ordering problems can be best understood in terms of choosing the maximum subgraph of a directed graph with certain desired properties. MAX CSP($<$) over $[q]$ is equivalent to the following problem: given a directed graph $G = (V, E)$, find $G' = (V, E')$ with $E' \subseteq E$ such that G' does not have a walk of length q and $|E'|$ is maximized. A walk of length q is a sequence of vertices $v_0 \to \cdots \to v_q$ (vertices can be repeated) such that $(v_i, v_{i+1}) \in E$ for $0 \leq i < q$. Note that a directed cycle results in a walk of any length. MAXIMUM ACYCLIC SUBGRAPH is to find G' such that G' does not have any cycle.

Instead of choosing a subset of the edges, we can try to find a subset of the vertices so that the induced subgraph has desired properties. Let MAX VERTEX CSP($<$) and MAXIMUM VERTEX ACYCLIC SUBGRAPH be the analogous problems where the goal is to choose the subset of variables (vertices) with the maximum weight so that the induced CSP is satisfiable (induced graph has the desired properties).

Despite of their similarity, MAX VERTEX CSP problems seem to be harder than their edge counterparts. There is no intuitive constant-factor approximation algorithm, like taking a random assignment in the edge problems. Svensson [Sve13] confirmed that these problems are indeed a lot harder to approximate, assuming the Unique Games Conjecture.

Theorem 4.4.1 (Theorem 1 of [Sve13]). *For any $k \geq 2$ and $\epsilon > 0$, given an instance $\mathcal{I}$ of* MAX VERTEX CSP*($<$) over the domain $[k-1]$, it is UG-hard to distinguish*

- $\mathsf{Opt}(\mathcal{I}) \geq (1-\epsilon)\frac{k-1}{k}$
- $\mathsf{Opt}(\mathcal{I}) \leq \epsilon$

Also, for any $\epsilon > 0$, given an instance $\mathcal{I}$ of MAXIMUM VERTEX ACYCLIC SUBGRAPH*, it is UG-hard to distinguish*

- $\mathsf{Opt}(\mathcal{I}) \geq 1-\epsilon$
- $\mathsf{Opt}(\mathcal{I}) \leq \epsilon$

We show that introducing hard constraints makes the edge versions almost as hard as the vertex counterparts. The reduction, introduced in Even et al. [ENSS98], is simple:

split each vertex v to a soft edge $v_{in} \to v_{out}$ and add a hard edge $u_{out} \to v_{in}$ for each edge $u \to v$. This is almost enough to show the following theorem.

Theorem 4.4.2 (Restatement of Theorem 4.1.5)**.** *For any $k \geq 2$ and $\epsilon > 0$, given an instance $\mathcal{I}$ of* HARD CSP*(<) over $[2k+1]$, it is UG–hard to distinguish the following cases.*

- $\mathsf{Opt}_H(\mathcal{I}) \geq (1-\epsilon)\frac{k-1}{k}$
- $\mathsf{Opt}_H(\mathcal{I}) \leq \epsilon$

Over the domain $\{1, 2, 3, 4\}$, it is also UG-hard to distinguish whether

$$\mathsf{Opt}_H(\mathcal{I}) \geq \frac{1-\epsilon}{2} \qquad \text{or} \qquad \mathsf{Opt}_H(\mathcal{I}) \leq \epsilon\ .$$

Also, for any $\epsilon > 0$, given an instance $\mathcal{I}$ of HARD CSP*(<) over the unbounded domain (an instance of* MAXIMUM ACYCLIC SUBGRAPH *with several edges that must be included), it is UG-hard to distinguish the following cases.*

- $\mathsf{Opt}_H(\mathcal{I}) \geq 1-\epsilon$
- $\mathsf{Opt}_H(\mathcal{I}) \leq \epsilon$

Proof. The last result about the unbounded domain (MAXIMUM ACYCLIC SUBGRAPH with hard edges) is immediate from the reduction above and the hardness of MAXIMUM VERTEX ACYCLIC SUBGRAPH.

We will now describe the first reduction from the instance $\mathcal{I}$ of MAX VERTEX CSP(<) over $[k-1]$ to the instance $\mathcal{I}'$ of HARD CSP(<) over $[2k+1]$ (in this proof, we identify each instance of the CSP with its underlying directed graph). We perform the same reduction as above, and add two special vertices s and t. Add a hard edge $s \to v_{in}$ if v_{in} has no incoming edge. Make a hard edge $v_{out} \to t$ if v_{out} has no outgoing edge.

There is a one-to-one correspondence between the vertices of $\mathcal{I}$ and the soft constraints of $\mathcal{I}'$. Let G denote a vertex-induced subgraph of $\mathcal{I}$ and G' be its corresponding edge subgraph of $\mathcal{I}'$. If there is a walk $(v_1, \dots, v_k)$ of length $k-1$ exists in G, let v_0 be any vertex which has a outgoing edge to v_1 ($v_{0,out} := s$ if v_1 has no incoming edge), and let v_{k+1} be any vertex which has a incoming edge from v_k ($v_{k+1,in} := t$ if v_k has no outgoing edge). Then $v_{0,out} \to v_{1,in} \to v_{1,out} \cdots \to v_{k,out} \to v_{k+1,in}$ is a walk of length $2k+1$ in G'. Similarly, since any walk of G' alternates soft edges and hard edges, a walk of length $2k+1$ in G' induces a walk of length $k-1$ in G. Therefore, there is a one-to-one correspondence between feasible solutions, and corresponding solutions have the same weight.

Finally, we show the hardness of HARD CSP($<$) over $\{1, 2, 3, 4\}$ by reducing MAXIMUM INDEPENDENT SET to it. Given an instance (undirected graph) $\mathcal{I}$, consider it as a directed graph with edges going both directions, and apply the above reduction to get $\mathcal{I}'$. There is again one-to-one correspondence between a vertex-induced subgraph $G \subseteq \mathcal{I}$ and a edge subgraph $G' \subseteq \mathcal{I}'$. If $(u, v) \in G$, $u_{in} \to u_{out} \to v_{in} \to v_{out} \to u_{in}$ forms a cycle of G', so it is not feasible. If there is a walk of length at least 4 in G', since it alternates soft and hard constraints, there are two neighboring vertices u and v so that $u_{in} \to u_{out}, v_{in} \to v_{out} \in G'$, so G is not an independent set. $\square$

It is easy to get a constant-factor approximation for HARD CSP($<$) over the domain $\{1, 2\}$. Each hard constraint uniquely determines the value of the variables participating, and randomly assigning remaining variables ensures that each soft constraint, if satisfiable at all, can be satisfied with some constant probability. A slight extension of this technique works for $\{1, 2, 3\}$ as well, in stark constrast to $\{1, 2, 3, 4\}$.

Theorem 4.4.3. *There is a constant-factor approximation algorithm for* HARD CSP*($<$) over* $\{1, 2, 3\}$.

Proof. Let $\mathcal{I}$ be an instance of HARD CSP($<$) over $\{1, 2, 3\}$. We also think $\mathcal{I}$ as a directed graph (V, E) such that for each constraint $u > v$ there is an edge $u \to v$. Each edge is either hard or soft, depending on the corresponding constraint. By virtue of certain combinations of hard constraints, namely paths of length 2, the value of some vertices becomes fixed; let F_1, F_2, F_3 be the set of vertices fixed to $1, 2, 3$ respectively. In the subgraph induced by $V \setminus (F_1 \cup F_2 \cup F_3)$, let B be the set of vertices v such that $v \to u$ is a hard edge for some u. Similarly, S be the set of vertices u such that $v \to u$ is hard for some v. Let R be the remaining vertices. Note there is no hard edge from $F_1 \cup F_2$ to $B \cup S \cup R$, and no hard edge from $B \cup S \cup R$ to $F_2 \cup F_3$, and from S to B (e.g. if there is a hard edge from S to B, we have a path of length 2 and these vertices should be fixed). The algorithm is the following.

1. Choose $a \in \{0, 1\}$ uniformly at random.
2. If $a = 0$, for each vertex $v \in B$, $v = 3$; for each vertex $v \in S$, choose $v \in \{1, 2\}$ uniformly and independently.
3. If $a = 1$, for each vertex $v \in B$, choose $v \in \{2, 3\}$ uniformly and independently; for each vertex $v \in S$, choose $v = 1$.
4. Regardless of a, for each vertex $v \in R$, choose $v \in \{1, 2, 3\}$ uniformly and independently.

Simple case analyses between possible soft edges between the sets yield the following:

- Each soft edge completely inside one of B, S, R is satisfied with probability at least $1/8$.
- Each soft edge whose one endpoint is in R and the other in one of B, S is satisfied by with probability at least $1/12$.
- Each soft edge to F_1 or from F_3 is satisfied with probability at least $1/4$.

Therefore, the above simple algorithm yields a $1/12$-approximation algorithm. $\square$

For fixed q, MAX CSP($<$) over $[q]$ admits a simple robust algorithm that finds a solution of value $(1-q\epsilon)$, given $(1-\epsilon)$-satisfiable instance: given a directed graph $G = (V, E)$, set one nonnegative variable y_e for each e, and for any walk of length q, add a constraint that the sum of the corresponding y's must be at least 1. Deleting e when $y_e \geq \frac{1}{q}$ gives a q-approximation algorithm for the minimization version. As this is a strong robust algorithm (as discussed in Section 4.2.2), it is preserved for HARD CSP($<$) over $[q]$.

Observation 4.4.1. *For fixed q,* HARD CSP*($<$), and consequently* MAX VERTEX CSP*($<$) over $[q]$ admits a robust algorithm.*

That gives the complete picture of the status these problems. The conclusion is that adding hard constraints to a problem makes it as hard as its vertex counterparts, which is much harder to approximate.

4.5 Discussion

MAX CUT is one of the simplest and well-studied CSPs. Its variants with balance or hard variants also inherit the desirable algorithmic properties of the original version — in particular both these variants admit robust satisfiability as well as constant-factor approximation algorithms. However, these algorithmic results do not extend to even slightly more general CSPs; in MAX LIN-MOD-2, even one of balance and hard variants rule out the possibility of a robust algorithm. Furthermore, HARD 2-SAT does not even have a constant-factor approximation algorithm, and the hardness result for BALANCE 2-SAT also holds for BALANCE HORN 2-SAT (which does not even capture MAX CUT).

These delicate differences even for CSPs of arity two over the Boolean domain (where satisfiability is always in P) suggest that adding balance or hard constraints highlights previously unnoticed characteristics with respect to the approximability of each predicate. This chapter settles the approximability of almost all interesting arity two CSPs over Boolean domain, demonstrating the effect of hard and balance constraints on some classic problems.

While our techniques are relatively simple twists to known ones (with the one underlying the inapproximability of HARD 2-SAT being the most significant), we think the body of results highlights subtle (and perhaps surprising) differences between CSPs in the presence of hard/balance constraints, and raises the challenge of extending our study to larger arity or domains. Attempts at such a generalization should be interesting by itself and also illuminate new aspects of CSPs.

In this regard, we record some simple observations on robust algorithms for HARD CSP(Π). In the weighted minimization version (MIN UNCUT, MIN CNF DELETION and FEEDBACK ARCS SET are minimization versions of some problems we studied in this chapter), where the goal is to *minimize* the weight of *unsatisfied* constraints, hard constraints do not add to the difficulty of the original version as can be seen via the simple trick of converting all hard constraints to soft constraints and giving them large enough weight. Therefore, a constant-factor approximation algorithm for the minimization version will yield a robust satisfiability algorithm for HARD CSP(Π) as well. In general we do not have a reduction in the reverse direction since the robust algorithm might be allowed runtime exponential in ϵ (e.g. $n^{O(\frac{1}{\epsilon})}$) on $(1-\epsilon)$-satisfiable instances. Also, when the approximation factor for the minimization version depends on the instance size (e.g. $O(\sqrt{\log n})$ for MIN CNF DELETION [ACMM05]), we do not automatically get a robust algorithm for the associated CSP with hard constraints.

We have an example, namely HARD 2-SAT, where constant-factor and robust approximations are both hard in the presence of hard constraints, and an example (HARD HORN 2-SAT) where robust approximation is possible but a constant-factor approximation is hard. It is easy to see that HARD LIN-MOD-2 admits an efficient constant factor approximation but not a robust algorithm; its ordinary version (without hard constraints), however, also has the same property. This raises the following interesting question: Is there a CSP which admits an efficient robust satisfiability algorithm, and whose hard version admits a constant-factor approximation algorithm but no robust algorithm? Or must a robust algorithm also exist for the hard version under these conditions?

Chapter 5

Symmetric CSP

5.1 Introduction

Constraint Satisfaction Problems (CSPs) are among the most fundamental and well-studied class of optimization problems. Given a fixed integer k and a predicate $Q \subseteq \{0,1\}^k$, an instance of CSP(Q) *without negation* is specified by a set of variables $X = \{x_1, \ldots, x_n\}$ on the domain $\{0,1\}$ and a set of constraints $\mathcal{C} = \{C_1, \ldots, C_m\}$, where each constraint $C_j = (x_{j,1}, \ldots, x_{j,k})$ is a k-tuple of variables. An assignment $X \to \{0,1\}$ satisfies C_j if $(x_{j,1}, \ldots, x_{j,k}) \in Q$. For an instance of CSP(Q) *with negation*, each constraint C_j is additionally given offsets $(b_{j,1}, \ldots, b_{j,k}) \in \{0,1\}^k$ and is satisfied if $(x_{j,1} \oplus b_{j,1}, \ldots, x_{j,k} \oplus b_{j,k}) \in Q$ where $\oplus$ denotes the addition in $\mathbb{F}_2$. The goal is to find an assignment that satisfies as many constraints as possible.

CSPs contain a large number of famous problems such as MAX SAT (with negation), and MAX CUT / MAX-SET-SPLITTING (without negation) by definition. They have always played a crucial role in the theory of computational complexity, as many breakthrough results such as the NP-completeness of 3SAT, the Probabilistically Checkable Proofs (PCP) theorem, and the Unique Games Conjecture (UGC) study hardness of a certain CSP.

Based on these works, recent works on approximability of CSPs focus on characterizing every CSP according to its approximation resistance. We define *random assignments* to be the class of algorithms that assign $x_i \leftarrow 1$ with probability α independently. A CSP is called *approximation resistant*, if for any $\epsilon > 0$, it is NP-hard to have a $(\rho^* + \epsilon)$-approximation algorithm, where ρ^* is the approximation ratio achieved by the best random assignment. Even assuming the UGC, the complete characterzation of approximation resistance has not been found, and previous works either change the notion of approximation resistance or study a subclass of CSPs to find a characterization, and more general results tend to suggest more complex characterizations.

This chapter considers a natural subclass of CSPs where a predicate Q is *symmetric* — for any permutation $\pi : [k] \to [k]$, $(x_1, \ldots, x_k) \in Q$ if and only if $(x_{\pi(1)}, \ldots, x_{\pi(k)}) \in Q$. Equivalently, for every such Q, there exists $S \subseteq [k] \cup \{0\}$ such that $(x_1, \ldots, x_k) \in Q$ if and only if $(x_1 + \cdots + x_k) \in S$. Let SCSP($S$) denote such a symmetric CSP. While this is a significant restriction, it is a natural one that still captures the following fundamental problems, such as MAX SAT, MAX NOT-ALL-EQUAL-SAT, MAX t-OUT-OF-k-SAT (with negation), and MAX CUT, MAX-SET-SPLITTING, DISCREPANCY MINIMIZATION (without negation). Except the work of Austrin and Håstad [AH13], many works on this line focused CSPs with negation, while we feel that the aforementioned problems without negation have a very natural interpretation as (hyper)graph coloring and are worth studying.

There is a simple sufficient condition to be approximation resistant due to Austrin and Mossel [AM09] with negation, and due to Austrin and Håstad [AH13] without negation. For SCSPs, we show that these simple sufficient conditions can be further simplfied and understood more intuitively, and suggest that they might also be necessary for and thus precisely characterize approximation resistance. We prove it for two natural special cases (which capture all problems mentioned in the last paragraph) for both SCSPs with / without negation, and provide reasons that we believe this is true at least for SCSPs without negation.

5.1.1 Related Work

Given the importance of CSPs and the variety of problems that can be formulated as a CSP, it is a natural task to classify all CSPs according to their computational complexity for some well-defined task. For the task of deciding satisfiability (i.e., finding an assignment that satisfies every constraint if there is one), the work of Schaefer [Sch78] gave a complete characterization on the Boolean domain in 1978.

However, such a classification seems much harder when we study approximability of CSPs. Since the seminal work of Håstad [Hås01], many natural problems have been proven to be approximation resistant. These examples include MAX 3-SAT / MAX 3-LIN (with negation) and MAX 4-SET-SPLITTING (without negation), and for Boolean CSPs of arity 3, putting together the hardness results of [Hås01] with the algorithmic results of Zwick [Zwi98a], it is known that a CSP is approximation resistant if and only if it is implied by parity. However, characterizing approximation resistance of every CSP for larger arity k is a harder task. The Ph.D. thesis of Hast [Has05] is devoted to this task for $k = 4$, and succeeds to classify 354 out of 400 predicates.

The advent of the Unique Games Conjecture (UGC) [Kho02b], though it is not as widely believed as $\mathbf{P} \neq \mathbf{NP}$, revived the hope to classify every CSP according to its approximation resistance. For CSPs with negation, the work of Austrin and Mossel [AM09] gave a simple sufficient condition to be approximation resistant, namely the existence of a balanced pairwise independent distribution that is supported on the satisfying assignments of the predicate. The work of Austrin and Håstad [AH13] proved a similar sufficient condition for CSPs without negation, and that if this condition is not met, this predicate (both with / without negation) is *useful* for some polynomial optimization — for every such Q, there is a k-variate polynomial $p(y_1, \ldots, y_k)$ such that if we are given an instance of CSP(Q) that admits a $(1-\epsilon)$-satisfying assignment, the altered problem, where we change each constraint C_j's payoff from $\mathbb{I}[(x_{j,1} \oplus b_{j,1}, \ldots, x_{j,k} \oplus b_{j,k}) \in Q]$ (where $\mathbb{I}[\cdot]$ is the indicator function) to $p(x_{j,1} \oplus b_{j,1}, \ldots, x_{j,k} \oplus b_{j,k})$, admits an approximation algorithm that does better than any random assignment.

Predicates that don't admit a pairwise independent distribution supported on their satisfying assignments can be expressed as the sign of a quadratic polynomial (see [AH13]). This motivates the study of the approximability of such predicates, though it is known that there are approximation resistant predicates that can be expressed as a quadratic threshold function and thus the sufficient condition of Austrin and Mossel [AM09] is not necessary for approximation resistance. Still this motivates the question of understanding which quadratic threshold functions can be approximated non-trivially.

Cheraghchi, Håstad, Isaksson, and Svensson [CHIS12] studied the simpler case of predicates which are the sign of a linear function with no constant term, obtaining algorithms beating the random assignment threshold of $1/2$ in some special cases. Austrin, Benabbas, and Magen [ABM12] conjecture that every such predicate can be approximated better than a factor $1/2$ and is therefore not approximation resistant. They prove that predicates that are the sign of *symmetric* quadratic polynomials with no constant term are not approximation resistant.

Assuming the UGC, the work of Austrin and Khot [AK13] gave a characterization of approximation resistance for even k-partite CSPs, and Khot, Tulsiani, and Worah [KTW14] gave a characterization of *strong approximation resistance* for general CSPs — strong approximation resistance roughly means hardness of finding an assignment that deviates from the performance of the random assignment in either direction (i.e., it is hard to also find an assignment saisfying a noticeably smaller fraction of constraints than the random assignment). These two works are notable in studying approximation resistance of general CSPs, but their characterizations become more complicated, which they suggest is necessary.

Without the UGC, even the existence of pairwise independent distribution supported on the predicate is not known to be sufficient for approximation resistance. Another line of work shows partial results either by using a stronger condition [Cha13], or by using a restricted model of computation (e.g., Sherali-Adams or Lasserre hierarchy of convex relaxations) [Tul09, BGMT12, BCK15].

5.1.2 Our Results

Our work was initially motivated by a simple observation that for symmetric CSPs, the sufficient condition to be approximation resistant by Austrin and Håstad [AH13] admits a more compact and intuitive two-dimensional description in $\mathbb{R}^2$.

Fix a positive integer k and denote $[k] = \{1, 2, \dots, k\}$. For $s \in [k] \cup \{0\}$, let $P(s) \in \mathbb{R}^2$ be the point defined by $P(s) := (\frac{s}{k}, \frac{s(s-1)}{k(k-1)})$. For any s, $P(s)$ lies on the curve $y = \frac{k}{k-1}x^2 - \frac{x}{k-1}$, which is slightly below the curve $y = x^2$ for $x \in [0, 1]$. Given a subset $S \subseteq [k] \cup \{0\}$, let $P_S := \{P(s) : s \in S\}$ and $\mathsf{conv}(P_S)$ be the convex hull of P_S. For symmetric CSPs, the condition of Austrin and Håstad depends on whether this convex hull intersects a certain curve or a point.

For SCSP(S) without negation, the condition becomes whether $\mathsf{conv}(P_S)$ intersects the curve $y = x^2$. If we let s_{min} and s_{max} be the minimum and maximum number in S respectively, by convexity of $y = \frac{k}{k-1}x^2 - \frac{x}{k-1}$, it is equivalent to that the line passing through $P(s_{min})$ and $P(s_{max})$ and $y = x^2$ intersect, which is again equivalent to (see Lemma 5.4.4)

$$\frac{(s_{max} + s_{min} - 1)^2}{k - 1} \geq \frac{4 s_{max} s_{min}}{k}. \tag{5.1}$$

A simple calculation shows that the above condition is implied by $(s_{max} - s_{min}) \geq \sqrt{2(s_{max} + s_{min})}$ which in turn holds if $(s_{max} - s_{min}) \geq 2\sqrt{k}$. This means that SCP($S$) is approximation resistant unless s_{min} and s_{max} are very close. See Figure 5.1 for an example.

For general CSP(Q) with $Q \subseteq \{0, 1\}^k$, Q is *positively correlated* if there is a distribution μ supported on Q and $p, \rho \in [0, 1]$ with $\rho \geq p^2$ such that $\Pr_\mu[x_i = 1] = p$ for every $i \in [k]$ and $\Pr_\mu[x_i = x_j = 1] = \rho$ for every $1 \leq i < j \leq k$. Austrin and Hastad [AH13] proved that CSP(Q) is approximate resistant if Q is positively correlated. For SCSP(S), Lemma 5.4.3 shows that (5.1) holds if and only if Q_S is positively correlated

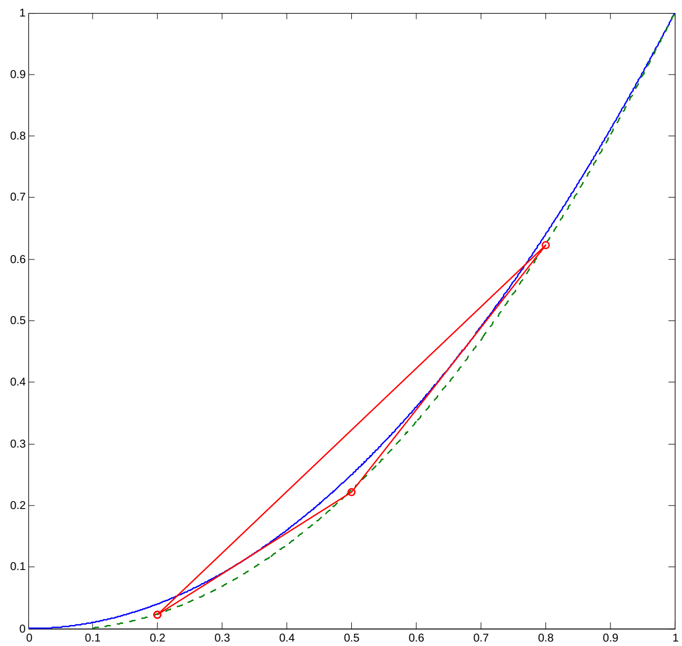

Figure 5.1: An example when $k = 10$ and $S = \{2, 5, 8\}$. The solid curve is $y = x^2$ and the dashed curve is $y = \frac{k}{k-1}x^2 - \frac{x}{k-1}$, where all $P(s)$ lie. In this case the triangle $\mathsf{conv}(P_S)$ intersects $y = x^2$, so SCSP(S) is approximation resistant.

where $Q_S := \{(x_1, \ldots, x_k) \in \{0,1\}^k : x_1 + \cdots + x_k \in S\}$.

We conjecture that this simple condition completely characterizes approximation resistance of symmetric CSPs without negation. Note that we exclude the cases where S contains 0 or k, since without negation, a trivial deterministic strategy to give the same value to every variable satisfies every constraint.

Conjecture 5.1.1. *For $S \subseteq [k-1]$, SCSP(S) without negation is approximation resistant if and only if* (5.1) *holds.*

The hardness claim, the "if" part, is currently proved only under the UGC, but our focus is on the algorithmic claim that the violation of (5.1) leads to an approximation algorithm that outperforms the best random assignment. Even though we were not formally able to prove Conjecture 5.1.1, we explain the rationale behind the conjecture and we prove it for the following two natural special cases in Section 5.2:

1. S is an *interval*: S contains every integer from s_{min} to s_{max}.

2. S is *even*: $s \in S$ if and only if $k - s \in S$.

Theorem 5.1.1. *If $S \subseteq [k-1]$ and S is either an interval or even, SCSP(S) without negation is approximation resistant if and only if* (5.1) *holds (the hardness claim, i.e., the "if" part, is under the Unique Games conjecture).*

For SCSP(S) with negation, the analogous condition is whether $\mathsf{conv}(P_S)$ contains a single point $(\frac{1}{2}, \frac{1}{4})$. This is equivalent to that $Q_S := \{(x_1, \ldots, x_k) \in \{0,1\}^k : x_1 + \cdots + x_k \in S\}$ is *balanced pairwise independent* (see Section 5.4). While it is tempting to pose a conjecture similar to Conjecture 5.1.1, we refrain from doing so due to the lack of evidence compared to the case without negation. However, we prove the following theorem which shows that the analogous characterization works at least for the two special cases introduced above.

Theorem 5.1.2. *If $S \subset [k] \cup \{0\}$ and S is either an interval or even, SCSP(S) with negation is approximation resistant if and only if $\mathsf{conv}(P_S)$ contains $(\frac{1}{2}, \frac{1}{4})$ (the hardness claim, i.e., the "if" part, is under the Unique Games conjecture).*

5.1.3 Techniques

We mainly focus on SCSPs without negation, and briefly sketch why the violation of (5.1) might lead to an approximation algorithm that outperforms the best random assignment. Let α^* be the probability that the best random assignment uses, and ρ^* be the expected fraction of constraints satisfied by it. Our algorithms follow the following general framework: sample correlated random variables $X_1, \ldots, X_n$, where each X_i lies in $[-\alpha^*, 1-\alpha^*]$, and independently round $x_i \leftarrow 1$ with probability $\alpha^* + X_i$.

Fix one constraint $C = (x_1, \ldots, x_k)$ (for SCSPs with negation, additionally assume that offsets are all 0). Using symmetry, the probability that it is satisfied by the above strategy can be expressed as

$$\rho^* + \sum_{l=1}^{k} c_l \mathbb{E}_{I \in \binom{[k]}{l}} [\prod_{i \in I} X_i].$$

For some coefficients $\{c_l\}_{l \in [k]}$. These coefficients c_l can be expressed by the following two ways.

- Let $f(\alpha) : [0,1] \to [0,1]$ the be probability that a constraint is satisfied by a random assignment with probability α. Then c_l is proportional to $f^{(l)}(\alpha^*)$, the l'th derivative of f evaluated at α^*.

- Let $Q = \{(x_1, \ldots, x_k) \in \{0,1\}^k : (x_1 + \cdots + x_k) \in S\}$ be the predicate associated with S. When $\alpha^* = \frac{1}{2}$, c_l is proportional to the Fourier coefficient $\hat{Q}(T)$ with $|T| = l$.

Given this observation, α^* for SCSPs without negation has nice properties since it should be a global maximum in the interval $[0,1]$. In particular, it should be a local maximum so that $c_1 = f'(\alpha) = 0$ and $c_2, f''(\alpha) \leq 0$. By modifying an algorithm by Austrin and Håstad [AH13], we prove that we can sample $X_1, \ldots, X_n$ such that the average second moment $\mathbb{E}[X_i X_j]$ is strictly negative if (5.1) does not hold. By scaling X_i's so that the product of at least three X_i's becomes negligible, this idea results in an approximation algorithm that outperforms the best random assignment, except in the degenerate case where $c_2 = f''(\alpha^*) = 0$ even though α^* is a local maximum. This is the main rationale behind Conjecture 5.1.1 and we elaborate this belief more in Section 5.2. It is notable that our conjectured characterization for the case without negation only depends on the minimum and the maximum number in S, while α^* also depends on other elements.

For SCSPs with negation where α^* is fixed to be $\frac{1}{2}$, the situation becomes more complicated since c_1 and $f'(\alpha)$ are not necessarily zero and there are many ways that $\mathsf{conv}(P_S)$ does not contain $(\frac{1}{2}, \frac{1}{4})$ (in the case of SCSPs without negation, the slope of the line separating $\mathsf{conv}(P_S)$ and $y = x^2$ is always positive, but it is not the case here). Therefore, a complete characterization requires understanding interactions among c_1, c_2, and the separating line. We found that the somewhat involved method of Austrin, Benabbas, and Magen [ABM12] gives a way to sample these $X_1, \ldots, X_n$ with desired first and second moments to prove our results when S exhibits additional special structures, but believe that a new set of ideas are required to give a complete characterization.

5.1.4 Organization

In Section 5.2, we study SCSPs without negation. We further elaborate our characterization in Section 5.2.1, and provide an algorithm in Section 5.2.2. We study SCSPs with negation in Section 5.3.

5.2 Symmetric CSPs without Negation

5.2.1 A 2-dimensional Characterization

Fix k and $S \subseteq [k-1]$. Our conjectured condition to be approximation resistant is that $\mathsf{conv}(P_S)$ intersects the curve $y = x^2$, which is equivalent to (5.1). Austrin and Håstad [AH13] proved that this simple condition is sufficient to be approximation resistant.

Theorem 5.2.1 ([AH13])**.** *Let $S \subseteq [k-1]$ be such that* (5.1) *holds. Then, assuming the Unique Games Conjecture, SCSP(S) without negation is approximation resistant.*

They studied general CSPs and their condition is more complicated than stated here. See Section 5.4 to see how it is simplified for SCSPs. We conjecture that for SCSPs, this condition is indeed equivalent to approximation resistance.

Conjecture 5.2.1 (Restatement of Conjecture 5.1.1)**.** *For $S \subseteq [k-1]$, SCSP(S) without negation is approximation resistant if and only if* (5.1) *holds.*

To provide our rationale behind the conjecture, we define the function $f : [0,1] \to [0,1]$ to be the probability that one constraint is satisfied by the random assignment that gives $x_i \leftarrow 1$ independently with probability α.

$$f(\alpha) = \sum_{s \in S} \binom{k}{s} \alpha^s (1-\alpha)^{k-s}$$

Let $\alpha^* \in [0,1]$ be a value that maximizes $f(\alpha)$, and $\rho^* := f(\alpha^*)$. There might be more than one α with $f(\alpha) = \rho^*$. In Section 5.2.2, we prove that S is not approximation resistant if there exists one such α^* with a negative second derivative.

Theorem 5.2.2. *$S \subseteq [k-1]$ be such that* (5.1) *does not hold and there exists $\alpha^* \in [0,1]$ such that $f(\alpha^*) = \rho^*$ and $f''(\alpha^*) < 0$. Then, there is a randomized polynomial time algorithm for SCSP(S) that satisfies strictly more than ρ^* fraction of constraints in expectation.*

Since $f(0) = f(1) = 0 < \rho^*$, every $\alpha \in [0,1]$ with $f(\alpha) = \rho^*$ must be a local maximum, so it should have $f'(\alpha) = 0$ and $f''(\alpha) \le 0$. If α is a local maximum, $f''(\alpha) = 0$ also implies $f'''(\alpha) = 0$, so ruling out this degeneracy at a global maximum gives the complete characterization!

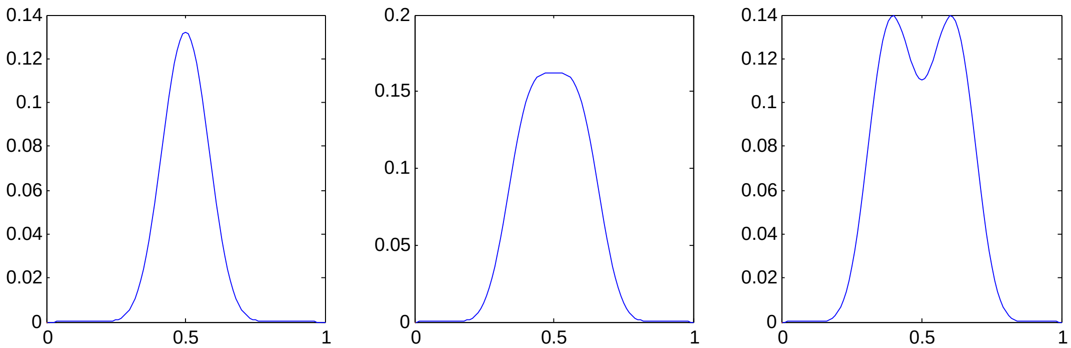

Figure 5.2: Examples for $k = 36$. **Left**: $S = \{18\}$, (5.1) is not satisfied, unimodal with $\alpha^* = \frac{1}{2}$, $f''(\frac{1}{2}) < 0$. **Middle**: $S = \{15, 21\}$, (5.1) is satisfied with equality, unimodal with $\alpha^* = \frac{1}{2}$, but $f''(\frac{1}{2}) = 0$. **Right**: $S = \{14, 22\}$, (5.1) is satisfied with slack, bimodal with two α^*, but $f''(\alpha^*) < 0$.

Ruling out this degeneracy at a global maximum does not seem to be closely related to general shape of $f(\alpha)$ or S. It might still hold even if $f(\alpha)$ has multiple global maxima, or S satisfies (5.1) so that SCSP(S) is approximation resistant.

However, examples in Figure 5.2 led us to believe that the condition (5.1) is also related to general shape of f. When S contains two numbers l and r with $l + r = k$, as two numbers become far apart, f becomes unimodal to bimodal, and the transition happens exactly when (5.1) starts to hold. Furthermore, the degenerate case $f'(\alpha^*) = f''(\alpha^*) = 0$ happens when (5.1) holds with equality. Intuitively, when two numbers l and r are far apart, the best strategy is to focus on only one of them (i.e. $\alpha^* \approx \frac{l}{k}$ or $\frac{r}{k}$), so f is bimodal. If l and r are close enough, it is better to target in the middle to satisfy both l and r, so f becomes unimodal with a large negative curvature at α^*.

Having more points between l and r seems to strengthen the above intuition, and removing the assumption that $l + r = k$ only seems to add algebraic complication without hurting the intuition. Thus, we propose the following stronger conjecture that implies Conjecture 5.1.1.

Conjecture 5.2.2. *If* (5.1) *does not hold,* $f(\alpha)$ *is unimodal in* $[0, 1]$ *with the unique maximum at* α^**, and* $f''(\alpha^*) < 0$.

While we are unable to formally prove Conjecture 5.2.2 for every S, we establish it for the case when S is either an interval (Section 5.2.3) or even (Section 5.2.4), thus proving Theorem 5.1.1.

5.2.2 Algorithm

Let $\alpha^* \in [0,1]$ be such that $f(\alpha^*) = \rho^*$ and $f''(\alpha^*) < 0$. Furthermore, suppose that S does not satisfy (5.1). We give a randomized approximation algorithm which is guaranteed to satisfy strictly more than ρ^* fraction of constraints in expectation, proving Theorem 5.2.2. Let $D := D(k)$ be a large constant determined later. Our strategy is the following.

1. Sample $X_1, \ldots, X_n$ from some correlated multivariate normal distribution where each X_i has mean 0 and variance at most σ^2 for some $\sigma := \sigma(k)$.

2. For each $i \in [n]$, set

$$X'_i = \begin{cases} -D\alpha^* & \text{if } X_i < -D\alpha^* \\ D(1-\alpha^*) & \text{if } X_i > D(1-\alpha^*) \\ X_i & \text{otherwise} \end{cases}$$

   so that $\alpha^* + \frac{X'_i}{D}$ is always in $[0,1]$.

3. Set $x_i \leftarrow 1$ independently with probability $\alpha^* + \frac{X'_i}{D}$.

Fix one constraint C and suppose that $C = (x_1, \ldots, x_k)$. We consider a multivariate polynomial

$$g(y_1, \ldots, y_k) := \sum_{T \subseteq [k], |T| \in S} \prod_{i \in T} (\alpha^* + \frac{y_i}{D}) \prod_{i \in [k] \setminus T} (1 - \alpha^* - \frac{y_i}{D}).$$

$g(X'_1, \ldots, X'_k)$ is equal to the probability that the constraint C is satisfied. By symmetry, for any $1 \le i_1 < \cdots < i_l \le k$, the coefficient of a monomial $y_{i_1} y_{i_2} \ldots y_{i_l}$ only depends on l. Let c_l be this coefficient.

Lemma 5.2.3. $c_l = \frac{(k-l)!}{k! D^l} f^{(l)}(\alpha^*)$.

Proof. Note that $g(y, y, \ldots, y) = f(\alpha^* + \frac{y}{D})$, which has the Taylor expansion

$$\sum_{l=0}^{k} \frac{f^{(l)}(\alpha^*)}{l!} (\frac{y}{D})^l.$$

Since g is multilinear, by symmetry, the coefficient of a monomial $y_{i_1} y_{i_2} \ldots y_{i_l}$ in $g(y_1, \ldots, y_k)$ is equal to the coefficient of y^l in $f(\alpha^* + \frac{y}{D})$ divided by $\binom{k}{l}$, which is $c_l = \frac{(k-l)!}{k! D^l} f^{(l)}(\alpha^*)$. $\square$

We analyze the overall performance of this algorithm. First we prove the following technical lemma about Gaussians.

Lemma 5.2.4. *Let $Y_1, \ldots, Y_l$ be sampled from a multivariate normal distribution where each Y_i has mean 0 and variance at most σ^2. Let $Y'_1, \ldots, Y'_l$ be such that*

$$Y'_i = \begin{cases} Y_i & \text{if } |Y_i| \le D \\ D & \text{if } Y_i > D \\ -D & \text{if } Y_i < -D \end{cases}$$

Then, for large enough D,

$$|\mathbb{E}[\prod_{i=1}^{l} Y_i] - \mathbb{E}[\prod_{i=1}^{l} Y'_i]| \le 2^l \cdot \sigma^l \cdot l! \cdot e^{-D/l}.$$

Proof. For each $i \in [l]$, let $Y''_i = Y'_i - Y_i$. Take D large enough so that

$$\mathbb{E}[|Y''_i|^l] = 2\int_{y=D}^{\infty} (y-D)^l \phi(y) \le 2\int_{y=D}^{\infty} y^l \phi(y) \le e^{-D}.$$

Also each Y_i, a normal random variable with mean 0 and variance σ, satisfies $\mathbb{E}[|Y_i|^l] \le \sigma^l \cdot l!$. We have

$$\begin{aligned}
& |\mathbb{E}[\prod_{i=1}^{l} Y_i] - \mathbb{E}[\prod_{i=1}^{l} Y'_i]| \\
= \quad & |\sum_{T\subseteq[l], T\neq[l]} \mathbb{E}[\prod_{i\in T} Y_i \prod_{i\notin T} Y''_i]| \\
\le \quad & \sum_{T\subseteq[l], T\neq[l]} \prod_{i\in T} (\mathbb{E}[|Y_i|^l])^{1/l} \prod_{i\notin T} (\mathbb{E}[|Y''_i|^l])^{1/l} \quad \text{By Generalized Hölder's inequality [Che01]} \\
\le \quad & 2^l \cdot \sigma^l \cdot l! \cdot e^{-D/l}.
\end{aligned}$$

□

Let $\mathcal{D}_l$ be the distribution on $\binom{[n]}{l}$ where we sample a constraint C uniformly at random, sample l distinct variables from $\binom{C}{l}$, and output their indices. We prove the following lemma, which implies that by taking large D, the effect of truncation from X_i to X'_i and the contribution of monomials of degree greater than two become small.

Lemma 5.2.5. *The expected fraction of constraints satisfied by the above algorithm is at least*

$$\rho^* + c_2 \binom{k}{2} \mathbb{E}_{(i,j)\sim\mathcal{D}_2}[X_i X_j] - O_k(\frac{1}{D^3}) = \rho^* + \frac{f''(\alpha^*)}{2D^2} \mathbb{E}_{(i,j)\sim\mathcal{D}_2}[X_i X_j] - O_k(\frac{1}{D^3}),$$

where $O_k(\cdot)$ is hiding constants depending on k.

Proof. For any $s \in \{1, \dots, k-1\}$, the function $f_s(\alpha) = \binom{k}{s}\alpha^s(1-\alpha)^{k-s}$ is unimodal in $[0,1]$ with the maximum at $\alpha = \frac{s}{k}$. As long as S does not contain 0 or k, $f(\alpha) = \sum_{s \in S} f_s(\alpha)$ cannot have a local maximum in $[0, \frac{1}{k})$ and $(1-\frac{1}{k}, 1]$, so we can assume that $\alpha^* \in [\frac{1}{k}, 1-\frac{1}{k}]$.

For any $1 \le l \le k$ and $1 \le i_1 < \cdots < i_l \le k$, we apply Lemma 5.2.4 (set $D \leftarrow \frac{D}{k}$),

$$|\mathbb{E}[\prod_{j=1}^{l} X_{i_j}] - \mathbb{E}[\prod_{j=1}^{l} X'_{i_j}]| \le 2^l \cdot \sigma^l \cdot l! \cdot e^{-D/kl}.$$

If we expand $f(\alpha) = \sum_{l=0}^{k} a_l \alpha^l$, each coefficient a_l has magnitude at most 2^k, which means that $|f^{(l)}(\alpha^*)|$ is bounded by $k2^k k!$. Therefore, any $|c_l|$ is at most $k2^k k!$. Let c_{max} be this quantity. Summing over this error for all monomials, the probability that a fixed constraint $C = \{x_1, \dots, x_k\}$ is satisfied is

$$\begin{aligned}
\mathbb{E}[g(X'_1, \dots, X'_k)] \ge \quad & \mathbb{E}[g(X_1, \dots, X_k)] - c_{max} \cdot 2^{2k} \cdot \sigma^k \cdot k! \cdot e^{-D/k^2} \\
= \quad & \rho^* + \sum_{l=1}^{k} c_l \sum_{1 \le i_1 < \cdots < i_l \le k} X_{i_1} X_{i_2} \dots X_{i_l} - O_k(e^{-D/k^2}) \\
= \quad & \rho^* + \sum_{l=1}^{k} c_l \sum_{1 \le i_1 < \cdots < i_l \le k} X_{i_1} X_{i_2} \dots X_{i_l} - O_k(e^{-D/k^2})
\end{aligned}$$

Averaging over m constraints, the expected fraction of satisfied constraints is at least

$$\begin{aligned}
& \rho^* + \sum_{l=1}^{k} c_l \binom{k}{l} \mathbb{E}_{(i_1,\dots,i_l)\sim\mathcal{D}_l}[X_{i_1} \dots X_{i_l}] - O_k(e^{-D/k^2}) \\
= \quad & \rho^* + c_2 \binom{k}{2} \mathbb{E}_{(i_1,i_2)\sim\mathcal{D}_2}[X_{i_1} X_{i_2}] + \sum_{l=3}^{k} c_l \binom{k}{l} \mathbb{E}_{(i_1,\dots,i_l)\sim\mathcal{D}_l}[X_{i_1} \dots X_{i_l}] - O_k(e^{-D/k^2}) \\
= \quad & \rho^* + c_2 \binom{k}{2} \mathbb{E}_{(i_1,i_2)\sim\mathcal{D}_2}[X_{i_1} X_{i_2}] - O_k(\frac{1}{D^3}) \\
= \quad & \rho^* + \frac{f''(\alpha^*)}{2D^2} \mathbb{E}_{(i,j)\sim\mathcal{D}_2}[X_i X_j] - O_k(\frac{1}{D^3}),
\end{aligned}$$

where the first equality follows from the fact that $\mathbb{E}[X_i] = 0$ for all i. Recall that $c_l = \frac{(k-l)!}{k!D^l} f^{(l)}(\alpha^*)$ so that $|c_l| = O_k(\frac{1}{D^l})$. $\square$

Therefore, if we have a way to sample $X_1, \ldots, X_n$ such that each X_i has mean 0 and variance at most σ^2, and $\mathbb{E}_{(i,j)\sim\mathcal{D}_2}[X_i X_j] < -\delta$ for some $\delta := \delta(k) > 0$, taking D large enough ensures that the algorithm satisfies strictly more than ρ^* fraction of constraints. We now show how to do such a sampling. Our basic intuition is that if S does not satisfy (5.1), there is no positively correlated distribution supported by $Q_S := \{(x_1, \ldots, x_k) : x_1 + \cdots + x_k\}$, which helps to find a distribution with negative correlation.

We assume that for some $\epsilon := \epsilon(k) > 0$, the given instance admits a solution that satisfies $(1-\epsilon)$ fraction of constraints. Otherwise, the random assignment with probability α^* guarantees the approximation ratio of $\frac{\rho^*}{1-\epsilon}$. The following lemma completes the proof of Theorem 5.2.2.

Lemma 5.2.6. *Suppose that S does not satisfy* (5.1)*. For sufficiently small $\epsilon, \delta > 0$ and sufficiently large σ all depending only on k, given an instance of SCSP(S) where $(1-\epsilon)$ fraction of constraints are simultaneously satisfiable, it is possible to sample $X_1, \ldots, X_n$ from a multivariate normal distribution such that each X_i has mean 0 and variance bounded by σ^2, and $\mathbb{E}_{(i,j)\sim\mathcal{D}_2}[X_i X_j] < -\delta$.*

Proof. Recall that (5.1) is equivalent to the fact that the line ℓ passing $P(s_{min})$ and $P(s_{max})$ intersects the curve $y = x^2$. Let a be the value such that the vector $(a, -1)$ is orthogonal to ℓ. a is strictly positive since ℓ has a positive slope. If ℓ and $y = x^2$ do not intersect, there is a line with the same slope as ℓ that strictly separates $y = x^2$ and $\{P(s) : s \in S\}$ — in other words, there exists $c \in \mathbb{R}$ such that

- $ax - y + c > \gamma > 0$ for $(x, y) \in \{P(s) : s \in S\}$.
- $ax - x^2 + c < 0$ for any $x \in \mathbb{R} \Rightarrow c < \frac{-a^2}{4}$.

Consider a constraint $C = (x_1, \ldots, x_k)$. Since $(\mathbb{E}_{i\in[k]}[x_i], \mathbb{E}_{i\neq j\in[k]}[x_i x_j]) = P(x_1 + \cdots + x_k)$, if C is satisfied,

$$a\mathbb{E}_{i\in[k]}[x_i] - \mathbb{E}_{i\neq j\in[k]}[x_i x_j] + c > \gamma.$$

Let

$$\eta := - \min_{x_1,\ldots,x_k\in\{0,1\}} \Big(a\mathbb{E}_{i\in[k]}[x_i] - \mathbb{E}_{i\neq j\in[k]}[x_i x_j] + c \Big).$$

We solve the following semidefinite programm (SDP):

$$\begin{array}{ll} \text{maximize} & a\mathbb{E}_{i\in\mathcal{D}_1}[\langle v_0, v_i\rangle] - \mathbb{E}_{i,j\in\mathcal{D}_2}[\langle v_i, v_j\rangle] + c \\ \text{subject to} & ||v_0|| = 1 \\ & \langle v_i, v_0\rangle = ||v_i||^2 \quad \text{for all } i \in [n] \end{array}$$

Note that $\langle v_i, v_0 \rangle = ||v_i||^2$ implies $||v_i|| \leq 1$. For any assignment to $x_1, \ldots, x_n$, setting $v_i = x_i v_0$ satisfies that $x_i = \langle v_0, v_i \rangle$ and $x_i x_j = \langle v_i, v_j \rangle$. Since at least $(1-\epsilon)$ fraction of constraints can be simultaneously satisfied, the optimum of the above SDP is at least $(1-\epsilon)\gamma - \epsilon\eta$. Given $\gamma > 0$ and η, take sufficiently small $\epsilon, \delta > 0$ such that $(1-\epsilon)\gamma - \epsilon\eta = \delta$. There are finitely many S (thus γ and η) for each k, so ϵ and δ can be taken to depend only on k. Given vectors $v_0, v_1, \ldots, v_n$, we sample $X_1, \ldots, X_n$ by the following simple procedure:

1. Sample a vector g whose coordinates are independent standard normal.
2. Let $X_i = \langle g, v_i - \frac{a}{2} v_0 \rangle$.

It is clear that $\mathbb{E}[X_i] = 0$ for each i, and $\mathbb{E}[X_i^2] = ||v_i - \frac{a}{2}v_0||^2 \leq (a+1)^2 + 1$, so taking $\sigma := \sigma(k)$ large enough ensures that the variance of each X_i is bounded by σ^2. We now compute the second moment.

$$\begin{aligned}
& \mathbb{E}_{i,j \sim \mathcal{D}_2}[X_i X_j] \\
= \; & \mathbb{E}_{i,j \sim \mathcal{D}_2}[\langle v_i - \frac{a}{2} v_0, v_j - \frac{a}{2} v_0 \rangle] \\
= \; & \mathbb{E}_{i,j \sim \mathcal{D}_2}[\langle v_i, v_j \rangle] - a \mathbb{E}_{i \in \mathcal{D}_1}[\langle v_i, v_0 \rangle] + \frac{a^2}{4} \\
< \; & \mathbb{E}_{i,j \sim \mathcal{D}_2}[\langle v_i, v_j \rangle] - a \mathbb{E}_{i \in \mathcal{D}_1}[\langle v_i, v_0 \rangle] - c \\
\leq \; & -((1-\epsilon)\gamma - \epsilon\eta) = -\delta,
\end{aligned}$$

where the first inequality follows from $c < -\frac{a^2}{4}$ and the second follows from the optimality of our SDP. $\square$

5.2.3 Case of Interval S

We study properties of $f(\alpha)$ when S is an interval — $S = \{s_{min}, s_{min}+1, \ldots, s_{max} - 1, s_{max}\}$, and prove Conjecture 5.2.2 for this case. One notable fact is that as long as S is an interval, the conclusion of Conjecture 5.2.2 is true even if S does satisfy (5.1) and becomes approximation resistant.

Lemma 5.2.7. *Suppose $S \subseteq [k-1]$ is an interval. Then, $f(\alpha)$ is unimodal in $[0,1]$ with the unique maximum at α^* and $f''(\alpha^*) < 0$.*

Proof. Let $l := s_{min}$ and $r = s_{max}$. Given

$$f(\alpha) = \sum_{s=l}^{r} \binom{k}{s} \alpha^s (1-\alpha)^{k-s}$$

and

$$f'(\alpha) = \sum_{s=l}^{r} \binom{k}{s} \Big(s\alpha^{s-1}(1-\alpha)^{k-s} - (k-s)\alpha^s(1-\alpha)^{k-s-1} \Big),$$

since $\binom{k}{s}(k-s) = \binom{k}{s+1}(s+1)$, we have

$$f'(\alpha) = \binom{k}{l} l\alpha^{l-1}(1-\alpha)^{k-l} - \binom{k}{r}(k-r)\alpha^r(1-\alpha)^{k-r-1}.$$

If $0 < \alpha < 1$, setting $\beta := \frac{\alpha}{1-\alpha}$ gives a unique non-zero solution to $f'(\beta) = 0$. This proves the unimodality. For the second derivative,

$$\begin{aligned} f''(\alpha) =& \binom{k}{l} l(l-1)\alpha^{l-2}(1-\alpha)^{k-l} - \binom{k}{l} l(k-l)\alpha^{l-1}(1-\alpha)^{k-l-1} + \\ & \binom{k}{r}(k-r)(k-r-1)\alpha^r(1-\alpha)^{k-r-2} - \binom{k}{r} r(k-r)\alpha^{r-1}(1-\alpha)^{k-r-1} \\ =& \binom{k}{l} l\alpha^{l-2}(1-\alpha)^{k-l-1}\Big((l-1)(1-\alpha) - (k-l)\alpha \Big) + \\ & \binom{k}{r}(k-r)\alpha^{r-1}(1-\alpha)^{k-r-2}\Big((k-r-1)\alpha - r(1-\alpha) \Big). \end{aligned}$$

Since $\frac{l-1}{k-1} < \frac{l}{k} \le \alpha^* \le \frac{r}{k} < \frac{r}{k-1}$,

$$(l-1)(1-\alpha^*) - (k-l)\alpha^* = (l-1) - (k-1)\alpha^* < 0$$

and

$$(k-r-1)\alpha^* - r(1-\alpha^*) = (k-1)\alpha^* - r < 0,$$

so that $f''(\alpha^*) < 0$. $\square$

5.2.4 Case of Even S

We study properties of $f(\alpha)$ when S is even — $s \in S$ if and only if $k - s \in S$, and prove Conjecture 5.2.2 for this case. We first simplify (5.1) for this setting. If we let $l := s_{min}$ and $r := s_{max} = k - l$, (5.1) is equivalent to

$$\frac{(l+r-1)^2}{k-1} \ge \frac{4lr}{k} \quad \Leftrightarrow \quad k(k-1) \ge 4lr \quad \Leftrightarrow \quad (r-l)^2 \ge k.$$

Therefore, (5.1) is equivalent to

$$r - l \ge \sqrt{k} \tag{5.2}$$

Lemma 5.2.8. *Suppose $S \subseteq [k-1]$ is even. If* (5.2) *does not hold, $f(\alpha)$ is unimodal in $[0,1]$ with the unique maximum at $\alpha^* = \frac{1}{2}$ and $f''(\alpha^*) < 0$.*

Proof. Given a even S, let $S_1 = \{s \in S : s \le k/2\}$. When we write f_S to denote the dependence of f on S, we can decompose $f_S(\alpha) = \sum_{s \in S_1} f_{\{s,k-s\}}(\alpha)$, so the following claim proves the lemma. □

Claim 5.2.9. *Let $l \le \frac{k}{2}$ and $r = k - l$ such that $r - l < \sqrt{k} \Leftrightarrow k(k-1) < 4lr$. Let $S = \{l, r\}$. f is unimodal with the unique maximum at $\frac{1}{2}$, and $f''(\frac{1}{2}) < 0$.*

Proof. Note that f is symmetric around $\alpha = 1/2$. If there exists a local maximum at $\alpha' \in (0, 1/2)$, f also has a local maximum at $(1-\alpha')$ with the same value, so there must exist a local minimum in $(\alpha', 1-\alpha')$. In particular, there is $\alpha \in (\alpha', 1-\alpha')$ such that $f'(\alpha) = 0$ and $f''(\alpha) \ge 0$. We prove that such an α cannot exist.

$$
\begin{aligned}
& f'(\alpha) = 0 \\
\Leftrightarrow \quad & \binom{k}{l}\alpha^{l-1}(1-\alpha)^{r-1}(l-k\alpha) + \binom{k}{r}\alpha^{r-1}(1-\alpha)^{l-1}(r-k\alpha) = 0 \\
\Leftrightarrow \quad & \frac{\binom{k}{l}\alpha^{l-1}(1-\alpha)^{r-1}}{\binom{k}{r}\alpha^{r-1}(1-\alpha)^{l-1}} = \frac{\binom{k}{l}(1-\alpha)^{r-l}}{\binom{k}{r}\alpha^{r-l}} = -\frac{(k\alpha - r)}{(k\alpha - l)}
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& f''(\alpha) \ge 0 \\
\Leftrightarrow \quad & \frac{\binom{k}{l}(1-\alpha)^{r-l}}{\binom{k}{r}\alpha^{r-l}} \ge -\frac{r(r-1)(1-\alpha)^2 - 2rl\alpha(1-\alpha) + l(l-1)\alpha^2}{l(l-1)(1-\alpha)^2 - 2rl\alpha(1-\alpha) + r(r-1)\alpha^2}.
\end{aligned}
$$

By symmetry, we can assume $\alpha \ge \frac{1}{2}$, so that $(k\alpha - l) \ge 0$ and $l(l-1)(1-\alpha)^2 - 2rl\alpha(1-\alpha) + r(r-1)\alpha^2 \ge 0$.

$$
\begin{aligned}
& \frac{(k\alpha - r)}{(k\alpha - l)} \le \frac{r(r-1)(1-\alpha)^2 - 2rl\alpha(1-\alpha) + l(l-1)\alpha^2}{l(l-1)(1-\alpha)^2 - 2rl\alpha(1-\alpha) + r(r-1)\alpha^2} \\
\Leftrightarrow \quad & (k\alpha - r)(l(l-1)(1-\alpha)^2 - 2rl\alpha(1-\alpha) + r(r-1)\alpha^2) \\
& \le (k\alpha - l)(r(r-1)(1-\alpha)^2 - 2rl\alpha(1-\alpha) + l(l-1)\alpha^2) \\
\Leftrightarrow \quad & \alpha^2(l^3 - r^3 - (l^2 - r^2) + rl(l-r) - 2k(l^2 - r^2) + 2k(l-r)) + \alpha(k(l^2 - r^2) - k(l-r)) - rl(l - \\
\Leftrightarrow \quad & \alpha^2(-k^2 + k) + \alpha(k^2 - k) - rl \ge 0 \qquad \text{divide by } (l-r) \text{ and use } l + r = k
\end{aligned}
$$

However, $\alpha^2(-k^2+k)+\alpha(k^2-k)-rl$ has a negative leading coefficient and its discriminant is

$$(k^2-k)^2-4rl(k^2-k)=(k^2-k)(k^2-k-4rl)<0$$

by the assumption of the claim. □

We do not formally prove the converse, but Figure 5.2 shows examples where it is tight. When (5.2) holds with equality, f still has the unique local maximum at $\frac{1}{2}$ but $f''(\frac{1}{2})=0$, and even when (5.2) holds with small slack, two local maxima start to appear. This phenomenon is one of the main reasons that we pose Conjecture 5.2.2. Though we were not able to formally prove for the general case, we believe that the violation of (5.1) not only allows us to sample random variables with desired second moments but also ensures that $f(\alpha)$ is a nice unimodal curve.

5.3 Approximability of Symmetric CSPs with Negation

Fix k and $S\subset[k]\cup\{0\}$. In this section, we consider SCSP(S) with negation and prove Theorem 5.1.1. Note that in this section we allow S to contain 0 or k. For example, famous MAX 3-SAT is 3-SCSP($\{1,2,3\}$). We still exclude the trivial case $S=[k]\cup\{0\}$.

The condition we are interested in is whether $\mathsf{conv}(P_S)$ contains $(\frac{1}{2},\frac{1}{4})$. In SCSPs with negation, the sufficient condition of Austrin and Mossel on general CSPs to be approximation resistant becomes equivalent to it. See Appendix 5.4 to see the equivalence.

Theorem 5.3.1 ([AH13]). *Fix k and let $S\subset[k]\cup\{0\}$ be such that $\mathsf{conv}(P_S)$ contains $(\frac{1}{2},\frac{1}{4})$. Then, assuming the Unique Games Conjecture, SCSP(S) with negation is approximation resistant.*

On the other hand, we now show that the algorithm of Austrin et al. [ABM12], which is inspired by Hast [Has05], can be used to show that if S is an interval or even and $\mathsf{conv}(P_S)$ does not contain $(\frac{1}{2},\frac{1}{4})$, SCSP($S$) is not approximation resistant.

Let $f:\{0,1\}^k\to\{0,1\}$ be the function such that $f(x_1,\ldots,x_k)=1$ if and only if $(x_1+\cdots+x_k)\in S$. Define the inner product of two functions as $\langle f,g\rangle=\mathbb{E}_{x\in\{0,1\}^k}[f(x)g(x)]$, and for $T\subseteq[k]$, let $\chi_T(x_1,\ldots,x_k)=\prod_{i\in T}(-1)^{x_i}$. It is well known that $\{\chi_T\}_{T\subseteq[k]}$ form an orthonormal basis and every function has a unique *Fourier expansion* with respect to this basis,

$$f=\sum_{T\subseteq[k]}\hat{f}(T)\chi_T,\qquad \hat{f}(T):=\langle f,\chi_T\rangle.$$

Define

$$f^{=d}(x) = \sum_{|T|=d} \hat{f}(S)\chi_T(x).$$

The main theorem of Austrin et al. [ABM12] is

Theorem 5.3.2 ([ABM12]). *Suppose that there exists $\eta \in \mathbb{R}$ such that*

$$\frac{2\eta}{\sqrt{2\pi}} f^{=1}(x) + \frac{2}{\pi} f^{=2}(x) > 0 \tag{5.3}$$

for every $x \in f^{-1}(1)$. Then there is a randomized polynomial time algorithm that approximates SCSP(S) better than the random assignment in expectation.

We compute $f^{=1}$ and $f^{=2}$.

$$\hat{f}(\{1\}) = \langle f, \chi_{\{1\}} \rangle = \frac{1}{2^k} \sum_{s \in S} \left(\binom{k-1}{s} - \binom{k-1}{s-1} \right)$$
$$\hat{f}(\{1,2\}) = \langle f, \chi_{\{1,2\}} \rangle = \frac{1}{2^k} \sum_{s \in S} \left(\binom{k-2}{s} - 2\binom{k-2}{s-1} + \binom{k-2}{s-2} \right)$$

By symmetry, $\hat{f}(T) =: \hat{f}_1$ is the same for all $|T| = 1$ and $\hat{f}(T) =: \hat{f}_2$ is the same for all $|T| = 2$. If we let $s = x_1 + \cdots + x_k$,

$$f^{=1}(x) = \hat{f}_1 \sum_{i \in [k]} (-1)^{x_i} = k\hat{f}_1 \mathbb{E}_{i \in [k]}[-2x_i + 1] = k\hat{f}_1(-2\frac{s}{k} + 1)$$
$$f^{=2}(x) = \hat{f}_2 \sum_{i \neq j} (-1)^{x_i + x_j} = \binom{k}{2} \hat{f}_2 \mathbb{E}_{i \neq j}[(-2x_i + 1)(-2x_j + 1)] = \binom{k}{2} \hat{f}_2(4\frac{s(s-1)}{k(k-1)} - 4\frac{s}{k} + 1).$$

When S is an interval. Let $S = \{l, l+1, \ldots, r-1, r\}$. If $r \leq \frac{k}{2}$, we have $(\frac{-2s}{k} + 1) \leq 0$ for all $s \in S$, so choosing η either large enough or small enough ensures (5.3). Similarly, if $l \geq \frac{k}{2}$, (5.3) holds. Therefore, we assume that $l < \frac{k}{2}$ and $r > \frac{k}{2}$, and compute $\hat{f}_2$.

$$\hat{f}_2 = \frac{1}{2^k} \sum_{s=l}^{r} \left(\binom{k-2}{s} - 2\binom{k-2}{s-1} + \binom{k-2}{s-2} \right)$$
$$= \frac{1}{2^k} \left(\binom{k-2}{l-2} - \binom{k-2}{l-1} + \binom{k-2}{r} - \binom{k-2}{r-1} \right)$$

Since $\binom{k-2}{l-1} > \binom{k-2}{l-2}$ for $0 < l < \frac{k}{2}$ and $\binom{k-2}{r-1} > \binom{k-2}{r}$ for $\frac{k}{2} < r < k$, $\hat{f}_2 < 0$ except when $l = 0$ and $r = k$ (i.e., $S = [k] \cup \{0\}$).

If $\mathsf{conv}(P_S)$ does not contain $(\frac{1}{2}, \frac{1}{4})$, there exist $\alpha, \beta \in \mathbb{R}$ such that for any $(a, b) \in \mathsf{conv}(P_S)$,

$$\alpha(a - \frac{1}{2}) + \beta(b - \frac{1}{4}) > 0.$$

If k is even, $s := \frac{k}{2} \in S$ and $P(s) = (\frac{1}{2}, \frac{s-1}{2(k-1)})$ where $\frac{s-1}{2(k-1)} < \frac{1}{4}$, which implies $\beta < 0$ since the above inequality should hold for all $s \in S$. When k is odd (let $k = 2s + 1$), s and $s + 1$ should be in S and

$$\frac{1}{2}\Big(P(s) + P(s+1)\Big) = (\frac{1}{2}, \frac{s^2}{k(k-1)}),$$

where $\frac{s^2}{k(k-1)} < \frac{1}{4}$. Therefore, we can conclude $\beta < 0$ in any case. For any $x \in f^{-1}(1)$ with $s = x_1 + \cdots + x_k$ and $P(s) = (a, b)$,

$$\begin{aligned}
& \frac{2\eta}{\sqrt{2\pi}} f^{=1}(x) + \frac{2}{\pi} f^{=2}(x) \\
= \quad & \frac{2\eta}{\sqrt{2\pi}} k\hat{f}_1(-2a+1) + \frac{2}{\pi}\binom{k}{2}\hat{f}_2(4b - 4a + 1) \\
= \quad & \frac{8}{\beta\pi}\binom{k}{2}\hat{f}_2\left(\frac{\frac{2\eta}{\sqrt{2\pi}}k\hat{f}_1}{\frac{8}{\beta\pi}\binom{k}{2}\hat{f}_2}(-2a+1) + \beta(b - a + \frac{1}{4})\right) \\
= \quad & \frac{8}{\beta\pi}\binom{k}{2}\hat{f}_2\left((-\frac{\alpha+\beta}{2})(-2a+1) + \beta(b - a + \frac{1}{4})\right) \quad \text{by adjusting } \eta \text{ so that } \frac{\frac{2\eta}{\sqrt{2\pi}}k\hat{f}_1}{\frac{8}{\beta\pi}\binom{k}{2}\hat{f}_2} = -\frac{\alpha+\beta}{2} \\
= \quad & \frac{8}{\beta\pi}\binom{k}{2}\hat{f}_2\left(\alpha(a - \frac{1}{2}) + \beta(b - \frac{1}{4})\right) \\
> 0. &
\end{aligned}$$

Therefore, (5.3) is satisfied if S is an interval and $\mathsf{conv}(S)$ does not contain $(\frac{1}{2}, \frac{1}{4})$.

When S is even. Given S, let $Q \in \{0,1\}^k$ be the predicate associated with S and $f : \{0,1\}^k \to \{0,1\}$ be the indicator function of Q. We want to show that when S is even,

$$\frac{2\eta}{\sqrt{2\pi}} f^{=1}(x) + \frac{2}{\pi} f^{=2}(x) > 0$$

is satisfied for any $x \in f^{-1}(1)$. When S is even,

$$\hat{f}_1 = \frac{1}{2^{k+1}} \sum_{s \in S} \left(\binom{k-1}{s} - \binom{k-1}{s-1} + \binom{k-1}{k-s} - \binom{k-1}{k-s-1} \right) = 0.$$

We compute the sign of the contribution of each s to $\hat{f}_2$.

$$\begin{aligned} & \binom{k-2}{s} - 2\binom{k-2}{s-1} + \binom{k-2}{s-2} \geq 0 \\ \Leftrightarrow \quad & (k-s)(k-s-1) - 2s(k-s) + s(s-1) \geq 0 \\ \Leftrightarrow \quad & 4s^2 - 4sk + k^2 - k \geq 0 \\ \Leftrightarrow \quad & s \leq \frac{k-\sqrt{k}}{2} \text{ or } s \geq \frac{k+\sqrt{k}}{2} \end{aligned}$$

We also consider the line passing $P(s)$ and $P(k-s)$. If we denote $t = k-s$, Its slope is

$$\frac{\frac{t(t-1)-s(s-1)}{k(k-1)}}{\frac{t-s}{k}} = \frac{t^2 - s^2 - (t-s)}{(k-1)(t-s)} = 1,$$

and the value of this line at $\frac{1}{2}$ is at least $\frac{1}{4}$ when

$$\begin{aligned} & \frac{s(s-1) + (k-s)(k-s-1)}{2k(k-1)} \geq \frac{1}{4} \\ \Leftrightarrow \quad & 2s(s-1) + 2(k-s)(k-s-1) \geq k(k-1) \\ \Leftrightarrow \quad & s \leq \frac{k-\sqrt{k}}{2} \text{ or } s \geq \frac{k+\sqrt{k}}{2}. \end{aligned}$$

Intuitively, if we consider the line of slope 1 that passes $(\frac{1}{2}, \frac{1}{4})$, $P(s)$ is below this line if $s \in (\frac{k-\sqrt{k}}{2}, \frac{k+\sqrt{k}}{2})$. Let $S_1 = S \cap \{0, 1, \ldots, \lceil \frac{k}{2} \rceil\}$. If S_1 contains a value $s_1 \leq \frac{k-\sqrt{k}}{2}$ and a value $s_2 \geq \frac{k-\sqrt{k}}{2}$ (including the case $s_1 = s_2 = \frac{k-\sqrt{k}}{2}$ is an integer in S_1), the line passing $P(s_1)$ and $P(k-s_1)$ passes a point $(\frac{1}{2}, t_1)$ for some $t_1 \geq \frac{1}{4}$ and the line passing $P(s_2)$ and $P(k-s_2)$ passes a point $(\frac{1}{2}, t_2)$ for some $t_2 \leq \frac{1}{4}$. Therefore, $\mathsf{conv}(P_S)$ contains a point $(\frac{1}{2}, \frac{1}{4})$ and S becomes balanced pairwise independent. We consider the remaining two cases.

1. $s < \frac{k-\sqrt{k}}{2}$ for all $s \in S_1$: $\hat{f}_2 > 0$ and for all $s \in S$, $-(\frac{s}{k} - \frac{1}{2}) + (\frac{s(s-1)}{k(k-1)} - \frac{1}{4}) > 0$. Therefore, for any $x \in f^{-1}$ with $s = x_1 + \cdots + x_k$,

$$\begin{aligned} & \frac{2\eta}{\sqrt{2\pi}} f^{=1}(x) + \frac{2}{\pi} f^{=2}(x) \\ = \quad & \frac{2}{\pi} f^{=2}(x) \\ = \quad & \frac{2}{\pi} \binom{k}{2} \hat{f}_2 (4\frac{s(s-1)}{k(k-1)} - 4\frac{s}{k} + 1) \\ > \quad & 0. \end{aligned}$$

2. $s > \frac{k-\sqrt{k}}{2}$ for all $s \in S_1$: $\hat{f}_2 < 0$ and for all $s \in S$, $-(\frac{s}{k} - \frac{1}{2}) + (\frac{s(s-1)}{k(k-1)} - \frac{1}{4}) < 0$. Similarly as above, for any $x \in f^{-1}$ with $s = x_1 + \cdots + x_k$, (5.3) is satisfied.

5.4 Austrin-Håstad Condition for Symmetric CSPs

This section explains how the condition of Austrin-Håstad [AH13] is simplified for SCSPs. They studied general CSPs where a predicate Q is a subset of $\{0,1\}^k$. Note that given $S \subseteq [k] \cup \{0\}$, SCSP($S$) is equivalent to CSP($Q$) where

$$Q = \{(x_1, \ldots, x_k) \in \{0,1\}^k : (x_1 + \cdots + x_k) \in S\} \tag{5.4}$$

Given Q, their general definition of *pairwise independence* and *positive correlation* is given below.

Definition 5.4.1. *Q is* balanced pairwise independent *if there is a distribution μ supported on Q such that* $\Pr_\mu[x_i = 1] = \frac{1}{2}$ *for every $i \in [k]$ and* $\Pr_\mu[x_i = x_j = 1] = \frac{1}{4}$ *for every $1 \le i < j \le k$.*

Definition 5.4.2. *Q is* positively correlated *if there is a distribution μ supported on Q and $p, \rho \in [0,1]$ with $\rho \ge p^2$ such that* $\Pr_\mu[x_i = 1] = p$ *for every $i \in [k]$ and* $\Pr_\mu[x_i = x_j = 1] = \rho$ *for every $1 \le i < j \le k$.*

We formally prove that their definitions have simpler descriptions in $\mathbb{R}^2$ for symmetric CSPs. Recall that given $s \in [k] \cup \{0\}$,

$$P(s) = (\frac{s}{k}, \frac{s(s-1)}{k(k-1)}) \in \mathbb{R}^2 \qquad \text{and} \qquad P_S := \{P(s) : s \in S\}.$$

Lemma 5.4.3. *Let $S \subseteq [k] \cup \{0\}$ and Q be obtained by* (6.4)*. Q is pairwise independent if and only if* $\mathsf{conv}(P_S)$ *contains $(\frac{1}{2}, \frac{1}{4})$, and Q is positively correlated if and only if* $\mathsf{conv}(P_S)$ *intersects the curve $y = x^2$.*

Proof. We first prove the second claim of the lemma. Let Q be positively correlated with parameters p, ρ ($\rho \geq p^2$) and the distribution μ such that $\Pr_\mu[x_i = 1] = p$ for all i, $\Pr_\mu[x_i = x_j = 1] = \rho$ and for all $i < j$. Let ν be the distribution of $x_1 + \cdots + x_k$ where $(x_1, \ldots, x_k)$ are sampled from μ.

$$(p, \rho) = (\mathbb{E}_i[x_i], \mathbb{E}_{i<j}[x_i x_j]) = (\mathbb{E}_{s\sim\nu}[\frac{s}{k}], \mathbb{E}_{s\sim\nu}[\frac{s(s-1)}{k(k-1)}]) = \mathbb{E}_{s\sim\nu}[P(s)],$$

proving that positive correlation of Q implies $(p, \rho) \in \mathsf{conv}(P_S)$. Since $P(s)$ is strictly below the curve $y = x^2$ for any $s \in [k-1]$ and (p, ρ) is on or above this curve, $\mathsf{conv}(P_S)$ must intersect $y = x^2$.

Suppose that $\mathsf{conv}(P_S)$ intersects the curve $y = x^2$. There exists a distribution ν on S such that $\mathbb{E}_{s\sim\nu}[P(s)] = (p, p^2)$. Let μ_s be the distribution on $\{0,1\}^k$ that uniformly samples a string with exactly s 1's. Let μ be the distribution where s is sampled from ν and $(x_1, \ldots, x_k)$ is sampled from μ_s. By definition, $\Pr_\mu[x_i = 1]$ and $\Pr_\mu[x_i = x_j = 1]$ do not depend on choice of indices,

$$\Pr_\mu[x_1 = 1] = \mathbb{E}_\mu[x_1] = \mathbb{E}_{s\sim\nu}\mathbb{E}_{x\sim\mu_s}[x_1] = \mathbb{E}_{s\sim\nu}[\frac{s}{k}] = p$$
$$\Pr_\mu[x_1 = x_2 = 1] = \mathbb{E}_\mu[x_1 x_2] = \mathbb{E}_{s\sim\nu}\mathbb{E}_{x\sim\mu_s}[x_1 x_2] = \mathbb{E}_{s\sim\nu}[\frac{s(s-1)}{k(k-1)}] = p^2,$$

implying that $(p, p^2) \in \mathsf{conv}(P_S)$.

The proof of the first claim is similar except that the curve $y = x^2$ is replaced by $(\frac{1}{2}, \frac{1}{4})$. $\square$

Lemma 5.4.4. $\mathsf{conv}(P_S)$ *intersects the curve $x = y^2$ if and only if*

$$\frac{(s_{max} + s_{min} - 1)^2}{k-1} \geq \frac{4 s_{max} s_{min}}{k} \ .$$

Proof. Let $l = s_{min}$ and $r = s_{max}$. The line passing $P(l)$ and $P(r)$ has a slope $\frac{\frac{r(r-1)-l(l-1)}{k(k-1)}}{\frac{r-l}{k}} = \frac{r+l-1}{k-1}$ and a y-intercept b such that

$$\frac{l(l-1)}{k(k-1)} = \frac{r+l-1}{k-1} \cdot \frac{l}{k} + b \Leftrightarrow b = \frac{l(l-1) - l(r+l-1)}{k(k-1)} = \frac{-lr}{k(k-1)}.$$

This line intersects $y = x^2$ if and only if

$$x^2 = \frac{r+l-1}{k-1}x - \frac{lr}{k(k-1)}$$

has a real root, which is equivalent to

$$(\frac{r+l-1}{k-1})^2 - \frac{4lr}{k(k-1)} \geq 0 \Leftrightarrow \frac{(r+l-1)^2}{k-1} \geq \frac{4lr}{k}. \qquad \square$$

Part II

Applied CSPs

Chapter 6

Unique Coverage

6.1 Introduction

Given a universe V of n elements and a collection E of m subsets of V, the UNIQUE COVERAGE problem asks to find $S \subseteq V$ to maximize the number of $e \in E$ that intersects S in exactly one element. When each $e \in E$ has size at most k, this problem is also known as 1-IN-k HITTING SET.

UNIQUE COVERAGE models numerous practical situations where each element represents a *service* and each subset represents a *customer* interested in the services it contains. We want to *activate* some services to satisfy customers, but customers want exactly one service from her list to be activated because more than one service may lead to confusion or high cost. These natural scenarios have been studied in many fields including wireless networks, radio broadcast, and envy-free pricing. We refer the reader to the work of Guruswami and Trevisan [GT05] and Demaine et al. [DFHS08] for a more detailed list of applications. Chalermsook et al. [CCKK12] showed an approximation-preserving reduction from UNIQUE COVERAGE to a special case of envy-free pricing called the TOLLBOOTH PRICING problem, so our result improves the hardness of TOLLBOOTH PRICING as well.

There is a simple $\Omega(\frac{1}{\log k})$-approximation algorithm for 1-IN-k HITTING SET. First, consider the case where each subset e has the same cardinality k (also known as 1-IN-k HITTING SET). Independently adding each $v \in V$ to S with probability $\frac{1}{k}$ will ensure that each set $e \in E$ will intersect S in exactly one element with probability $(1 - \frac{1}{k})^{k-1}$, which approaches $\frac{1}{e}$ as k grows. For the general case where each subset has cardinality at most k (assume k is a power of 2), randomly choosing a value $l \in \{2, 4, 8, \ldots, k\}$ first and independently adding each $v \in V$ to S with probability $\frac{1}{l}$ will give an $\Omega(\frac{1}{\log k})$-approximation algorithm. If there exists $S \subseteq V$ that intersects *every* subset in exactly one element, solving the standard LP relaxation and independently rounding with the resulting solution will guarantee a factor $1/e$-approximation even if the subsets have different sizes [GT05].

These approximation algorithms highlight interesting theoretical aspects of this problem. 1-IN-k HITTING SET can be naturally interpreted as a Constraint Satisfaction Problem (CSP) where each element $v \in V$ becomes a variable taking a value from $\{0, 1\}$ ($v \leftarrow 1$ corresponds to $v \in S$), and each subset becomes a constraint. Each constraint $e = (v_1, \ldots, v_l)$ is satisfied by an assignment $\sigma : V \to \{0, 1\}$ if and only if $\sigma(v_1) + \cdots + \sigma(v_l) = 1$. An $\Omega(1)$-approximation for 1-IN-Ek HITTING SET and an $\Omega(\frac{1}{\log k})$-approximation for 1-IN-k HITTING SET exhibit an example where mixing predicates of different arities decreases the best approximation ratio significantly. The second $\Omega(1)$-approximation when every subset can be intersected exactly once shows a rare example where *perfect completeness* of a CSP allows a much better approximation. When k is a growing function of n, as pointed out in [DFHS08], UNIQUE COVERAGE is one of few natural *maximization* problems for which the tight approximation threshold is (semi)logarithmic (i.e., $\Omega(\log^c n)$ for some $0 < c \leq 1$).

There are even more theoretically interesting developments from the hardness side. Demaine, Feige, Hajiaghayi, and Salavatipour [DFHS08] showed it is hard to approximate UNIQUE COVERAGE within a factor of $\Omega(\frac{1}{\log^\epsilon n})$ for some constant $\epsilon > 0$ depending on δ, assuming that $\mathbf{NP} \not\subseteq \mathbf{BPTIME}(2^{n^\delta})$ for some constant $\delta > 0$. Their second result proved that the inapproximability can be strengthened to $\Omega(\frac{1}{\log^{1/3-\epsilon} n})$ for any $\epsilon > 0$ assuming Feige's Random 3SAT Hypothesis [Fei02]. For 1-IN-k HITTING SET for constant k, Guruswami and Zhou [GZ12] recently proved that the $\Omega(\frac{1}{\log k})$-approximation is optimal, assuming Khot's Unique Games Conjecture [Kho02b]. Since many other problems whose strong inapproximabilities are known only under Feige's or Khot's conjecture, it was open whether we were able to *bypass* these conjectures to show almost optimal inapproximability only assuming $\mathbf{P} \neq \mathbf{NP}$ or $\mathbf{NP} \not\subseteq \mathbf{QP}$.

Our main contribution in this chapter is a positive answer to this question. For 1-IN-k HITTING SET for constant k, we prove the following theorem.

Theorem 6.1.1. *Assuming $\mathbf{P} \neq \mathbf{NP}$, for large enough constant k, there is no polynomial time algorithm that approximates 1-IN-k HITTING SET within a factor better than $O(\frac{1}{\log k})$.*

This result bypasses the Unique Games Conjecture to show that the simple $\Omega(\frac{1}{\log k})$-approximation algorithm is the best polynomial time algorithm up to a constant factor. For

UNIQUE COVERAGE, we prove that following theorem. Recall that

$$\mathbf{QP} = \cup_{c \in \mathbb{N}} \mathbf{DTIME}(2^{\log^c n}).$$

Theorem 6.1.2. *For any $\epsilon > 0$, assuming $\mathbf{NP} \not\subseteq \mathbf{DTIME}(2^{\log^{O(1/\epsilon)} n})$, there is no polynomial time algorithm that approximates* UNIQUE COVERAGE *within a factor better than $\frac{1}{\log^{1-\epsilon} n}$. In particular, for any fixed $\epsilon > 0$, unless $\mathbf{NP} \not\subseteq \mathbf{QP}$, there is no polynomial time algorithm that approximates* UNIQUE COVERAGE *within a factor better than $\frac{1}{\log^{1-\epsilon} n}$.*

Compared to the first result of Demaine et al., we replace their assumption $\mathbf{NP} \not\subseteq \mathbf{BPTIME}(2^{n^\delta})$ for *some* δ by a much weaker assumption $\mathbf{NP} \not\subseteq \mathbf{QP}$ and at the same time show an improved (and near-optimal) inapproximability factor, which is near-optimal and also improves their second result conditioned on the Random 3SAT Hypothesis.

Besides these improvements, our proof is also significantly simpler than previous works. The result of Guruswami and Zhou for constant k is obtained by constructing a gap instance for a semidefinite programming (SDP) relaxation for the problem, and using the sophisticated result of Raghavendra [Rag08] that converts an SDP gap to a Unique Games hardness. Demaine et al. first showed a reduction from an intermediate problem called *Balanced Bipartite Independent Set* (BBIS) to UNIQUE COVERAGE, and used the Random 3SAT Hypothesis or Khot's Quasirandom PCP [Kho06] to prove hardness of BBIS. Our two theorems are corollaries of one simple reduction from the basic Label Cover, whose hardness relies only on the PCP theorem and the Parallel Repetition Theorem.

6.1.1 Techniques

While UNIQUE COVERAGE can be interpreted as a CSP, it also seems similar to the Max k-Coverage problem, where, given a set system (V, E), we want to find a subset $S \subseteq V$ with $|S| = k$ that intersects as many $e \in E$ as possible.[1] Max k-Coverage is tightly related to the more famous Set Cover problem and admits an $\frac{e-1}{e}$-approximation algorithm which is proved to be tight [LY94, Fei98]. It can be also interpreted as a variant of CSPs where each element becomes a variable taking a value from $\{0, 1\}$, and each subset becomes a constraint that is satisfied if at least one of its variables is assigned 1, and we additionally require that at most k variables have to be assigned 1.

A weaker but simple inapproximability of Max k-Coverage can be proved via the Label Cover problem. An instance of Label Cover consists of a biregular bipartite graph $G =$

[1]Max k-Coverage is usually stated in terms of the dual set system, where we want to find a subcollection $E' \subseteq E$ of subsets with $|E'| = k$ that maximizes the number of elements covered.

$(U_G \cup V_G, E_G)$ where each edge $e = (u, v)$ is associated with a projection $\pi_e : [R] \mapsto [L]$ for some positive integers R and L, and we look for a labeling $l : U_G \cup V_G \mapsto [R]$ that satisfies as many $e \in E_G$ as possible ($e = (u, v)$ is satisfied when $\pi_e(l(v)) = l(u)$). Given an instance of Label Cover, the reduction to Max k-Coverage makes every (vertex, label)-pair of Label Cover as an element of the set system, and for each projection $e = (u, v) \in E_G$ and $b \in \{0, 1\}^L$, there is a subset corresponding to (e, b) containing $\{(u, j) : b_j = 0\} \cup \{(v, j) : b_{\pi_e(j)} = 1\}$. It is a simple but useful exercise to check that if a labeling l satisfies every projection, its *canonical set* $\{(v, l(v)) : v \in U_G \cup V_G\}$ will intersect every subset exactly once. However, it is also easy to see that for any labeling l, its canonical set will intersect at least half of subsets exactly once.

In order to prove a stronger inapproximability result, we have a subset for each tuple $(e_1, \ldots, e_q, b)$ for various values of q where $e_1, \ldots, e_q$ share an endpoint in U_G. If l satisfies *all* $e_1, \ldots, e_q$, its canonical set will intersect $(e_1, \ldots, e_q, b)$ in exactly one element for many (but not all) b, but if l does not even *approximately* satisfy $e_1, \ldots, e_q$, there no way to intersect many subsets in exactly one element. Even though our technique is different from traditional hardness results for Max-CSPs (e.g., no long code consisting of variables), the idea of *probabilistic checking* (i.e., subsets having weights summing up to 1, and the instance is interpreted as a probabilistic procedure where we sample a subset e according to weights and check $|S \cap e| = 1$) conceptually simplifies the proof and technically makes the reduction efficient by appealing to various derandomization methods based on bounded independence.

6.1.2 Preliminaries

An instance of UNIQUE COVERAGE is simply a set system. We view the set system as a hypergraph $H = (V_H, E_H)$, where V_H is the universe of elements and E_H is a collection of hyperedges. Unless stated otherwise, every $\log$ in this chapter indicates a logarithm base 2. We use $a \sim \mathcal{D}$ to indicate that a random variable a is sampled from a distribution $\mathcal{D}$. When a random variable a is sampled uniformly from a set A, we write $a \in A$. For a positive integer m, we denote $[m] := \{1, 2, \ldots, m\}$.

6.2 Reduction from LABEL COVER

Our main reduction is from LABEL COVER introduced in Chapter 3.1. Given an instance of Label Cover $G = (U_G \cup V_G, E_G)$ with projections $\{\pi_e\}_{e \in E_G}$ with parameters R, L, D, d, we produce an instance $H = (V_H, E_H)$ of UNIQUE COVERAGE. The set of vertices V_H is defined to be $V_G \times [R]$. In the following, we describe a probabilistic procedure to sample a hyperedge e. E_H is defined to be the set of hyperedges with nonzero probability, with these probabilities as weights. We abuse notation and let E_H also denote the distribution. There are three distributions used to describe the entire procedure.

1. Let Q be a positive integer to be determined later. We will take Q to be a power of 2 and $Q < D$. Let $\mathcal{D}$ be a uniform distribution on $\{2, 4, 8, \ldots, Q\}$.

2. For each $u \in U_G$, let $\mathcal{D}_{u,Q}$ be a *uniform pairwise independent* distribution on $(v_1, \ldots, v_Q) \in N(u)^Q$ such that

$$\Pr_{(v_1,\ldots,v_Q)\sim\mathcal{D}_{u,Q}}[v_i = v, v_j = v'] = \frac{1}{D^2} \quad \text{for all } i \neq j \in [Q] \text{ and } v, v' \in N(u),$$

   and its support has size D^2. Note that it implies that $\Pr[v_i = v] = \frac{1}{D}$ for all $i \in [Q]$ and $v \in N(u)$.

   Claim 6.2.1. *Such a distribution $\mathcal{D}_{u,Q}$ exists.*

   Proof. $\mathcal{D}_{u,Q}$ can be described by the following standard procedure. Fix a bijection f from $N(u)$ to the finite field $\mathbb{F}_D$ (recall that D is a power of 5), another injective mapping g from $[Q]$ to a subset of $\mathbb{F}_D$, sample $a, b \in \mathbb{F}_D$ independently, and output $v_i \leftarrow f^{-1}(a \cdot g(i) + b)$. It is a standard fact that this distribution is uniform pairwise independent. □

3. Let $\mathcal{D}_L$ be a uniform 4-wise independent distribution on $(c_1, \ldots, c_L) \in \{0,1\}^L$ such that for all $(j_1, j_2, j_3, j_4) \in \binom{[L]}{4}$ and $(b_1, b_2, b_3, b_4) \in \{0,1\}^4$,

$$\Pr_{(c_1,\ldots,c_L)\sim\mathcal{D}_L}[c_{j_i} = b_i \text{ for } 1 \leq i \leq 4] = \frac{1}{2^4},$$

   and its support has size $2L^2$.

   Claim 6.2.2. *Such a distribution $\mathcal{D}_L$ exists.*

   Proof. $\mathcal{D}_L$ can be described by the following procedure. Fix a bijection σ from $[L]$ to the finite field $\mathbb{F}_L$ (recall that L is a power of 2). Sample $a, b \in \mathbb{F}_L$ and $d \in \mathbb{F}_2$ independently, and output $c_i \leftarrow \mathrm{Tr}(a \cdot \sigma(i)^3 + b \cdot \sigma(i)) + d$ where Tr is the Trace map from $\mathbb{F}_L$ to $\mathbb{F}_2$. This distribution is uniform over the codewords of a binary linear code whose dual is a linear code (specifically, a BCH code) of minimum distance at least 5, and is therefore uniform 4-wise independent. As this explicit form of the 4-wise independent distribution may not be that widely known, let us give some details. For a linear code $C \subseteq \mathbb{F}_L^n$, define $\mathrm{Tr}(C) = \{(\mathrm{Tr}(c_1), \mathrm{Tr}(c_2), \cdots, \mathrm{Tr}(c_n)) \mid (c_1, c_2, \ldots, c_n) \in C\}$; note that $\mathrm{Tr}(C) \subseteq \mathbb{F}_2^n$ is a binary linear code. Then Delsarte's identity states that $\mathrm{Tr}(C)^\perp = C^\perp \cap \mathbb{F}_2^n$ (see, for instance, [Sti08, Chap. 9]). Apply this to C being the Reed-Solomon code of block length L and dimension 4 over $\mathbb{F}_L$, i.e., $C = \{(p(\alpha))_{\alpha \in \mathbb{F}_L} \mid p \in \mathbb{F}_L[X],\ \deg(p) \leq 3\}$. Then it is a well-known fact that $C^\perp$ is the Reed-Solomon code of dimension $L-4$, i.e., $C^\perp = \{(p(\alpha))_{\alpha \in \mathbb{F}_L} \mid p \in \mathbb{F}_L[X],\ \deg(p) \leq L-5\}$. So $C^\perp$, and hence also $C^\perp \cap \mathbb{F}_2^L$ (which equals $\mathrm{Tr}(C)^\perp$), has minimum distance at least 5. Finally, we observe that $\mathrm{Tr}(C) = \{\big(\mathrm{Tr}(a\alpha^3 + b\alpha) + d\big)_{\alpha \in \mathbb{F}_L} \mid a, b \in \mathbb{F}_L, d \in \mathbb{F}_2\}$ since $\mathrm{Tr}(\alpha^2) = \mathrm{Tr}(\alpha)$. $\square$

Given these distributions, a random hyperedge e is sampled by the following procedure:

- Sample $u \in U_G$.
- Sample $q \sim \mathcal{D}$.
- Sample $(v_1, \ldots, v_Q) \sim \mathcal{D}_{u,Q}$. Note that only $v_1, \ldots, v_q$ are used in the reduction. This slightly redundant sampling reduces the number of distributions involved and simplifies our analysis.
- Sample $(c_1, \ldots, c_L) \sim \mathcal{D}_L$.
- For $j \in [L]$, consider a *block* of vertices $\cup_{i=1}^q (\{v_i\} \times \pi^{-1}_{(u,v_i)}(j))$. Every block has cardinality at most qd. It has exactly qd vertices when $v_1, \ldots, v_Q$ are pairwise distinct.
  - If $c_j = 0$, add d vertices in $\{v_1\} \times \pi^{-1}_{(u,v_1)}(j)$ to e.
  - If $c_j = 1$, add the entire block to e.

Note that the maximum cardinality of any hyperedge is RQ, $|V_H| = |V_G| \cdot R$, and the total number of hyperedges with nonzero probability is bounded by $s := |U_G| \cdot \log Q \cdot D^2 \cdot 2L^2$. Also, our definition of weights ensures that the weight of each hyperedge is an integer multiple of $\frac{1}{s}$, so one can view this weighted instance as an unweighted instance with exactly s hyperedges. Since we use s as the size of the instance throughout the paper, the same hardness results hold for unweighted UNIQUE COVERAGE.

6.2.1 Completeness

Lemma 6.2.3. *If the instance $(G, \{\pi_e\}_e)$ of Label Cover admits a labeling l that satisfies every projection, there exists $S \subseteq V_H$ such that $\Pr_{e \sim E_H}[|e \cap S| = 1] \geq \frac{1}{2}$.*

Proof. Given a labeling $l : U_G \cup V_G \mapsto [R]$, let $S := \cup_{v \in V_G}\{(v, l(v))\}$. From the above probabilistic procedure to sample a hyperedge e, every choice of u, q, $v_1, \ldots, v_q$ satisfies $\pi_{(u,v_1)}(l(v_1)) = \cdots = \pi_{(u,v_q)}(l(v_q))$. In particular, S and $\cup_{i=1}^{q}(\{v_i\} \times [R])$ intersect in exactly one block corresponding to $l(u) \in [L]$. Therefore, if $c_{l(u)} = 0$, which happens with probability $\frac{1}{2}$, $(v_1, l(v_1))$ is the only element in $S \cap e$. □

6.2.2 Soundness

For soundness, we prove that if the Label Cover instance $(G, \{\pi_e\})$ does not admit a good labeling, the UNIQUE COVERAGE instance H does not have a good solution either.

Lemma 6.2.4. *If every labeling $l : U_G \cup V_G \mapsto [R]$ satisfies at most ϵ fraction of projections, for every $S \subseteq V_H$, $\Pr_{e \sim E_H}[|e \cap S| = 1] \leq 2Q\sqrt{\epsilon} + \frac{Q^2}{D} + O\big(\frac{1}{\log Q}\big)$.*

Proof. Fix $S \subseteq V_H$. We construct a partial labeling $l : V_G \mapsto [R]$ as follows. For each $v \in V_G$ and $j \in [R]$, we say that v *picks* j if $(v, j) \in S$. If v picked at least one label, we choose an arbitrarily picked label j and set $l(v) = j$. Otherwise, we set $l(v) = \emptyset$, which means that every projection that includes v will not be satisfied. Note that we have not defined labels for U_G yet.

For $q \in \{2, 4, 8, \ldots, Q\}$, $u \in U_G$, and $v_1, \ldots, v_q \in N(u)$, we say that $(u, v_1, \ldots, v_q)$ is *weakly satisfied* by a partial labeling to V_G if there exist $1 \leq i < j \leq q$ such that $l(v_i) \neq \emptyset, l(v_j) \neq \emptyset$, and $\pi_{(u,v_i)}(l(v_i)) = \pi_{(u,v_j)}(l(v_j))$. Note that if the Label Cover instance admitted a labeling l^* that satisfied every projection, every tuple $(u, v_1, \ldots, v_q)$ with $v_1, \ldots, v_q \in N(u)$ would satisfy $\pi_{(u,v_1)}(l^*(v_1)) = \cdots = \pi_{(u,v_q)}(l^*(v_q))$. The following claim shows that since the Label Cover instance does not admit a good labeling, if we sample $u, v_1, \ldots, v_Q$ as in the reduction, $(u, v_1, \ldots, v_Q)$ is unlikely to be even weakly satisfied.

Claim 6.2.5. *Suppose we sample $u \in U_G$ and $(v_1, \ldots, v_Q) \sim \mathcal{D}_{u,Q}$. The probability that $(u, v_1, \ldots, v_Q)$ is weakly satisfied is at most $Q^2\epsilon$.*

Proof. Fix $1 \leq i < j \leq Q$. We will show that the probability that $\pi_{(u,v_i)}(l(v_i)) = \pi_{(u,v_j)}(l(v_j))$ is at most ϵ, which implies the claim by taking the union bound over $\binom{Q}{2}$ pairs.

Given a labeling $l' : U_G \cup V_G \mapsto [R]$, since the Label Cover instance is biregular, the fraction of satisfied projections by l' is equal to $\Pr_{u,v_i}[\pi_{(u,v_i)}(l'(v_i)) = l'(u)]$ ($u, v_1, \ldots, v_Q$ are sampled as above). Now let l' be the randomized extension to l, where $l'(v) = l(v)$ for $v \in V_G$ and each $u \in U_G$ independently picks one more random neighbor $v \in N(u)$ and sets $l'(u) \leftarrow \pi_{(u,v)}(l(v))$. The expected the fraction of satisfied projections by this random labeling l' is equal to

$$\begin{aligned}\Pr_{u,v_i,v}[\pi_{(u,v_i)}(l'(v_i)) = l'(u)] &= \Pr_{u,v_i,v}[\pi_{(u,v_i)}(l(v_i)) = \pi_{(u,v)}(l(v))] \\ &= \Pr_{u,v_i,v_j}[\pi_{(u,v_i)}(l(v_i)) = \pi_{(u,v_j)}(l(v_j))],\end{aligned}$$

where the last equality uses the fact that $\mathcal{D}_{u,Q}$ is uniform pairwise independent, so that given u, the pair (v_i, v) has the same joint distribution as (v_i, v_j). This quantity is at most ϵ, since every labeling satisfies at most ϵ fraction of projections. $\square$

By an averaging argument, the fraction of $u \in U_G$ such that

$$\Pr_{(v_1,\ldots,v_Q)\sim\mathcal{D}_{u,Q}}[(u, v_1, \ldots, v_Q) \text{ is weakly satisfied}] > Q\sqrt{\epsilon}$$

is at most $Q\sqrt{\epsilon}$. Call these u *bad*, and fix a *good* u. Since the probability that $v_i = v_j$ is exactly $\frac{1}{D}$ for fixed $i \neq j$,

$$\Pr_{(v_1,\ldots,v_Q)\sim\mathcal{D}_{u,Q}}[\exists i \neq j \text{ s.t. } v_i = v_j] \leq \binom{Q}{2}\frac{1}{D} \leq \frac{Q^2}{D}. \tag{6.1}$$

Fix $q \in \{2, 4, 8, \ldots, Q\}$. For fixed u, by the definition of weak satisfaction, the probability of weak satisfaction decreases as the number of considered neighbors q decreases, i.e.,

$$\begin{aligned}&\Pr_{(v_1,\ldots,v_Q)\sim\mathcal{D}_{u,Q}}[(u, v_1, \ldots, v_q) \text{ is weakly satisfied}] \\ \leq &\Pr_{(v_1,\ldots,v_Q)\sim\mathcal{D}_{u,Q}}[(u, v_1, \ldots, v_Q) \text{ is weakly satisfied}] \\ \leq &\, Q\sqrt{\epsilon}.\end{aligned} \tag{6.2}$$

Fix $v_1, \ldots, v_q \in N(u)$ such that $(u, v_1, \ldots, v_q)$ is not weakly satisfied and $v_1, \ldots, v_q$ are pairwise distinct. Let $p := |\{i \in [q] : l(v_i) \neq \emptyset\}|$. The fact that $(u, v_1, \ldots, v_q)$ is not weakly satisfied implies that for at least p values of $j \in [L]$, the corresponding block $\cup_{i=1}^{q}(\{v_i\} \times \pi^{-1}_{(u,v_i)}(j))$ intersects S. Consider the probabilistic procedure to sample a hyperedge e as in the reduction.

Claim 6.2.6. *Given u, q, $v_1, \ldots, v_q$, and p satisfying the conditions above, the probability that $|e \cap S| = 1$ is at most*

$$\begin{cases} 0 & p = 0 \\ 1 & 1 \le p \le 4 \\ O\left(\frac{1}{p^2}\right) & p > 4. \end{cases}$$

Proof. The above upper bounds are clear when $p = 0$ or $1 \le p \le 4$. For $p \ge 5$, note that if at least 2 of p different blocks see $c_j = 1$ and decide to add the whole block to e, $|e \cap S| \ge 2$. Therefore, if we let $\{i \in [q] : l(v_i) \ne \emptyset\} = \{i_1, \ldots, i_p\}$, $\Pr[|e \cap S| = 1]$ is at most the probability that $c_{i_1} + \cdots + c_{i_p} \le 1$ when $(c_1, \ldots, c_L) \sim \mathcal{D}_L$. We use the following concentration inequality for the sum of k-wise independent random variables by Bellare and Rompel [BR94].

Theorem 6.2.1 ([BR94])**.** *Let k be an even integer, and let X be the sum of n k-wise independent random variables taking values in $[0, 1]$. Let $\mu = \mathbb{E}[X]$ and $a > 0$. Then we have*

$$\Pr[|X - \mu| > a] < 1.1\left(\frac{nk}{a^2}\right)^{k/2}.$$

Applying the above theorem with $n \leftarrow p, \mu = p/2, k \leftarrow 4, a \leftarrow p/4$ gives

$$\Pr[c_{i_1} + \cdots + c_{i_p} \le 1] < 1.1\left(\frac{64}{p}\right)^2 = O\left(\frac{1}{p^2}\right).$$

□

Let $\alpha := \alpha(u)$ be the fraction of $v \in N(u)$ such that $l(v) \ne \emptyset$. For $1 \le i \le q$, let $b_i \in \{0, 1\}$ be the random variable such that $b_i = 1$ if and only if $l(v_i) \ne \emptyset$. By pairwise independence of $\{v_1, \ldots, v_q\}$, $\{b_1, \ldots, b_q\}$ are also pairwise independent, and $\Pr[b_i = 1] = \alpha$ for each i. Let $\mathcal{B}_{u,q,\alpha}$ be the distribution on $b_1 + \cdots + b_q$. By (6.1), (6.2), and Claim refclaim:p, for fixed good u and q, $\Pr_e[|e \cap S| = 1 \mid u, q]$ is at most

$$Q\sqrt{\epsilon} + \frac{Q^2}{D} + \sum_{p=1}^{q} \Pr_{X \sim \mathcal{B}_{u,q,\alpha}}[X = p] \cdot O\left(\frac{1}{p^2}\right). \tag{6.3}$$

We now consider the expected value of (6.3) over $q \sim \mathcal{D}$, with u still fixed.

Claim 6.2.7.

$$\mathbb{E}_{q \sim \mathcal{D}}\left[\sum_{p=1}^{q} \Pr_{X \sim \mathcal{B}_{u,q,\alpha}}[X = p] \cdot O\left(\frac{1}{p^2}\right)\right] = O(\frac{1}{\log Q}).$$

Proof. Since $\Pr_{X\sim\mathcal{B}_{u,q,\alpha}}[X=p]=0$ when $p>q$,

$$\begin{aligned}\mathbb{E}_{q\sim\mathcal{D}}\left[\sum_{p=1}^{q}\Pr_{X\sim\mathcal{B}_{u,q,\alpha}}[X=p]\cdot O\left(\frac{1}{p^2}\right)\right] &= \mathbb{E}_{q\sim\mathcal{D}}\left[\sum_{p=1}^{Q}\Pr_{X\sim\mathcal{B}_{u,q,\alpha}}[X=p]\cdot O\left(\frac{1}{p^2}\right)\right] \\ &= \sum_{p=1}^{Q} O\left(\frac{1}{p^2}\right)\mathbb{E}_{q\sim\mathcal{D}}\left[\Pr_{X\sim\mathcal{B}_{u,q,\alpha}}[X=p]\right].\end{aligned}$$

Since $\sum_{p=1}^{Q} O\left(\frac{1}{p^2}\right)=O(1)$, it suffices to prove that for any $1\le p\le Q$,

$$\mathbb{E}_{q\sim\mathcal{D}}\left[\Pr_{X\sim\mathcal{B}_{u,q,\alpha}}[X=p]\right]=O\big(\frac{1}{\log Q}\big). \tag{6.4}$$

We analyze it by considering how $\Pr[X=p]$ changes as q gets smaller or larger. For the lower tail where $q\alpha\le\frac{p}{2}$, let y be the biggest integer such that $2^y\in[2,Q]$ and $\alpha 2^y\le\frac{p}{2}$. For every $x=y,y-1,\dots,1$, by Markov's inequality,

$$\Pr_{X\sim\mathcal{B}_{u,2^x,\alpha}}[X=p]\le\Pr_{X\sim\mathcal{B}_{u,2^x,\alpha}}[X\ge p]\le\frac{\alpha 2^x}{p}.$$

By our choice of y, when $x=y$, $\frac{\alpha 2^x}{p}\le\frac{1}{2}$, and it decreases by a factor of 2 as we decrease x by 1. Therefore,

$$\begin{aligned}&\Pr_q\left[q\alpha\le\frac{p}{2}\right]\cdot\mathbb{E}_q\left[\Pr_{X\sim\mathcal{B}_{u,q,\alpha}}[X=p]\,\middle|\,q\alpha\le\frac{p}{2}\right] \\ =\ &\sum_{x=y}^{1}\Pr_q[q=2^x]\cdot\mathbb{E}_q\left[\Pr_{X\sim\mathcal{B}_{u,q,\alpha}}[X=p]\,\middle|\,q=x\right] \\ \le\ &\frac{1}{\log Q}\sum_{x=y}^{1}\big(\frac{\alpha 2^x}{p}\big)\le O\big(\frac{1}{\log Q}\big).\end{aligned} \tag{6.5}$$

For the upper tail where $q\alpha\ge 2p$, let y be the smallest integer such that $2^y\in[2,Q]$ and $\alpha 2^y\ge 2p$. For every $x=y,y+1,\dots,\log Q$, let X be a random variable sampled from $X\sim\mathcal{B}_{u,2^x,\alpha}$. By pairwise independence of $(b_1,\dots,b_q)$, $\mathsf{Var}[X]\le\mathbb{E}[X]=\alpha 2^x$. By

Chebyshev's inequality,

$$\begin{aligned}
& \Pr_{X\sim\mathcal{B}_{u,2^x,\alpha}}[X=p] \\
\leq\ & \Pr_{X\sim\mathcal{B}_{u,2^x,\alpha}}[X\leq p] \\
\leq\ & \Pr_{X\sim\mathcal{B}_{u,2^x,\alpha}}[|X-\mathbb{E}[X]|\geq|\mathbb{E}[X]-p|] \\
\leq\ & \frac{\alpha 2^x}{(\alpha 2^x-p)^2}.
\end{aligned}$$

By our choice of y, for any $x \geq y$, $\frac{\alpha 2^x}{(\alpha 2^x-p)^2} \leq \frac{\alpha 2^x}{(\frac{\alpha 2^x}{2})^2} = \frac{4}{\alpha 2^x}$, and it is decreased by a factor of 2 as we increase x by 1. Therefore,

$$\mathbb{E}_q\left[\Pr_{X\sim\mathcal{B}_{u,q,\alpha}}[X=p]\,\middle|\, q\alpha\geq 2p\right]\Pr_q\left[q\alpha\geq 2p\right]\leq\frac{1}{\log Q}\sum_{x=y}^{\log Q}\frac{4}{\alpha 2^x}\leq O\big(\frac{1}{\log Q}\big). \quad (6.6)$$

Finally,

$$\Pr[\frac{p}{2}\leq q\leq 2p]\leq O\big(\frac{1}{\log Q}\big). \quad (6.7)$$

Equations (6.5), (6.6), (6.7) imply (6.4), which completes the proof of the claim. □

Therefore, for a good u, $\Pr_e[|e\cap S|=1\,|\,u]$ is at most

$$Q\sqrt{\epsilon}+\frac{Q^2}{D}+O\big(\frac{1}{\log Q}\big),$$

and the overall probability $\Pr_e[|e\cap S|=1]$ is at most

$$2Q\sqrt{\epsilon}+\frac{Q^2}{D}+O\big(\frac{1}{\log Q}\big), \quad (6.8)$$

as desired in the Lemma. □

6.3 Main Results

We compose our reduction from Label Cover to UNIQUE COVERAGE with the standard reduction from 3SAT to Label Cover. We restate Theorem 3.1.1 that shows the properties of the reduction from 3SAT to Label Cover.

Theorem 6.3.1 (Restatement of Theorem 3.1.1)**.** *There exists an absolute constant $\tau < 1$ such that the following is true. For any positive integer $r > 0$, there is a reduction that given an instance ϕ of 3SAT with n variables, outputs an instance of Label Cover $(G, \{\pi_e\}_e)$ with $|U_G|, |V_G| = n^{O(r)}$, $R = 10^r, L = 2^r, d = D = 5^r$ in time $n^{O(r)}$, and satisfies the following.*

- *Completeness: If ϕ is satisfiable, there exists a labeling that satisfies every projection.*
- *Soundness: If ϕ is not satisfiable, every labeling satisfies at most τ^r fraction of projections.*

Let $\gamma > 1$ be an absolute constant such that $\gamma\tau^{1/2} < \frac{1}{\gamma}$ and $\frac{\gamma^2}{5} < \frac{1}{\gamma}$, and for each r, let $Q = Q(r)$ be the largest power of 2 at most γ^r. We run our reduction given the Label Cover instance $(G, \{\pi_e\}_e)$ to produce an instance of UNIQUE COVERAGE $H = (V_H, E_H)$. Recall that $|V_H| = |V_G| \cdot R = n^{O(r)}$, and $|E_H| = |U_G| \cdot \log Q \cdot D^2 \cdot 2L^2 = n^{O(r)}$, and the cardinality of each hyperedge is at most RQ.

If ϕ is satisfiable, $(G, \{\pi_e\}_e)$ admits a labeling that satisfies every constraint, so by Lemma 6.2.3, there exists $S \subseteq V_H$ such that the total weight of the hyperedges intersecting S in exactly one element is at least $\frac{1}{2}$. If ϕ is not satisfiable, every labeling of $(G, \{\pi_e\}_e)$ satisfies at most $\epsilon = \tau^r$ fraction of projections, and by Lemma 6.2.4, for any $S \subseteq V_H$, the total weight of hyperedges intersecting S in exactly one element is at most

$$2Q\sqrt{\epsilon} + \frac{Q^2}{D} + O\big(\frac{1}{\log Q}\big).$$

As r increases, (6.8) becomes

$$\begin{aligned} & 2Q\sqrt{\epsilon} + \frac{Q^2}{D} + O\big(\frac{1}{\log Q}\big) \\ \leq \quad & 2(\gamma\tau^{1/2})^r + \big(\frac{\gamma^2}{5}\big)^r + O(\frac{1}{\log Q}) \\ \leq \quad & \frac{3}{Q} + O(\frac{1}{\log Q}) = O(\frac{1}{\log Q}) = O(\frac{1}{\log(RQ)}), \end{aligned}$$

using the fact that $\log(RQ) = \Theta(\log Q)$.

6.3.1 1-IN-k HITTING SET for Constant k

We set parameters to show inapproximability of 1-in-k for constant k, proving Theorem 6.1.1. Given a constant k, take the largest r such that $\frac{k}{20} \leq RQ \leq k$ (R is always a power of 10 and Q is a power of 2). Since r is a constant, the combined reduction from 3SAT to 1-IN-k HITTING SET runs in time polynomial in n. Therefore, if we approximate 1-IN-k HITTING SET within a factor better than $O(\frac{1}{\log(RQ)}) = O(\frac{1}{\log k})$ in polynomial time in $|V_H|$, we can decide whether a given formula ϕ is satisfiable or not in time polynomial in n.

Note that the choice of r given k depends on the absolute constant τ in Theorem 3.1.1, and the gap of $O(\frac{1}{\log k})$ depends on another absolute constant hidden in (6.4). Therefore, our result requires k to be larger than some function of these two constants to show APX-hardness. But even when $k = 2$, 1-IN-k HITTING SET already captures the Max Cut problem and is APX-hard.

6.3.2 UNIQUE COVERAGE

We now set parameters to show inapproximability of UNIQUE COVERAGE, proving Theorem 6.1.2. Given $\epsilon > 0$, let $r = \log^{1/\epsilon} n$. For some absolute constant $\alpha > 1$, the combined reduction from 3SAT in UNIQUE COVERAGE runs in time $n^{\alpha \log^{1/\epsilon} n} = 2^{\alpha \log^{1/\epsilon+1} n}$, which is quasi-polynomial in n. Note that $|V_H| \leq 2^{\alpha \log^{1/\epsilon+1} n}$ and $RQ \geq 2^{\beta r} = 2^{\beta \log^{1/\epsilon} n}$ for another absolute constant $\beta > 0$. Therefore,

$$\log RQ \geq \beta \log^{1/\epsilon} n = \beta \cdot \alpha^{-\frac{1}{1+\epsilon}} \cdot (\alpha \log^{1/\epsilon+1} n)^{\frac{1}{1+\epsilon}} = \Omega(\log^{\frac{1}{1+\epsilon}} |V_H|),$$

so if we approximate UNIQUE COVERAGE within a factor better than $O(\frac{1}{\log^{1-\epsilon} |V_H|})$ in time polynomial in $|V_H|$, we can decide whether a given formula ϕ is satisfiable or not in quasi-polynomial time in n.

Chapter 7

Graph Pricing

7.1 Introduction

Consider the following natural problem for a seller with a profit-maximization objective. The seller has n types of items $1, \ldots, n$, each with unlimited copies, and there are m customers $1, \ldots, m$. Each customer j has her own budget b_j and a subset of items $e_j \subseteq \{1, \ldots, n\}$ that she is interested in. Customers are ***single-minded*** in a sense that each customer j buys all items in e_j if the sum of the prices does not exceed her budget (i.e. $b_j \geq \sum_{i \in e_j} p(i)$, where $p(i)$ indicates the price of item i), in which the seller gets $\sum_{i \in e_j} p(i)$ from the customer. Otherwise, the customer does not buy anything and the seller gets no profit from this customer. The goal of the seller is to set a nonnegative price to each item to maximize her profit from m customers.

This problem was proposed by Guruswami et al. [GHK$^+$05], and has received much attention. Let k be the maximum cardinality of any e_i. Approximability of this problem achieved by polynomial time algorithms for large k and n is relatively well-understood now. There is a polynomial time algorithm that guarantees $O(\min(k, (n \log n)^{1/2}))$ fraction of the optimal solution, while we cannot hope for an approximation ratio better than $\Omega(\min(k^{1-\epsilon}, n^{1/2-\epsilon}))$ for any $\epsilon > 0$ under the Exponential Time Hypothesis [CLN13].

The special case $k = 2$ has also been studied in many works separately. The instance can be nicely represented by a graph, with vertices as items and edges as customers, so this problem is called the GRAPH (VERTEX) PRICING problem. The fact that this case can be represented as a graph not only gives a theoretical simplification, but also makes the problem flexible to model other settings. For example, Lee et al. [LBA$^+$07, LBA$^+$08] independently suggested the same problem from the networking community, motivated by the study of pricing traffic between different levels of internet service providers under the presence of peering.

The best known approximation algorithm for a general instance of GRAPH PRICING, which guarantees $\frac{1}{4}$ of the optimal solution, is given by Balcan and Blum [BB07] and Lee et al. [LBA$^+$07] The algorithm is simple enough to state here. First, assign 0 to each vertex with probability half independently. For each remaining vertex v, assign the price which maximizes the profit between v and its neighbors already assigned 0. This simple algorithm has been neither improved nor proved to be optimal. GRAPH PRICING is APX-hard [GHK$^+$05], but the only strong hardness of approximation result rules out an approximation algorithm with a guarantee better than $\frac{1}{2}$ [KKMS09] under the Unique Games Conjecture (UGC) (via reduction from MAXIMUM ACYCLIC SUBGRAPH).

The $\frac{1}{4}$-approximation algorithm is surprisingly simple and does not even rely on the power of a linear programming (LP) or semidefinite programming (SDP) relaxation. The efforts to exploit the power of LP relaxations to find a better approximation algorithm have produced positive results for special classes of graphs. Krauthgamer et al. [KMR11] studied the case where all budgets are the same (but the graph might have a self-loop), and proposed a $\frac{5+\sqrt{2}}{6+\sqrt{2}} \approx 0.86$-approximation algorithm based on a LP relaxation. In general case, the standard LP is shown to have an integrality gap close to $\frac{1}{4}$ [KKMS09]. Therefore, it is natural to consider hierarchies of LP relaxations such as the Sherali-Adams hierarchy [SA90] (see [CT12] for a general survey and [GTW13, YZ14] for recent algorithmic results using the Sherali-Adams hierarchy). Especially, Chalermsook et al. [CKLN13] recently showed that there is a FPTAS when the graph has bounded treewidth, based on the Sherali-Adams hierarchy. However, the power of the Sherali-Adams hierarchy and SDP, as well as the inherent hardness of the problem, was not well-understood in general case.

7.1.1 Our Results

In this chapter, we show that any polynomial time algorithm that guarantees a ratio better than $\frac{1}{4}$ must be powerful enough to refute the Unique Games Conjecture.

Theorem 7.1.1. *Under the Unique Games Conjecture, for any $\epsilon > 0$, it is NP-hard to approximate* GRAPH PRICING *within a factor of $\frac{1}{4} + \epsilon$.*

By the results of Khot and Vishnoi [KV05] and Raghavendra and Steurer [RS09] that convert a hardness under the UGC to a SDP gap instance, our result unconditionally shows that even a SDP-based algorithm will not improve the performance of a simple algorithm. For the Sherali-Adams hierarchy, we prove that even polynomial rounds of the Sherali-Adams hierarchy has an integrality gap close to $\frac{1}{4}$.

Theorem 7.1.2. *Fix $\epsilon > 0$. There exists $\delta > 0$ such that the integrality gap of n^δ-rounds of the Sherali-Adams hierarchy for* GRAPH PRICING *is at most $\frac{1}{4} + \epsilon$.*

One possible way to prove Theorem 7.1.2 is to *compose* our reduction from UNIQUE GAMES in Theorem 7.1.1 with the integrality gap of the Sherali-Adams hierarchy for UNIQUE GAMES found by Charikar et al. [CMM09]. Our proof extends techniques of [CMM09] directly for our problem to have a more intuitive and efficient gap instance.

Our result is based on an interesting generalization of MAX DICUT, which we call GENERALIZED MAX DICUT. It is parameterized by a positive integer $T \geq 1$. An instance consists of a directed graph $D = (V, A)$ and a label on each edge $l_A : E \to \{1, \ldots, T\}$, where the goal is to assign to each vertex v a label $l_V(v)$ from $\{0, \ldots, T\}$ to maximize the number of satisfied edges — each edge (u, v) is satisfied if $l_V(u) = 0$ and $l_V(v) = l_A(u, v)$.

This problem shares many properties with GRAPH PRICING, including a simple combinatorial $\frac{1}{4}$-approximation algorithm. There is an approximation-preserving reduction from GENERALIZED MAX DICUT(T) on directed acyclic graphs (DAGs) to GRAPH PRICING for any T. We prove the following theorems that it is hard to improve upon this simple algorithm for large T even on DAGs, which immediately imply Theorem 7.1.1 and 7.1.2.

Theorem 7.1.3. *Under the Unique Games Conjecture, it is NP-hard to approximate* GENERALIZED MAX DICUT*(T) on directed acyclic graphs within a factor of $\frac{1}{4} + O(\frac{1}{T^{1/4}})$.*

Theorem 7.1.4. *Fix T and $\epsilon > 0$. There exists $\delta > 0$ such that the integrality gap of n^δ-rounds of the Sherali-Adams for* GENERALIZED MAX DICUT*(T) is at most $\frac{T+1}{4T}(1 + \epsilon)$. Furthermore, the same result holds even when the graph is acyclic.*

It is also interesting to compare the above results to other arity two Constraint Satisfaction Problems (CSPs), since whether the domain is Boolean (e.g. MAX CUT, MAX 2-SAT [GW95]) or not (e.g. 2-CSP with bounded domain [Hås08], UNIQUE GAMES [CMM06]), SDP-based algorithms give a strictly better guarantee than LP-based or combinatorial algorithms. As discussed above, our result unconditionally says that a SDP-based algorithm cannot outperform a simple combinatorial algorithm for this arity two CSP (as T increases).[1]

[1] Formally, (approximation ratio of the SDP-based algorithm) / (approximation ratio of the best known combinatorial algorithm) = $1 + O(\frac{1}{T^{1/4}})$ for GENERALIZED MAX DICUT. For UNIQUE GAMES with T labels, the SDP-based algorithm of Charikar et al. [CMM06], which satisfies roughly $T^{-\epsilon/(2-\epsilon)}$ fraction of constraints in an $(1 - \epsilon)$-satisfiable instance, performs better than the random assignment by any constant factor as T increases.

7.1.2 Related Work and Our Techniques

Formulation of GENERALIZED MAX DICUT. Our conceptual contribution is the introduction of GENERALIZED MAX DICUT as a CSP that captures the complexity of GRAPH PRICING. It is inspired by the work of Khandekar et al. [KKMS09], and our reduction is the almost same as their reduction from MAXIMUM ACYCLIC SUBGRAPH to GRAPH PRICING.

In the natural formulation of GRAPH PRICING as a CSP, each vertex is assigned an (half-)integer price from 0 to B for the maximum budget B, and each customer becomes multiple constraints on two variables since the payoff linearly depends on the prices. It is shown in [KKMS09] that a half-integral optimal solution always exists for integral budgets, so this is a (almost) valid relaxation. However, as each customer becomes multiple constraints with different payoffs, it seems hard to apply current techniques developed for well-studied CSPs to this formulation.

Khandekar et al.'s main idea was to use two well-known CSPs — MAS for the hardness of approximation and MAX DICUT on directed acyclic graphs for the integrality gap of the standard LP. The former is harder to approximate, and the latter has the lower optimum.[2] GENERALIZED MAX DICUT seems to combine ingredients of both problems needed for GRAPH PRICING. It certainly inherits properties of MAX DICUT including low integral optima, but is much harder to approximate than MAX DICUT by Theorem 7.1.3.

Uniques Games-Hardness. Proving hardness of GENERALIZED MAX DICUT on general graphs is relatively straightforward — proposing a dictatorship test with high completeness and low soundness, and plugging it into the *recipe* of Khot et al. [KKMO07] to deduce the hardness result. The dictatorship test is an instance of GENERALIZED MAX DICUT with the set of vertices $\{0, ..., T\}^R$ (called *hypercube*) for some $R \in \mathbb{N}$. The main question in constructing a dictatorship test is how to sample $(x, y) \in \{0, \dots, T\}^2$, which induces a distribution on $\{0, \dots, T\}^2$. In GENERALIZED MAX DICUT, 0 is the only special label such that every directed edge is satisfied only if its tail is assigned 0. The simple combinatorial algorithm samples 0 heavily — the marginal distribution satisfies $\Pr[x = 0] \geq 0.5$, while the solution to the Sherali-Adams hierarchy constructed in Theorem 7.1.4 treats 0 as other labels, having $\Pr[x = 0] = \frac{1}{T+1}$. The latter distribution had a disadvantage that x and y are perfectly correlated — the value of x determines the

[2] Under the Unique Games Conjecture, the best inapproximability ratio is 0.5 for MAXIMUM ACYCLIC SUBGRAPH [GHM$^+$11] and 0.874 for MAX DICUT [Aus10]. For the lower bound on integral optima, the maximum acyclic subgraph always has at least half of edges, while there is a directed acyclic graph where every directed cut cannot have more than $\frac{1}{4} + \epsilon$ fraction of edges for any $\epsilon > 0$.

value of y.

To show the hardness based on the UGC (roughly equivalent to constructing a solution that *fools* SDP), we found that $\Pr[x = 0] = \frac{1}{T^{1/4}}$ is enough. In this case, we can ensure that the probability that dictators pass the test is large, while x and y behave almost independently. Based on the low correlation, we use the result of Mossel [Mos10] to show low soundness.

The resulting dictatorship test is not a DAG. To fix this problem, the final dictatorship test has the vertex set $V \times [T]^R$ for some DAG $D = (V, A)$. For each edge $(u, v) \in A$, the above dictatorship test is performed so that each edge of the dictatorship test goes from the hypercube associated with u to the one with v. This idea of keeping the dictatorship test acyclic is used in Svensson [Sve13], where he takes (the undirected version of) D to be a complete graph. We take a nontrivial DAG found by Alon et al. [ABG$^+$07] where any directed cut has at most $\frac{1}{4} + o(1)$ fraction of edges. In the soundness case, if every hypercube is *pseudorandom*, the soundness analysis of an individual dictatorship test associated with each edge gives a rounding algorithm that finds a large directed cut in D, which contradicts the choice of D.

This style of argument, composing the dictatorship test with a certain instance and solving this instance by the soundness analysis, resembles that of Raghavendra [Rag08] for CSPs, Guruswami et al. [GHM$^+$11] for ordering CSPs, Kumar et al. [KMTV11] for strict CSPs, and Guruswami and Saket [GS10] for k-uniform k-partite Hypergraph Vertex Cover. While they require the instance to have a good fractional solution (LP or SDP) but the low integral optimum, we only need the low integral optimum (of even a simpler problem) and our individual dictatorship test ensures completeness and part of soundness. We hope that this two-level technique — constructing a simple dictatorship test for each edge and composing it with a certain instance with purely combinatorial properties — makes it easier to bypass the barrier of finding a gap instance and prove hardness for many other problems, especially those with structured instances.

Sherali-Adams Gap. On the integrality gap of GENERALIZED MAX DICUT on a DAG, our work generalizes the work of Charikar et al. [CMM09], which showed a similar result for MAX CUT, in several directions. The first obstacle is to find a DAG with a low integral optimum which is amenable to construct a good solution to the Sherali-Adams hierarchy. Previous works which obtained lower bounds for the Sherali-Adams hierarchy [ABLT06, dlVKM07, CMM09] used $G(n, p)$, but $G(n, p)$ with an consistent orientation will not result in a low integral optimum. Instead, we show that sparsifying the aforementioned graph constructed in Alon et al. [ABG$^+$07], which is already a DAG with a low integral optimum, gives other desired properties as well.

Given a set S of k vertices, we define a local distribution on the events $\{l_V(v) = i\}_{v \in S, i \in T}$. One caveat of the above approach is that local distributions obtained might be inconsistent, in a sense that S and S' might induce different marginal distributions on $S \cap S'$. Charikar et al.'s main idea is to embed them into l_2 and use hyperplane rounding to produce consistent ones. The most technical part of our work is to extend the hyperplane rounding to work for non-Boolean domains. It is a complicated task in general, but we use the fact that the embedding is explicitly constructed for two adjacent vertices and it exhibits some symmetry, so that we can analyze the performance of our rounding. For $T = 1$, our result matches that of [CMM09].

7.1.3 Organization

Section 7.2 introduces problems and notations formally. Section 7.4 and Section 7.5 present Unique Games-hardness and Sherali-Adams integrality gaps of GENERALIZED MAX DICUT respectively, which can be combined with the reduction in Section 7.3 to give the same results for GRAPH PRICING.

7.2 Preliminaries

For any positive integer n, let $[n]^0 := \{0, 1, 2, \ldots, n\}$ and $[n] := \{1, 2, \ldots, n\}$. Given a sequence of numbers $a_1, ..., a_n$, let $\max2_j[a_j]$ be the second largest number among a_j's.

GRAPH PRICING. An instance of GRAPH PRICING consists of an undirected (possibly contain parallel edges) graph $G = (V, E)$ with budgets $b : E \to \mathbb{R}^+$ and weights $w : E \to \mathbb{R}^+$. Our goal is to find a pricing $p : V \to \mathbb{R}^+ \cup \{0\}$ to maximize

$$\mathsf{Val}(p) := \sum_{e=(u,v) \in E} w(e)(p(u) + p(v))\mathbb{I}[p(u) + p(v) \leq b(e)]$$

where $\mathbb{I}[\cdot]$ is the indicator function. Let $\mathsf{Opt}(G, b, w) := \max_p \mathsf{Val}(p)$.

Remark 7.2.1. *This definition of* GRAPH PRICING *above coincides with General* GRAPH PRICING *defined in Khandekar et al. [KKMS09]. They presented an additional reduction from General* GRAPH PRICING *to* GRAPH PRICING *with no parallel edge and $w(e) = 1$. Throughout this paper, we use the definition above and allow weights and parallel edges for simplicity. In practice, weights can be naturally interpreted as the number of customers interested in the same pair.*

Remark 7.2.2. *Another well-known pricing problem assumes that each customer will buy the cheapest item of her interest if she can afford it, which means that the value of the pricing p becomes*

$$\mathsf{Val}(p) := \sum_{e=(u,v)} w(e) \min(p(u), p(v)) \mathbb{I}[\min(p(u), p(v)) \le b(e)].$$

This is called unit-demand pricing. *Its approximability is similar to that of our single-minded pricing, including algorithms / hardness results for k-Hypergraph Pricing for large k [CLN13], and a simple $\frac{1}{4}$-approximation algorithm for* GRAPH PRICING $(k = 2)$. *Indeed,* GENERALIZED MAX DICUT *is also reducible to Unit-demand* GRAPH PRICING *and Theorem 7.1.1 and 7.1.2 hold for it as well. We focus on Single-minded* GRAPH PRICING *here.*

GENERALIZED MAX DICUT. Fix a positive integer T. An instance of GENERALIZED MAX DICUT(T) consists of a digraph $D = (V, A)$ with a label $l_A : A \to [T]$ and a weight $w : A \to \mathbb{R}^+$ on each edge. Assume that the sum of weights is normalized to 1. (u, v) denotes the edge of D from u to v. We allow parallel edges from u to v if they have different labels (if parallel edges have the same label, simply merge them). Our goal is to find a labeling $l_V : V \to [T]^0$ (note vertices can be assigned 0, while edges are not) to maximize the weight of *satisfied* edges — (u, v) is satisfied when $l_V(u) = 0$ and $l_V(v) = l_A(u, v)$. Note than when $T = 1$, the problem becomes MAX DICUT. Given an instance $D = (V, A)$, l_A, and w, let $\mathsf{Opt}(G, l_A, w)$ be the maximum weight of edges satisfied by any labeling of vertices. Given an assignment $l_V : V \to [T]^0$ to the vertices, let $\mathsf{Val}(l_V)$ be the weight of edges satisfied by l_V. Note that unlike GRAPH PRICING, the value of any assignment is normalized between 0 and 1. The *normalized outdegree*, denoted by ndeg, is defined to be $[\sum_u (\max_{(u,v)\in A} w(u, v))]^{-1}$. In unweighted instances (i.e. $w(e) = \frac{1}{|A|}$ for all e), $\mathsf{ndeg} \ge \frac{|A|}{|V|}$.

Sherali-Adams Hierarchy. In its most intuitive and redundant form, a feasible solution to the r-rounds of the Sherali-Adams hierarchy for a CSP with the domain $[q]^0$ consists of $\sum_{i=1}^{r} \binom{n}{r}(q+1)^r$ variables $\{x_S(\alpha)\}$ for each subset of variables S with cardinality at most r, and $\alpha \in ([q]^0)^S$. Each $x_S(\alpha)$ can be interpreted as the probability that the variables in S are assigned α. Therefore, it is required to satisfy the following natural conditions: (1) $x_S(\alpha) \ge 0$ for all S, α. (2) $\sum_{\alpha\in([q]^0)^S} x_S(\alpha) = 1$ for all S. (3) $\sum_{\alpha\in([q]^0)^{S'\setminus S}} x_{S'}(\alpha \circ \beta) = x_S(\beta)$ for all $S \subseteq S', \beta \in ([q]^0)^S$, where $\alpha \circ \beta \in ([q]^0)^{S'\cup S}$ denote the joint assignment to the variables in S'.

The r-rounds of the Sherali-Adams hierarchy for GRAPH PRICING and GENERALIZED MAX DICUT(T) can be obtained by choosing an appropriate domain and an objective function, while using the constraints given above. For GRAPH PRICING, if we choose the domain to be $[B]^0$ where B is the maximum budget, the objective function is the following.

$$\sum_{e=(u,v)} w(e) \sum_{(i,j)\in([B]^0)^2, i+j\le b(u,v)} (i+j)\cdot x_{(u,v)}(i,j)$$

Since $p(v)$ can be real, it is not clear whether this is a relaxation, even when the budgets are integers. [KKMS09] shows that there is a half-integral optimal solution. The maximum budget B can be exponentially big in the size of an instance, and a standard trick is to consider only the powers of $(1+\epsilon)$ as valid prices. It loses at most ϵ fraction of the optimum. Our gap instance and proposed solution to the hierarchy have the marginal on each vertex supported by a constant number of prices, so they are applicable to any choice of the domain.

For GENERALIZED MAX DICUT(T), the domain is $[T]^0$, and the objective function is

$$\sum_{(u,v)\in A} w(u,v) x_{(u,v)}(0, l_A(u,v)).$$

Given an instance and a relaxation, we define the *integrality gap* to be the integral optimum divided by the value of the best solution to the relaxation. Since both our problems are maximization problems, it is at most 1 and a small number indicates a large gap.

7.3 Reduction from GENERALIZED MAX DICUT to GRAPH PRICING

Theorem 7.3.1. *For any $T > 0$, there is a polynomial time reduction from an instance $(D = (V, A), l_A, w_{\mathsf{GMD}})$ of* GENERALIZED MAX DICUT*(T), where D is acyclic and* $\mathsf{ndeg} \ge \frac{1}{\epsilon}$*, to an instance (G, b, w_{GP}) of* GRAPH PRICING *such that* $\mathsf{Opt}(D, l_A, w_{\mathsf{GMD}}) \le \mathsf{Opt}(G, b, w_{\mathsf{GP}}) \le \mathsf{Opt}(D, l_A, w_{\mathsf{GMD}}) + 3\epsilon$.

Proof. Fix an instance $(D = (V, A), l_A, w_{\mathsf{GMD}})$ of GENERALIZED MAX DICUT(T) with $n = |V|$ and $m = |A|$. Let G be the underlying undirected graph of D. Our reduction from GENERALIZED MAX DICUT on directed acyclic graphs to GRAPH PRICING is almost the same as the one in Khandekar et al. [KKMS09] with some simplification. Let M be a large number which will be fixed later.

The resulting instance of GRAPH PRICING is based on the same graph G. Since D is acyclic, there is an injective function $s : V \to [n]$ such that for each edge $(u, v) \in A$, $s(u) > s(v)$. For each edge $(u, v) \in A$, $b(u, v) = M^{Ts(v)+l_A(u,v)-1}$ and $w_{\mathsf{GP}}(u, v) = \frac{w_{\mathsf{GMD}}(u,v)}{b(u,v)}$.

To avoid confusion, let $\mathsf{Opt}_{\mathsf{GMD}}$, $\mathsf{Val}_{\mathsf{GMD}}$ denote Opt, Val for GENERALIZED MAX DICUT instances, and $\mathsf{Opt}_{\mathsf{GP}}$ and $\mathsf{Val}_{\mathsf{GP}}$ for GRAPH PRICING instances. Fix a labeling $l_V : V \to [T]^0$. The corresponding *canonical solution* $p : V \to \mathbb{R}^+ \cup \{0\}$ defined by

$$p(v) = \begin{cases} M^{Ts(v)+l_V(v)-1} & \text{if } l_V(v) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

gives $\mathsf{Val}_{\mathsf{GP}}(p) \geq \mathsf{Val}_{\mathsf{GMD}}(l_V)$ — for each $(u, v) \in A$ satisfied by l_V, p gets $p(v)w_{\mathsf{GP}}(u, v) = w_{\mathsf{GMD}}(u, v)$. Therefore, $\mathsf{Opt}_{\mathsf{GP}}(G, b, w_{\mathsf{GP}}) \geq \mathsf{Opt}_{\mathsf{GMD}}(D, l_A, w_{\mathsf{GMD}})$. The following lemma shows that the converse is almost true.

Lemma 7.3.2 ([KKMS09]). *For any p, $\mathsf{Val}_{\mathsf{GP}}(p) \leq \mathsf{Opt}_{\mathsf{GMD}}(D, l_A, w_{\mathsf{GMD}}) + \frac{1}{M} + 2\epsilon$.*

Proof. Given p, we define the *principal part* of $\mathsf{Val}_{\mathsf{GP}}(p)$ as

$$\sum_{(u,v)\in A} w_{\mathsf{GP}}(u, v)p(v)\mathbb{I}[p(u) + p(v) \leq b(u, v)].$$

Note that for each directed edge, only the price of its head contributes.

We first bound the principal part of $\mathsf{Val}_{\mathsf{GP}}(p)$. For a vertex v, the only edges where $w_{\mathsf{GP}}(u, v)p(v) > \frac{w_{\mathsf{GMD}}(u,v)}{M}$ satisfy $M^{Ts(v)+l_A(u,v)-2} < p(v) \leq M^{Ts(v)+l_A(u,v)-1}$. If there is such an edge, let $l'_V(v) = l_A(u, v)$. Otherwise, let $l'_V(v) = 0$. Fix an edge (u, v) where $w_{\mathsf{GP}}(u, v)p(v) > \frac{w_{\mathsf{GMD}}(u,v)}{M}$. $l'_V(v) = l_A(u, v)$ by above. If $l'_V(u) \neq 0$, it means $p(u) > M^{Ts(u)-1} \geq M^{Ts(v)+T-1} \geq b(u, v)$, so (u, v) contributes 0 to the principal part of $\mathsf{Val}_{\mathsf{GP}}(p)$. Therefore, for each edge (u, v) that contributes more than $\frac{1}{M}$ to the principal part of $\mathsf{Val}_{\mathsf{GP}}(p)$, l'_V satisfies (u, v). Therefore, the principal part of $\mathsf{Val}_{\mathsf{GP}}(p)$ is at most $\mathsf{Opt}_{\mathsf{GMD}}(D, l_A, w_{\mathsf{GMD}}) + \frac{1}{M}$.

For the non-principal part of $\mathsf{Val}_{\mathsf{GP}}(p)$, for each vertex u, and we bound

$$\sum_{(u,v)\in A} w_{\mathsf{GP}}(u, v)p(u)\mathbb{I}[p(u) + p(v) \leq b(u, v)] \leq \sum_{(u,v)\in A, p(u)\leq b(u,v)} w_{\mathsf{GMD}}(u, v)\frac{p(u)}{b(u, v)}.$$

Note that all edges (u, v) have different $b(u, v)$, and any two differ by at least a factor of M. Let $w_u := \max_{(u,v)\in A} w_{\mathsf{GMD}}(u, v)$. Therefore, the right hand side can be bounded by

$w_u(1 + \frac{1}{M} + \frac{1}{M^2} + \dots) \le 2w_u$, where

$$\sum_u w_u = \frac{1}{\mathsf{ndeg}} \le \epsilon.$$

This shows that the non-principal part of $\mathsf{Val}_{\mathsf{GP}}(p)$ is at most 2ϵ, proving the lemma. $\square$

Taking $M \ge \frac{1}{\epsilon}$ proves the theorem. $\square$

7.4 Approximability of GENERALIZED MAX DICUT

Recall that GENERALIZED MAX DICUT(1) is exactly the well-known MAX DICUT problem, which admits a 0.874-approximation algorithm [LLZ02] as any MAX 2-CSP over the Boolean domain. As T increases, however, the best approximation ratio for MAX 2-CSP over the domain of size $T + 1$ can be at most $O(\frac{\log T}{\sqrt{T}})$ [Cha13], so viewing it as a general MAX 2-CSP does not yield a constant-factor approximation algorithm.

There is a simple $\frac{1}{4}$-approximation algorithm, similar to the one for GRAPH PRICING — assign 0 to each vertex with probability half independently and assign nonzero values to the remaining vertices greedily. The proof is based on the fact that we can easily find the optimal solution once the set of vertices assigned 0 is given. For small T, we can do a little better based on a standard LP relaxation. The proof is given in Section 7.6.

Theorem 7.4.1. *There is a polynomial time approximation algorithm for* GENERALIZED MAX DICUT*(T) that guarantees $\frac{1}{4} + \Omega(\frac{1}{T})$ of the optimal solution.*

However, we prove that for large T, it is Unique Games-hard to improve the approximation ratio from $\frac{1}{4}$ to a better constant.

Theorem 7.4.2 (Restatement of Theorem 7.1.3)**.** *Under the Unique Games Conjecture, it is NP-hard to approximate* GENERALIZED MAX DICUT*(T) on directed acyclic graphs within a factor of $\frac{1}{4} + O(\frac{1}{T^{1/4}})$.*

Together with the reduction shown in Theorem 7.3.1, it immediately implies Theorem 7.1.1 for GRAPH PRICING. Besides working on DAGs, the reduction also requires that ndeg be large, but it can be easily ensured by taking an UNIQUE GAMES instance with large degree. See Section 7.4.3 to see the full details.

The theorem is proved by proposing a *dictatorship test* with high completeness and low soundness, combined with the standard technique to convert a dictatorship test to a hardness result based on the Unique Games Conjecture [KKMO07]. Constructing the dictatorship test has two components — a simple dictatorship test based on correlation and Gaussian geometry, and composing it with a designated DAG.

7.4.1 Dictatorship Test

Consider the hypercube $([T]^0)^R$ where $[T]^0 = \{0, 1, \ldots, T\}$. Let $\Omega_1 = \Omega_2 = [T]^0$. For $t \in [T]$, $\mathbf{P}^t$ is a probability measure on $\Omega_1 \times \Omega_2$. Let $\mathbf{P}$ be the marginal on Ω_i in $\mathbf{P}^t$ (which does not depend on t and i). We want to ensure that $\mathbf{P}(0) = \delta$, $\mathbf{P}(j) = \frac{1-\delta}{T}$ for $j \in [T]$ where $\delta = \frac{1}{T^{1/4}}$. Let $\mathbf{P}'$ be the distribution on Ω_1 such that $\mathbf{P}'(0) = (\frac{1}{1-\frac{1-\delta}{T}})(\delta - \frac{1-\delta}{T})$, $\mathbf{P}'(j) = (\frac{1}{1-\frac{1-\delta}{T}})(\frac{1-\delta}{T})$ (subtract $\frac{1-\delta}{T}$ from $\mathbf{P}(0)$ and renormalize). $\mathbf{P}^t$ is defined by the following procedure to sample (x, y). Sample y according to $\mathbf{P}$. If $y = t$, set $x = 0$. Otherwise, sample x from $\mathbf{P}'$ independently. It is easy to see that the marginal of both x and y is $\mathbf{P}$. We show that (x, y) are almost independent as T increases. Recall from Definition 3.3.5 that given a distribution $\mathbf{Q}$ on $\Omega_1 \times \Omega_2$, the correlation between two correlated spaces is defined as

$$\rho(\Omega_1, \Omega_2; \mathbf{Q}) = \sup\left\{\mathsf{Cov}[f, g] : f : \Omega_1 \to \mathbb{R}, g : \Omega_2 \to \mathbb{R}, \mathsf{Var}[f] = \mathsf{Var}[g] = 1\right\}.$$

Lemma 7.4.3. *For any t, $\rho(\Omega_1, \Omega_2; \mathbf{P}^t) \leq \sqrt{\frac{2}{T\delta}}$.*

Proof. Let $f : \Omega_1 \to \mathbb{R}$ be the function satisfying $\mathbb{E}[f] = 0$, $\mathbb{E}[f^2] = 1$. Let L be the Markov operator defined in Section 2.1 of Mossel [Mos10] such that

$$(Lf)(y) = \mathbb{E}[f(X)|Y = y]$$

for $y \in \Omega_2$ and $(X, Y) \in \Omega_1 \times \Omega_2$ is distributed according to $\mathbf{P}^t$. By Lemma 2.8 of [Mos10],

$$\rho(\Omega_1, \Omega_2) = \sup_f \sqrt{\mathbb{E}[(Lf)^2]}.$$

Let $f(i) = a_i$, $(Lf)(i) = b_i$ for $i \in [T]^0$. $b_t = a_0$ and all the other b_i's are equal to $\mathbb{E}_{\mathbf{P}'_1}[f]$, which is equal to $(\frac{1}{1-\frac{1-\delta}{T}})(\mathbb{E}_{\mathbf{P}_1}[f] - \frac{1-\delta}{T}a_0) = (\frac{1}{1-\frac{1-\delta}{T}})(-\frac{1-\delta}{T}a_0)$.

$$\begin{aligned}
\mathbb{E}[(Lf)^2] &= (\frac{1-\delta}{T})a_0^2 + (1-\frac{1-\delta}{T})[(\frac{1}{1-\frac{1-\delta}{T}})(-\frac{1-\delta}{T}a_0)]^2 \\
&= (\frac{1-\delta}{T})a_0^2 + (\frac{1}{1-\frac{1-\delta}{T}})(\frac{1-\delta}{T}a_0)^2 \\
&= (\frac{1-\delta}{T})a_0^2[1 + (\frac{1}{1-\frac{1-\delta}{T}})(\frac{1-\delta}{T})] \\
&\leq \frac{2}{T}a_0^2 \\
&\leq \frac{2}{T\delta}
\end{aligned}$$

Since $\delta a_0^2 \leq \mathbb{E}[f^2] \leq 1$. □

Another component of the dictatorship test is the directed acyclic graph $D = (V, A)$ of Alon et al. [ABG+07], where every directed cut has size at most $(\frac{1}{4} + o(1))|A|$. Fix a graph $D = (V, A)$ such that every dicut cuts at most $(\frac{1}{4} + \frac{1}{T^{1/4}})|A|$ edges. Note that the size of this graph depends only on T. We now describe the dictatorship test. The prover is expected to provide $F_v : ([T]^0)^R \to ([T]^0)$ for each $v \in V$.

1. Choose $(u, v) \in A$ and $t \in [T]$ uniformly at random.
2. For each $i \in [R]$, pick (x_i, y_i) according to $\mathbf{P}^t$.
3. Accept if $F_u(x) = 0$ and $F_v(y) = t$.

This dictatorship test can be naturally interpreted as an instance of GENERALIZED MAX DICUT(T) with the vertex set $V \times ([T]^0)^R$. The weight of edge $((u, x), (v, y))$ with label t is equal to the probability that it is sampled, and a labeling $l : V \times ([T]^0)^R \mapsto [T]^0$ passes with probability $\mathsf{Val}(l)$ (by $F_v(x) = l(v, x)$).

7.4.2 Completeness and Soundness

The ith dictator function is $D_i : ([T]^0)^R \to [T]^0$ given by $D_i(x_1, \ldots, x_R) = x_i$. The purpose of the above dictatorship test is to allow dictatorship functions to be accepted with high probability while penalizing functions *far from* any dictator. The following lemma for completeness is immediate from the test — for any fixed t and i, $\Pr[x_i = 0, y_i = t] = \Pr[y_i = t] = \frac{1-\delta}{T}$.

Lemma 7.4.4 (Completeness). *Suppose that for some i, $F_v = D_i$ for all $v \in V$. The above test accepts with probability $\frac{1-\delta}{T}$.*

For $v \in V$ and $t \in [T]^0$, let $F_{v,t} : ([T]^0)^R \to \{0,1\}$ be defined such that $F_{v,t}(x) = 1$ iff $F_v(x) = t$, and $\mu_{v,t} := \Pr[F_v(x) = t] = \mathbb{E}[F_{v,t}(x)]$ where $x \sim \mathbf{P}$. For each $F_{v,t}$ and $i \in [R]$, let $\mathsf{Inf}_i[F_{v,t}]$ and $\mathsf{Inf}_i^{\leq d}(F_{v,t})$ the influence and the low-degree influence defined in Definition 3.3.5 and 3.3.7.

Lemma 7.4.5 (Soundness). *For large enough T, there exist τ and d (depending on T) such that if $\mathsf{Inf}_i^{\leq d}(F_{v,t}) \leq \tau$ for all $i \in [R]$, $t \in [T]^0$, and $v \in V$, the probability of accepting is at most $\frac{1}{4T} + \frac{4}{T^{5/4}}$.*

Proof. Applying Theorem 3.3.10 (set $\epsilon \leftarrow \frac{1}{T^{5/4}}$ and $\alpha = \Theta(\frac{1}{T^2})$), the probability of accepting is at most

$$\mathbb{E}_{(u,v)\in A}[\mathbb{E}_{t\in[T]}[\mathbb{E}_{(x,y)\sim(\mathbf{P}^t)^{\otimes R}}[F_{u,0}(x)F_{v,t}(y)]]] \leq \mathbb{E}_{(u,v)\in A}[\mathbb{E}_{t\in[T]}[\overline{\Gamma}_\rho(\mu_{u,0}, \mu_{v,t}) + \frac{1}{T^{5/4}}]].$$

The following lemma, whose proof is given in Section 7.7, shows that it is at most

$$\mathbb{E}_{(u,v)\in A}[\overline{\Gamma}_\rho(\mu_{u,0}, \frac{1-\mu_{v,0}}{T})] + \frac{1}{T^{5/4}}.$$

Lemma 7.4.6. *Fix $\rho, a \in (0,1)$. The function $f(b) := \overline{\Gamma}_\rho(a,b)$ is concave.*

The following lemma, whose proof is again given in Section 7.7, shows that it is at most

$$\mathbb{E}_{(u,v)\in A}[\mu_{u,0}\frac{(1-\mu_{v,0})}{T} + \frac{2}{T^{5/4}}] + \frac{1}{T^{5/4}} = \frac{1}{T}\mathbb{E}_{(u,v)\in A}[\mu_{u,0}(1-\mu_{v,0})] + \frac{3}{T^{5/4}}.$$

Lemma 7.4.7. *For large enough T and $\delta = \frac{1}{T^{1/4}}$, the following holds. For any $a \in [0,1], b \in [0, \frac{1}{T}]$ and $\rho \in (0, \sqrt{\frac{2}{T\delta}})$, $\overline{\Gamma}_\rho(a,b) \leq ab + \frac{2}{T^{5/4}}$.*

Given $\{\mu_{v,0}\}_{v\in V}$, imagine the rounding algorithm which puts $v \in S$ with probability $\mu_{v,0}$ independently. The expected fraction of edges from S to $V \setminus S$ is $\mathbb{E}_{(u,v)\in A}[\mu_{u,0}(1 - \mu_{v,0})]$, which is at most the fractional size of maximum dicut of D. Since we took D to satisfy that $\mathbb{E}_{(u,v)\in A}[\mu_{u,0}(1-\mu_{v,0})] \leq \frac{1}{4} + \frac{1}{T^{1/4}}$, the probability of accepting is at most $\frac{1}{4T} + \frac{4}{T^{5/4}}$ as desired. Note that the probabilities of accepting in completeness and soundness differ by a factor of $\frac{\frac{1}{4T}+\frac{4}{T^{5/4}}}{\frac{1-\delta}{T}} = \frac{\frac{1}{4}+\frac{4}{T^{1/4}}}{1-\frac{1}{T^{1/4}}} = \frac{1}{4} + O(\frac{1}{T^{1/4}})$. $\square$

7.4.3 Reduction from UNIQUE GAMES

In this subsection, we introduce the reduction from the UNIQUE GAMES to GENERALIZED MAX DICUT(T), using the dictatorship test constructed.

Theorem 7.4.8 (Restatement of Theorem 7.1.3)**.** *Under the Unique Games Conjecture, it is NP-hard to approximate* GENERALIZED MAX DICUT*(T) on directed acyclic graphs within a factor of* $\frac{1}{4} + O(\frac{1}{T^{1/4}})$.

Proof. Given an instance of $\mathcal{L}(G(U \cup W, E), [R], \{\pi(v,w)\}_{(v,w)\in E})$ of UNIQUE GAMES, we construct an instance $\mathcal{D}(\mathcal{V}, \mathcal{A}), l_{\mathcal{A}}$ of GENERALIZED MAX DICUT(T). For $x \in ([T]^0)^R$ and a permutation $\pi : [R] \to [R]$, let $x \circ \pi \in ([T]^0)^R$ be defined by $(x \circ \pi)_i = (x)_{\pi^{-1}(i)}$. Let $D = (V, A)$ be the fixed-size graph where the maximum dicut has at most $(\frac{1}{4} + \frac{1}{T^{1/4}})$ fraction of edges.

- $\mathcal{V} = U \times V \times ([T]^0)^R$.
- Sample $w \in W$ uniformly at random and its neighbors u_1, u_2 uniformly and independently. Sample $t \in [T]$, $(v_1, v_2) \in A$, and $x, y \in ([T]^0)^R$ from the dictatorship test. Add an edge $((u_1, v_1, x \circ \pi_{u_1,w}), (u_2, v_2, y \circ \pi_{u_2,w}))$ to $\mathcal{A}$ with label t. The weight is equal to the probability that this edge is sampled.

Completeness. Suppose that $\mathsf{Val}_{\mathsf{UG}}(l) \geq 1 - \alpha$ for some labeling $l : U \cup W \to [R]$.

Set $l_{\mathcal{V}}(u, v, (x_1, \ldots, x_R)) = x_{l(u)}$. For w, u_1, u_2 sampled as above, with probability $1 - 2\alpha$, $\pi(u_1, w)^{-1}(l(u_1)) = \pi(u_2, w)^{-1}(l(u_2))$. In that case, by Lemma 7.4.4,

$$\begin{aligned}
&\Pr_{v_1,v_2,t,x,y}[l_{\mathcal{V}}(u_1, v_1, x \circ \pi_{u_1,w}) = 0, l_{\mathcal{V}}(u_2, v_2, y \circ \pi_{u_2,w}) = t] \\
=& \Pr_{v_1,v_2,t,x,y}[(x \circ \pi_{u_1,w})_{l(u_1)} = 0, (y \circ \pi_{u_2,w})_{l(u_2)} = t] \\
=& \Pr_{v_1,v_2,t,x,y}[(x)_{\pi(u_1,w)^{-1}(l(u_1))} = 0, (y)_{\pi(u_2,w)^{-1}(l(u_2))} = t] \\
\geq& \frac{1-\delta}{T}.
\end{aligned}$$

Therefore, $\mathsf{Val}_{\mathsf{GMD}}(l_{\mathcal{V}}) \geq \frac{(1-2\alpha)(1-\delta)}{T}$.

Soundness. For each $u \in U, v \in V$ and $t \in [T]^0$, let $F_{u,v,t} : ([T]^0)^R \to \{0, 1\}$ be defined by

$$F_{u,v,t}(x) = 1 \text{ if and only if } l_{\mathcal{V}}(u, v, x) = t.$$

Similarly, for each $w \in W, v \in V$ and $t \in [T]^0$, let $H_{w,v,t} : ([T]^0)^R \to [0,1]$ be the function defined by

$$H_{w,v,t}(x) = \mathbb{E}_{(u,w)\in E}[F_{u,v,t}(x \circ \pi(u,w))] = \Pr_{(u,w)\in E}[l_{\mathcal{V}}(u, v, x \circ \pi(u,w)) = t].$$

Suppose that there exists l_V such that $\mathsf{Val}_{\mathsf{GMD}}(l_V) \geq \frac{1}{4T} + \frac{5}{T^{5/4}}$. For at least $\frac{1}{T^{5/4}}$ fraction of w, an edge of $\mathcal{A}$ sampled by first choosing w is satisfied with probability more than $\frac{1}{4T} + \frac{4}{T^{5/4}}$. By Lemma 7.4.5, there exist τ and d, such that, for each such w, we have $\mathsf{Inf}_i^{\leq d}[H_{w,v,t}] > \tau$ for some i, v and t. Set $l_{\mathcal{V}}(w) = i$. For other w's, choose $l_V(w)$ arbitrarily.

From the representation of influences in terms of Fourier coefficients (equation (6) of Khot et al. [KKMO07]),

$$\tau < \mathsf{Inf}_i^{\leq d}[H_{w,v,t}] \leq \mathbb{E}_{(u,w)\in E}[\mathsf{Inf}_{\pi(u,w)(i)}^{\leq d}[F_{u,v,t}]]$$

and we conclude that $\tau/2$ fraction of neighbors u of w have $\mathsf{Inf}_{\pi(u,w)(i)}^{\leq d}(F_{u,v,t}) \geq \tau/2$. We choose $l_{\mathcal{V}}(u)$ uniformly from

$$\left\{ i : \mathsf{Inf}_i^{\leq d}[F_{u,v,t}] \geq \tau/2 \text{ for some } t, v \right\}.$$

Since $\sum_i \mathsf{Inf}_i^{\leq d}[F_{u,t}] \leq d$, there are at most $\frac{2(T+1)d|V|}{\tau}$ of candidate i's for each u. If u have no candidate, choose $l_V(u)$ arbitrarily. The above strategy satisfies $(\frac{1}{T^{5/4}})(\frac{\tau}{2})(\frac{\tau}{2(T+1)d|V|})$ fraction of constraints in expectation. Taking α small enough completes the proof of the theorem. $\square$

Now, we present the full proof of our main theorem.

Theorem 7.4.9 (Restatement of Theorem 7.1.1)**.** *Under the Unique Games Conjecture, for any $\epsilon > 0$, it is NP-hard to approximate* GRAPH PRICING *within a factor of $\frac{1}{4} + \epsilon$.*

Proof. Given $\epsilon > 0$, let T large enough so that $\frac{1}{T^{1/4}} < \frac{\epsilon}{2}$. Theorem 7.1.3 tells that it is hard to distinguish

- Completeness: $\mathsf{Opt}_{\mathsf{GMD}} \geq \frac{1}{T} - \frac{2}{T^{4/5}} = \frac{1-O(\epsilon)}{T}$.
- Soundness: $\mathsf{Opt}_{\mathsf{GMD}} \leq \frac{1}{4T} + \frac{5}{T^{4/5}} = \frac{1+O(\epsilon)}{4T}$.

Let $t = \frac{T}{\epsilon}$. We can assume that each vertex in the UNIQUE GAMES instance is of degree at least t, since duplicating each vertex v into t copies $v_1, ..., v_t$ and duplicating each constraint (u, v) into t^2 copies $(u_i, v_j)_{1 \le i,j \le t}$ preserves the optimum. Therefore, the instance of GENERALIZED MAX DICUT obtained from the above UNIQUE GAMES instance will have $\mathsf{ndeg} \ge t$. Theorem 7.3.1 shows that it is NP-hard to distinguish

- Completeness: $\mathsf{Opt}_{\mathsf{GP}} \ge \mathsf{Opt}_{\mathsf{GMD}} = \frac{1-O(\epsilon)}{T}$.
- Soundness: $\mathsf{Opt}_{\mathsf{GP}} \le \mathsf{Opt}_{\mathsf{GMD}} + \frac{1}{t} = \frac{1+O(\epsilon)}{4T} + \frac{\epsilon}{T} = \frac{1+O(\epsilon)}{4T}$.

□

7.5 Integrality Gaps for GENERALIZED MAX DICUT

Fix a positive integer T and $\epsilon \in (0, \frac{1}{100})$. We present an instance of GENERALIZED MAX DICUT(T) $(D = (V, A), l_A)$ (we only deal with unweighted instances in this section and omit w) such that D is acyclic, $|V| \le \epsilon |A|$ (so that $\mathsf{ndeg} \ge \frac{1}{\epsilon}$), and a solution to n^{δ}-rounds of the Sherali-Adams hierarchy such that the integrality gap is at most $\frac{T+1}{4T}(1+\epsilon)$. This result almost matches a simple $\frac{1}{4}$-approximation algorithm.

Through the reduction given in Theorem 7.3.1, we also prove Theorem 7.1.2 — a bad integral solution is guaranteed by the reduction, a good solution to the Sherali-Adams hierarchy is obtained by the mapping $l_V(u) = i$ to $p(u) = M^{Ts(u)+i-1}$ (if $i \ne 0$) or 0 (otherwise). The budget in the resulting instance is an integer exponential in the size of instances, and our gap works even for a strong linear programming hierarchy where there is a variable for each vertex v and an integer price i.

The rest of this section is devoted to the proof of Theorem 7.1.4.

7.5.1 Obtaining a Good Instance

Our graph D is obtained by randomly sparsifying the graph $D_* = (V, A_*)$ constructed in Alon et al. [ABG+07], followed by an appropriate postprocessing. D_* is a directed acyclic graph with n vertices and $m_* = \Theta(n^{\frac{5}{3}})$ edges. Its underlying undirected graph $G_* = (V, E_*)$ is a simple graph with the same number of vertices and edges, with the maximum degree $\Delta_* = \Theta(n^{\frac{2}{3}})$. Actually, $V = [n]$ and $(u, v) \in E$ only if $|u - v| \le r$ where $r := \Theta(n^{\frac{2}{3}})$. It has the property that any directed cut has size at most $\frac{m_*}{4} + o(m_*)$ edges.

The first version of $D = (V, A)$ is constructed as the following. $V := V_* = [n]$, and for each edge $(u, v) \in A_*$, put $(u, v) \in A$ with probability $p := \frac{\Delta}{\Delta_*}$ for some Δ to be fixed later. Let $G = (V, E)$ be the underlying undirected graph of D. l_A is obtained by assigning each $l(u, v)$ a random number from $[T]$.

Like previous integrality gap constructions for MAX CUT and VERTEX COVER (e.g. [ABLT06, dlVKM07, STT07, CMM09]) , D must be postprocessed to be amenable to have a Sherali-Adams solution with a large value. Intuitively, we need to have the underlying undirected graph G *locally sparse* — if we look at a neighborhood of a certain vertex, the graph almost looks like a tree. We use the notion of [CMM09] to measure how locally sparse the graph is.

Definition 7.5.1. *We say that G' is l-path decomposable if every 2-connected subgraph H of G' contains a path of length l such that every vertex of the path has degree 2 in H.*

The first version of the instance already has $\mathsf{Opt}(D, l_A) \approx \frac{1}{4T}$ with high probability. In order to make the instance locally sparse, we additionally need to remove some of the edges, but the fraction of removed edges is so small that it does not affect $\mathsf{Opt}(D, l_A)$ too much. As a result, we get the following theorem.

Theorem 7.5.2. *Given T and $\epsilon, \mu > 0$, there exist constants Δ, δ and $l = \Theta(\log n)$ (all constants depending on T and ϵ, μ) such that there is an instance of* GENERALIZED MAX DICUT*(T) (D, l_A) with the underlying undirected graph G with the following properties.*

- *Acyclicity: D is a DAG.*
- *Low integral optimum: $\mathsf{Opt}(D, l_A) \leq \frac{1+\epsilon}{4T}$.*
- *Almost regularity: Maximum degree of G is at most 2Δ, and G has at least $\Omega(\Delta n)$ edges.*
- *Local sparsity: For $k \leq n^{\delta}$, every induced subgraph of G on $(2\Delta)^l k$ vertices is l-path decomposable.*
- *Large noise: For $k \leq n^{\delta}$, $(1 - \mu)^{l/10} \leq \frac{\mu}{5k}$.*

The last condition, large noise, is needed to ensure that in a LP solution, even though adjacent vertices are very correlated to give a large value, far away vertices behave almost independently. The meaning of each condition will be elaborated in later sections.

Proof. Our graph D is obtained by randomly sparsifying the graph $D_* = (V, A_*)$ constructed in Alon et al. [ABG+07] after an appropriate postprocessing. D_* is a directed acyclic graph with n vertices and $m_* = \Theta(n^{\frac{5}{3}})$ edges. Its underlying undirected graph $G_* = (V, E_*)$ is a simple graph with the same number of vertices and edges, with the maximum degree $\Delta_* = \Theta(n^{\frac{2}{3}})$. Actually, $V = [n]$ and $(u, v) \in E$ only if $|u - v| \leq r$ where $r := \Theta(n^{\frac{2}{3}})$. It has the property that any directed cut has size at most $\frac{m_*}{4} + o(m_*)$ edges.

The first version of $D = (V, A)$ is constructed as the following. $V := V_* = [n]$, and for each edge $(u, v) \in A_*$, put $(u, v) \in A$ with probability $p := \frac{\Delta}{\Delta_*}$ for some Δ to be fixed later. Let $G = (V, E)$ be the underlying undirected graph of V. l_A is obtained by assigning each $l(u, v)$ a random number uniformly sampled from $[T]$.

Integral Solution. The following lemma shows that if Δ is big enough, $\mathsf{Opt}(D, l_A)$ is close to $\frac{1}{4T}$.

Lemma 7.5.3. *If G satisfies the above four properties and $\Delta = \Omega(\frac{T \log T}{\epsilon^2})$, then D and l_A obtained by the above process satisfies $\mathsf{Opt}(D, l_A) \leq \frac{1+4\epsilon}{4T}$ with high probability.*

Proof. Fix one assignment $l_V : V \to [T]^0$. For any edge $(u, v) \in A_*$ call it a *candidate* when $l_V(u) = 0, l_V(v) \neq 0$. Note that the number of candidate edges is at most the cardinality of the maximum directed cut of D_*, which is at most $\frac{1+o(1)}{4} m_*$.

For each candidate edge (u, v), the probability that $(u, v) \in A$ with $l_A(u, v) = l_V(v)$ is $\frac{1}{T}$. Therefore, the expected number of satisfied edges is at most $\frac{(1+o(1))\Delta m_*}{4\Delta_* T}$. By Chernoff bound, the probability that it is bigger than $\frac{(1+\epsilon)pm_*}{4T}$ is bounded by $\exp(-\Omega(\frac{\epsilon^2 pm_*}{T})) = \exp(-\Omega(\frac{\epsilon^2 \Delta n}{T}))$. By taking union bound over $(T+1)^n$ different l_V's, the probability that there exists an assignment with more than $\frac{(1+\epsilon)pm_*}{4T}$ satisfied edges is at most

$$\exp(-\Omega(\frac{\epsilon^2 \Delta n}{T})) * \exp(n \log(T+1)) \leq n^{-1}$$

for $\Delta := \Omega(\frac{T \log T}{\epsilon^2})$. Similarly, we can conclude that $|A| \geq (1-\epsilon)m_* p$ with high probability. Therefore, $\mathsf{Opt}(D, l_A)$ is at most $\frac{(1+\epsilon)}{4T(1-\epsilon)} \leq \frac{1+4\epsilon}{4T}$ with high probability. $\square$

The above lemma is the only place where it is desirable to have large $|A| = |E|$. For the rest of this subsection, we are going to delete some edges of D (and G) to satisfy desired properties. Note that in any case, the number of edges deleted is much less than $\epsilon p m_*$ so that each deletion does not hurt the above lemma.

Maximum Degree Control. Since the maximum degree in G_* is Δ_*, expected degree of each vertex $v \in V$ in G is at most $p\Delta_* = \Delta$. Call a vertex $v \in V$ *bad* if it has degree more than 2Δ in G, and call an edge $(u, v) \in E$ bad if either u or v is bad. Fix an edge (u, v). The probability that (u, v) becomes bad given $(u, v) \in E$ is at most $2\exp(-\frac{\Delta}{4})$. The expected number of bad edges is at most $2\exp(-\frac{\Delta}{4})pm_*$, and by Markov's inequality, with probability at least half, the number of bad edges is at most $4\exp(-\frac{\Delta}{3})pm_*$.

Deleting all bad edges guarantees that the maximum degree of G is at most 2Δ, and with probability at least half, we delete only $4\exp(-\frac{\Delta}{3})pm_*$ edges, which is much smaller than ϵpm_* since $\Delta = \Omega(\frac{1}{\epsilon^2})$.

Girth Control. The expected number of cycles of length i is bounded by

$$n(2r)^{i-1}p^i = n(2r)^{i-1}(\frac{\Delta}{\Delta^*})^i \leq \frac{n(C\Delta)^i}{\Delta_*}$$

for some absolute constant C. When $i = O(\frac{\log n}{\log \Delta})$ the above quantity becomes less than $n^{0.5}$. Assume $l = O(\frac{\log n}{\log \Delta})$ (it will be fixed even smaller than that later). Summing over $i = 4, \ldots, l$ ensures that the expected number of cycles of length up to l is at most $O(n^{0.6})$, and it is less than $O(n^{0.7})$ with high probability. Removing one edge for each cycle of length up to l ensures that G has girth at least l.

Local Sparsity Control. Let $\eta = \frac{1}{3l}$ for some l fixed later. We want to show that there exists $\overline{\Gamma} > 0$ such that every subgraph G' of G induced on $t \leq n^{\overline{\Gamma}}$ vertices have only $(1 + \eta)t$ edges.

For $4 \leq t \leq 1/\eta$, we count the number of connected subgraphs of G_* with t vertices and $t + 1$ edges.

Lemma 7.5.4. *The number of connected subgraphs of G_* with t vertices and $t + 1$ edges is bounded by $2nt^2\Delta_*^{t-1}$.*

Proof. For each of such subgraphs, the only possible degree sequences are $(4, 2, 2, 2, \ldots)$ or $(3, 3, 2, 2, \ldots)$. Assume that it is $(4, 2, 2, 2, \ldots)$. Let v be the vertex with degree 4. There is a sequence of $t + 2$ vertices $(v, \ldots, v, \ldots, v)$ representing an Eulerian tour (not necessarily unique). The number of such sequences is bounded by $nt\Delta_*^{t-1}$ (n for guessing v, t for guessing where v occurs in the middle of the sequence, Δ_*^{t-1} for the other vertices).

Assume that the degree sequence is $(3, 3, 2, 2, \ldots)$, and u, v be the vertices of degree 3. Take a sequence of $t + 2$ vertices representing an Eulerian path from u to v (either

$(u, \ldots, u, \ldots, v, \ldots, v)$ or $(u, \ldots, v, \ldots, u, \ldots, v)$). The number of such sequences is bounded by $nt^2\Delta_*^{t-1}$ (n for guessing u, t^2 for guessing positions of u and v in the middle of the sequence, Δ_*^{t-1} for the other vertices including v). $\square$

Therefore, the probability that there exists a subgraph of G with t vertices and $t+1$ edges for $4 \le t \le 1/\eta = 3l$ is

$$\sum_{t=4}^{3l} 2nt^2\Delta_*^{t-1}p^{t+1} = \sum_{t=4}^{3l} \frac{2nt^2\Delta^{t+1}}{\Delta_*^2} \le \frac{n}{\Delta_*^2}(9l)^2\Delta^{3l+1} \le n^{-0.1}$$

for $l = O(\log n / \log \Delta)$, since $\frac{n}{\Delta_*^2} = O(n^{-\frac{1}{3}})$.

For $t > 1/\eta = 3l$, we count the number of subgraphs of G_* with t vertices and $(1+\eta)t$ edges. It is upper bounded by (the number of connected subtrees on t vertices) * (the number of possibilities to choose other $\eta t + 1$ edges out of $\binom{t}{2}$ pairs). The number of unlabeled rooted trees on t vertices is $C\alpha^t$ for some constants C and α [Ott48], so the number of connected subtrees on t vertices is bounded by $Cn\alpha^t\Delta_*^{t-1}$. Therefore, the total number of such subgraphs is

$$Cn\alpha^t\Delta_*^{t-1}\binom{\frac{t(t+1)}{2}}{\eta t+1} \le Cn\alpha^t\Delta_*^{t-1}\binom{t^2}{2\eta t} \le Cn\alpha^t\Delta_*^{t-1}(\frac{et}{2\eta})^{2\eta t}.$$

The probability that such a graph exists in G is at most

$$Cn\alpha^t\Delta_*^{t-1}(\frac{et}{2\eta})^{2\eta t}(\frac{\Delta}{\Delta_*})^{(1+\eta)t} \le \frac{n}{\Delta_*}(C_1\Delta^2)^t(C_2\frac{l^2t^2}{\Delta_*})^{t/3l}.$$

Let $A = C_1\Delta^2$ and $B = C_2\frac{l^2t^2}{\Delta_*}$. The above quantity is at most

$$\frac{n}{\Delta_*}A^tB^{t/3l} = (\frac{n}{\Delta_*}A^{3l}B)(AB^{1/3l})^{t-3l}.$$

Assume $t \le n^{\overline{\Gamma}}$ for some $\overline{\Gamma} \in (0, 0.1)$ and $l = O(\frac{\log n}{\log \Delta})$ be such that $\frac{n}{\Delta_*}A^{3l}B = \frac{C_2l^2t^2n(C_1\Delta^2)^{3l}}{\Delta_*^2} \le n^{-0.1}$, which also implies $AB^{1/3l} \le 1$. Summing over $t = 3l, \ldots, n^{\overline{\Gamma}}$, the probability that such a graph exists is bounded by $o(1)$.

Putting Them Together. In Section 7.5.1, we mentioned that the resulting graph should be amenable to have a Sherali-Adams solution with a large value, and introduced the notion of path-decomposability to measure it. The following lemma of Arora et al. [ABLT06] shows that our construction satisfies that every subgraph of G induced on at most $t \le n^{\overline{\Gamma}}$ vertices is l-path decomposable.

Lemma 7.5.5 ([ABLT06]). *Let $l \geq 1$ be an integer and $0 < \eta < \frac{1}{3l-1}$, and let H be a 2-connected graph with t vertices and at most $(1+\eta)t$ edges. Then H contains a path of length at least $l+1$ whose internal vertices have degree 2 in H.*

Finally, δ and l are fixed based on the other parameters to satisfy the requirements of the theorem.

Lemma 7.5.6. *There exists $\delta > 0$ and l (depending on T, ϵ, Δ, μ, $\overline{\Gamma}$) such that for any $k \leq n^{\delta}$, the following holds.*

1. $(1-\mu)^{\frac{l}{10}} \leq \frac{\mu}{5k}$.
2. Every induced subgraph of G on $(2\Delta)^l k$ vertices is l-path decomposable.

Proof. The first condition is implied by $l \geq C\delta \log n$ for some constant C depending on μ. The second condition is implied by $(2\Delta)^l k \leq n^{\overline{\Gamma}} \Leftrightarrow l \leq C'(\overline{\Gamma} - \delta)\log n$ for another constant C' depending on Δ. When we control girth and local sparsity, l is required to be $O(\frac{\log n}{\log \Delta})$. Therefore, by taking δ a small enough constant depending on T, ϵ, Δ, μ, and $\overline{\Gamma}$ (all of which depend on T, ϵ), we can ensure that such l exists. □

Therefore, there exist constants Δ, δ and $l = \Theta(\log n)$ (all constants depending on T, ϵ, μ) that satisfy all the requirements given in the theorem. □

7.5.2 Constructing (Inconsistent) Local Distributions

Let $D = (V, A)$, l_A, and $G = (V, E)$ be the instance of GENERALIZED MAX DICUT(T) and its underlying undirected graph constructed as above. In this subsection, given a set of $k \leq n^{\delta}$ vertices $S = \{v_1, \dots, v_k\}$ we give a distribution on events

$$\{l_V(v_1) = x_1, \dots, l_V(v_k) = x_k\}_{x_1,\dots,x_k \in [T]^0}.$$

The local distributions we construct in this subsection are not consistent; for different sets S and S', the marginal distribution on $S \cap S'$ from the distribution on S can be different from the same marginal from the distribution on S' (albeit they are close). This problem is fixed in the next subsection.

Let $d(u, v)$ be the shortest distance between u and v in G and $V' \subseteq V$ be the set of vertices whose shortest distance to S is at most l. Let G' and D' be the subgraph of G and D induced on V', respectively. Since $|V'| \leq (2\Delta)^l k$, G' is l-path decomposable by Theorem 7.5.2. Note that if $d(u, v) < l$, $d(u, v)$ is also the shortest distance between u and

v in G'. By the definition, a l-path decomposable graph does not have a cycle of length l, so if $d(u, v) < \frac{l}{2}$, the shortest path between u and v must be unique.

We begin by establishing a fact that when G' is path-decomposable (intuitively looks similar to a tree), there is a distribution on the partitions of V (i.e. multicuts) such that close vertices are unlikely to be separated but far vertices are likely to be separated. If G' is a tree, it is obtained by deleting each edge independently with probability μ. The noise parameter μ will be fixed later depending only on T and ϵ, so is asymptotically greater than $\frac{1}{l} = O(\frac{1}{\log n})$.

Theorem 7.5.7 ([CMM10]). *Suppose $G' = (V, E)$ is an l-path decomposable graph. Let $L = \lfloor l/9 \rfloor; \mu \in [1/L, 1]$. Then there exists a probabilistic distribution of multicuts of G' (or in other words random partition of G' in pieces) such that the following properties hold. For every two vertices u and v,*

1. *If $d(u, v) \leq L$, then the probability that u and v are separated by the multicut (i.e. lie in different parts) equals $1 - (1 - \mu)^{d(u,v)}$; moreover, if u and v lie in the same part, then the unique shortest path between u and v also lies in that part.*
2. *If $d(u, v) > L$, then the probability that u and v are separated by the multicut is at least $1 - (1 - \mu)^L$.*
3. *Every piece of the multicut partition is a tree.*

Based on this random partitioning, we define the distribution on the vertices in S (actually in V'). For each piece which is a tree, pick an arbitrary vertex v in the tree, choose $l_V(v)$ uniformly at random, and propagate this label to *weakly satisfy* every edge in the tree — an undirected edge $(u', v') \in E$ (swap u' and v' if necessary to assume $(u', v') \in A$) is weakly satisfied when $l_V(v') - l_V(u') = l_A(u', v')$ over $\mathbb{Z}_{T+1}$. Note that this definition is necessary for the original definition of satisfaction, but not sufficient.

It is clear that the choice of root in each tree does not matter, and the marginal distribution of each $l_V(v)$ is uniform on $[T]^0$. For vertices u and v with $d(u, v) \leq L$, we say that label i for u and i' for v *match* if $l_V(u) = i, l_V(v) = i'$ can be extended to weakly satisfy every edge on the unique shortest path between u and v (there are $T+1$ such pairs). If u and v are close, $l_V(u)$ and $l_V(v)$ will be correlated in a sense that if i and i' match, $l_V(u) = i$ almost implies $l_V(v) = i'$, while it is not the case when u and v are far apart. The following corollary formalizes this intuition.

Corollary 7.5.8. *Suppose $G' = (V', E')$ is an l-path decomposable graph. Let $L = \lfloor l/9 \rfloor$; $\mu \in [1/L, 1]$. Then there exists a random mapping $r : V' \to [T]^0$ such that*

1. *If* $d := d(u, v) \leq L$ *then*

$$\Pr[r(u) = i, r(v) = i'] = \begin{cases} \frac{(1-\mu)^d}{(T+1)} + \frac{1-(1-\mu)^d}{(T+1)^2} & \textit{if } i \textit{ and } i' \textit{ match} \\ \frac{1-(1-\mu)^d}{(T+1)^2} & \textit{otherwise} \end{cases}$$

2. *If* $d > L$ *then* $\frac{1-(1-\mu)^L}{(T+1)^2} \leq \Pr[r(u) = i, r(v) = i'] \leq \frac{1-(1-\mu)^L}{(T+1)^2} + \frac{(1-\mu)^L}{T+1}$ *for any* $i, i' \in [T]^0$.

Proof. r is defined by the following process: sample a distribution of multicuts as Theorem 7.5.7. Each piece is a tree, so we can pick an arbitrary vertex w and give a value $l_V(w)$ uniformly from $[T]^0$ and propagate along the tree to weakly satisfy every edge. Note that the distribution does not depend on the choice of the initial vertex.

Suppose $d(u, v) \leq L$, which ensures that if u and v are in the same piece, the only path connecting u and v in the piece is the shortest path in G. If i and i' are match labels,

$$\Pr[r(u) = i, r(v) = i'] = \Pr[u, v \text{ in the same piece}] \cdot \frac{1}{T+1} + \Pr[u, v \text{ separated}] \cdot \frac{1}{(T+1)^2}.$$

If i and i' are nonmatching labels,

$$\Pr[r(u) = i, r(v) = i'] = \Pr[u, v \text{ in the same piece}] \cdot 0 + \Pr[u, v \text{ separated}] \cdot \frac{1}{(T+1)^2}.$$

If $d(u, v) > L$, $\Pr[r(u) = i, r(v) = i']$ is lower bounded by $\frac{\Pr[u \text{ and } v \text{ are separated}]}{(T+1)^2}$, and upper bounded by $\frac{\Pr[u \text{ and } v \text{ are separated}]}{(T+1)^2} + \frac{\Pr[u \text{ and } v \text{ are not separated}]}{T+1}$. The separation guarantee in Theorem 7.5.7 proves the lemma. $\square$

Definition 7.5.9. *For any vertices* $u \neq v$ *and* $i, i' \in [T]^0$, *let* $\rho(u(i), v(i')) := \Pr[r(u) = i, r(v) = i']$ *if* $d(u, v) \leq L$, *or* $\frac{1}{(T+1)^2}$ *otherwise.* $\rho(v(i), v(i)) := \frac{1}{T+1}$ *and* $\rho(v(i), v(i')) := 0$ *for* $i \neq i'$. *Since the shortest path between* u *and* v *is unique when* $d(u, v) \leq L$, ρ *is uniquely defined given* G, D, l_A *and does not depend on* S, V', G', D' *which induce a local distribution.*

Definition 7.5.10. *Fix a set of* k *vertices* $S = \{v_1, \ldots, v_k\}$. *For any vertex* $u, v \in S$ *and* $i, j \in [T]^0$, *let* $\nu_S(u(i), v(i')) := \Pr[x(u) = i, x(v) = i']$ *in the local distribution on* S *defined by* r *in Corollary 7.5.8.*

7.5.3 Geometric Embedding and Rounding

In this subsection, we still fix a set of k vertices $S = \{v_1, \ldots, v_k\}$ and produce a distribution on the events $\{l_V(v_1) = x_1, \ldots, l_V(v_k) = x_k\}_{x_1,\ldots,x_k \in [T]^0}$. The difference from the last subsection is that the resulting distributions become consistent — the marginal distribution on $S \cap S'$ does not depend on the choice of its superset (S or S') that is used to obtain a larger local distribution.

Embedding. Consider ρ and ν_S defined in the last subsection. ρ and ν_S both capture the pairwise distribution between the events $\{l_V(v) = x\}_{v \in S, x \in [T]^0}$, but each of them has its own defects. ν_S depends on the choice of S, so does not yield consistent local distributions. ρ does not depend on S, but for far vertices, Corollary 7.5.8 does not guarantee any local distribution consistent with it. However, they are close in a sense — they are identical when $d(u, v) \leq L$ and differ by at most $\frac{(1-\mu)^L}{T+1}$ otherwise.

The main idea of Charikar et al. [CMM09] is to interpret ρ and ν_S as pairwise distances between events and embed ρ to l_2 with small error. It is based on the fact that ρ and ν_S are close for any S and ν_S is readily embeddable to l_2. Since the embedding into l_2 is uniquely defined by the pairwise distances and ρ does not depend on the choice of S, geometric rounding schemes based on the embedding yield consistent local distributions. Let $v(i)$ be the vector corresponding to the event $l_V(v) = i$. Our goal is to construct $k(T+1)$ vectors $\{v(i)\}_{v \in S, i \in [T]^0}$ such that $u(i) \cdot v(i') \approx \rho(u(i), v(i'))$. Following the above intuition, the following lemma says that this embedding is possible with error depending on μ.

Lemma 7.5.11. *There exist $k(T+1)$ vectors $\{v(i)\}_{v \in S, i \in [T]^0}$ such that $\|v(i)\|_2^2 = \mu + \frac{1}{T+1}$ and $u(i) \cdot v(i') = \frac{\mu}{2} + \rho(u(i), v(i'))$.*

Proof. For each $u(i)$, we construct two vectors $u(i)_1$ and $u(i)_2$ and finally merge them by $u(i) := u(i)_1 \oplus u(i)_2$. $u(i)_2$ is the indicator random variable for the event $l_V(u) = i$, where the distribution follows ν_S. Since ν_S is based on an actual distribution on the events, the vectors $\{v(i)_2\}_{v \in V, i \in [T]^0}$ are embeddable into l_2 with $\|v(i)_2\|_2^2 = \Pr[l_V(v) = i] = \frac{1}{T+1}$ and $u(i)_2 \cdot v(i')_2 = \nu_S(u(i), v(i'))$. The first group of vectors $\{v(i)_1\}_{v \in V, i \in [T]^0}$ *convert* these inner products from ν_S to ρ with small error.

The following lemma says that a metric space can be isometrically embeddable into l_2 if all pairwise distances are similar.

Lemma 7.5.12 ([CMM10]). *Consider a metric space (Y, α) on t points. If for every two distinct points u and v: $|\alpha(u, v) - \beta| \leq \frac{\beta}{2t}$ for some $\beta > 0$, then (Y, α) is isometrically embeddable into l_2.*

We add a vector O (so that we have $k(T+1)+1$ vectors) and set the following distance requirements.

1. $\|v(i)_1 - O\|_2 = \sqrt{\mu}$ for all u_i.
2. $\|u(i)_1 - v(i')_1\|_2 = \sqrt{\mu - 2\rho(u(i), v(i')) + 2\nu_S(u(i), v(i'))}$ for all $u(i), v(i')$.

Note that $|\rho(u(i), v(i')) - \nu_S(u(i), v(i'))| \leq \frac{(1-\mu)^L}{T+1} \leq \frac{\mu}{5(T+1)k}$, where the last inequality follows from Theorem 7.5.2. This implies

$$||\|u(i)_1 - v(i')_1\|_2 - \sqrt{\mu}| \leq \sqrt{\mu}(1 - \sqrt{1 - \frac{1}{2.5(T+1)k}}) \leq \sqrt{\mu} \cdot \frac{1}{2((T+1)k+1)}.$$

By Lemma 7.5.12, there are vectors $\{u(i)_1, v(i)_1\}_i$ and O that meet the above distance requirements. Without loss of generality, assume that O is the origin. Defining $u(i) := u(i)_1 \oplus u(i)_2$ satisfies

1. $\|u(i)\|_2^2 = \mu + \frac{1}{T+1}$.
2. $u(i) \cdot v(i') = u(i)_1 \cdot v(i')_1 + u(i)_2 \cdot v(i')_2 = \frac{2\mu - \|u(i)_1 - v(i')_1\|_2^2}{2} + \nu_S(u(i), v(i')) = \frac{\mu}{2} + \rho(u(i), v(i'))$.

□

Rounding and Analyzing adjacent vertices. Given $k(T+1)$ vectors $\{v(i)\}_{v \in S, i \in [T]^0}$, our rounding scheme is one of the most natural ways to choose one out of $(T+1)$ vectors — take a random Gaussian vector g and for each vertex v, set $l_V(v) = i$ such that $v(i) \cdot g$ is the maximum over all i. Since the inner products of these vectors depend only on ρ (which does not depend on the choice of S), it gives a consistent local distribution.

Fix adjacent vertices v and u (without loss of generality assume $(u, v) \in A$). It only remains to show that $\Pr[l_V(u) = 0, l_V(v) = l_A(u, v)] \approx \frac{1}{T+1}$. For any pair of adjacent vertices, we can write $2(T+1)$ vectors explicitly. They are just two sets of $T+1$ orthonormal vectors, very closely correlated — there are $T+1$ pairs $(u(i), v(i'))$, $i' - i = l_A(u, v)$ in $\mathbb{Z}_{T+1}$, such that $u(i) \approx v(i')$. With this symmetric structure and a suitable choice of the noise parameter μ, we can analyze the performance of our rounding.

Lemma 7.5.13. *There exists μ depending on T and ϵ such that, in the above rounding scheme, the probability that $l_V(u) = 0$ and $l_V(v) = l_A(u, v)$ is at most $\frac{1-12\epsilon}{T+1}$.*

Proof. For notational simplicity, assume $l_A(u,v) = 0$ — which is not allowed in actual instances. Then $u(i)$ and $v(i)$ become *matching* vectors — $\rho(u(i), v(i)) = \frac{1-\mu}{T+1} + \frac{\mu}{(T+1)^2}$ and $\rho(u(i), v(j)) = \frac{\mu}{(T+1)^2}$ for $i \neq j$. The following is the list of all possible inner products between $2(T+1)$ vectors.

1. $\|u(i)\|_2^2 = \mu + \frac{1}{T+1}$.
2. $u(i) \cdot u(j) = \frac{\mu}{2}$ for $i \neq j$.
3. $u(i) \cdot v(i) = \frac{\mu}{2} + \frac{1-\mu}{T+1} + \frac{\mu}{(T+1)^2}$.
4. $u(i) \cdot v(j) = \frac{\mu}{2} + \frac{\mu}{(T+1)^2}$ for $i \neq j$.

Even though we used Lemma 7.5.12 as a black-box to obtain the current embedding, we can explicitly represent $u(i), v(i)$'s in the Euclidean space. They can be represented as a linear combination of $(T+1) + (T+1)^2 + 2(T+1) + 1$ orthogonal vectors (with different lengths), which can be classified into the following four categories:

- $a(i)$ for $i \in [T]^0$: Length $\sqrt{\frac{1-\mu}{T+1}}$. Denotes the event that (u,v) is not deleted and $l_V(u) = l_V(v) = i$.
- $b(i,j)$ for $i, j \in [T]^0$: Length $\sqrt{\frac{\mu}{(T+1)^2}}$. Denotes the event that (u,v) is deleted and $l_V(u) = i$, $l_V(v) = j$.
- $c(i), c'(i)$ for $i \in [T]^0$: Length $\sqrt{\frac{\mu}{2}}$. One of them is assigned for each of $2(T+1)$ vectors.
- d: Length $\sqrt{\frac{\mu}{2}}$. Common for all vectors.

Let

$$u(i) := a(i) + \sum_j b(i,j) + c(i) + d$$
$$v(i) := a(i) + \sum_j b(j,i) + c'(i) + d.$$

It is straightforward to check that the following representation of $u(i)$ and $v(i)$ satisfy all the inner product requirements.

For each vector $u(i)$, we denote the random variable equal to the inner product of $u(i)$ and g by $U(i)$. Similarly, define $V(i), A(i), B(i,j), C(i), C'(i), D(i)$ for $v(i)$, $a(i)$, $b(i,j)$, $c(i)$, $c'(i)$, $d(i)$ respectively. Each random variable follows the Gaussian distribution with mean 0 and standard deviation same with the length of the corresponding vector. Furthermore, the inner products of two vectors is the same with the covariance of corresponding random variables. The following lemma shows that our consistent local distributions actually satisfy each edge with probability close to $\frac{1}{T+1}$, proving Theorem 7.1.4.

Lemma 7.5.14. *Fix* $i \in [T]^0$ *and* $0 < \epsilon < 1/24$. *If* $\mu \leq \frac{\epsilon^2}{256(T+1)\log^2(\frac{T+1}{\epsilon})}$,

$$\Pr[l_V(u) = i, l_V(v) = i] \geq \frac{1-12\epsilon}{T+1}.$$

Proof. We compute the probability that u and v are assigned the same label i. Let

$$M := \max\left[\max_j[\sum_k B(j,k)], \max_j[\sum_k B(k,j)], \max_j[C(j)], \max_j[C'(j)]\right].$$

Then,

$$\begin{aligned}
& \Pr[l_V(u) = i, l_V(v) = i] \\
\geq\ & \Pr[A(i) = \max_j[A(j)]] \cdot \Pr\left[M \leq \frac{A(i) - \max_{j \neq i}[A(j)]}{4} \middle| A(i) = \max_j[A(j)]\right] \\
\geq\ & \frac{1}{T+1} \Pr\left[M \leq \frac{\max_j[A(j)] - \max 2_j[A(j)]}{4}\right]
\end{aligned}$$

We argue that the above quantity is close to $\frac{1}{T+1}$ by showing that each of 4 quantities

$$\max_j[\sum_k B(j,k)], \max_j[\sum_k B(k,j)], \max_j[C(j)], \max_j[C'(j)]$$

is greater than $\frac{\max_j[A(j)] - \max 2_j[A(j)]}{4}$ with small probability. Note that $\sum_k B(j,k)$ follows the Gaussian distribution with mean 0 and variance $\frac{\mu}{T+1}$, which is much less than that of $C(j)$. Since $C(j)$ and $C'(j)$ follow the same distribution, it is enough to show that $\max_j[C(j)] > \frac{\max_j[A(j)] - \max 2_j[A(j)]}{4}$ with small probability. The following claim proves the lemma. $\square$

Claim 7.5.15. *Let* $0 < \epsilon < 1/4$. *If* $\mu \leq \frac{\epsilon^2}{256(T+1)\log^2(\frac{T+1}{\epsilon})}$,

$$\Pr[\max_j[C(j)] > \frac{\max_j[A(j)] - \max 2_j[A(j)]}{4}] < 3\epsilon.$$

Proof. The above probability can be rewritten as

$$\Pr[\sqrt{\frac{\mu}{2}}\max_j[g_j] > \sqrt{\frac{1-\mu}{T+1}}\frac{\max_j[g'_j] - \max2_j[g'_j]}{4}]$$

where $g_0, \ldots, g_T, g'_0, \ldots, g'_T$ are independent standard Gaussian random variables.

Let $x \geq \sqrt{\mu \log \frac{T+1}{\epsilon}}$. By Lemma 7.5.16,

$$\Pr[\sqrt{\frac{\mu}{2}}\max_j[g_j] > x] < \epsilon.$$

Let $x \leq \frac{\epsilon}{8\sqrt{\log \frac{T+1}{\epsilon}}}\sqrt{\frac{1-\mu}{T+1}}$. By Lemma 7.5.17,

$$\Pr[\sqrt{\frac{1-\mu}{T+1}}\frac{\max_j[g'_j] - \max2_j[g'_j]}{4} < x] < 2\epsilon.$$

The fact that $\mu \leq \frac{\epsilon^2}{256(T+1)\log^2(\frac{T+1}{\epsilon})}$ ensures that there is x that satisfies the both Lemma 7.5.16 and 7.5.17. Taking union bound proves the lemma. $\square$

It remains to prove the following two lemmas about Gaussians. We prove them in Section 7.7 using some basic properties of Gaussians.

Lemma 7.5.16. *Let $g_1, \ldots, g_n$ ($n \geq 2$) be independent standard Gaussian random variables and $0 < \epsilon < 1$. If $x \geq \sqrt{2\log \frac{n}{\epsilon}}$,*

$$\Pr[\max_j[g_j] \leq x] \geq 1 - \epsilon.$$

Lemma 7.5.17. *Let $g_1, \ldots, g_n$ ($n \geq 2$) be independent standard Gaussian random variables and $0 < \epsilon < 1/4$. If $x \leq \frac{\epsilon}{2\sqrt{\log \frac{n}{\epsilon}}}$,*

$$\Pr[\max_j[g_j] - \max_j 2[g_j] \geq x] \geq (1 - 2\epsilon).$$

$\square$

This finishes the construction of a solution to the n^δ-rounds of the Sherali-Adams hierarchy with value $\frac{1-12\epsilon}{T+1}$. Since $\mathsf{Opt}(V, l_A) \leq \frac{1+\epsilon}{4T}$ by Theorem 7.5.2, it proves Theorem 7.1.4 and Theorem 7.1.2.

7.6 $(\frac{1}{4}+\Omega(\frac{1}{T}))$-Approximation Algorithm for GENERALIZED MAX DICUT

In this section, we propose an approximation algorithm for GENERALIZED MAX DICUT(T) that guarantees $(\frac{1}{4}+\frac{1}{16T})$ fraction of the optimal solution, proving Theorem 7.4.1. It is based on the 2-rounds of the Sherali-Adams hierarchy (also known as the standard LP), defined as the following:

$$\begin{array}{lll} \text{maximize} & \sum_{(u,v)\in A} x_{(u,v)}(0, l_A(u,v)) & \\ \text{subject to} & \sum_{\alpha\in([T]^0)^S} x_S(\alpha) = 1 & \text{for all } S \subseteq V, |S| \le 2 \\ & \sum_{j\in[T]^0} x_{(u,v)}(i,j) = x_u(i) & \text{for all } u \neq v, i \in [T]^0 \end{array}$$

The algorithm is almost identical to the simple $\frac{1}{4}$-approximation algorithm. For each vertex v, independently set $l_V(v)=0$ with probability $\frac{1+x_v(0)}{2}$, and $l_V(v)=i$ ($i\neq 0$) with probability $\frac{x_v(i)}{2}$. Equivalently, we assign each vertex 0 with probability half and follow its marginal x_v with probability half.

For each edge $(u,v)\in A$, let $c = c(u,v) := x_{(u,v)}(0, l_A(u,v))$ so that the value the solution $\{x_S(\alpha)\}$ to the LP is $\mathbb{E}_{(u,v)}[c(u,v)] \ge \mathsf{Opt}$. The probability that (u,v) is satisfied is

$$(\frac{1+x_u(0)}{2})(\frac{x_v(l_A(u,v))}{2}) \ge \frac{c}{4}+\frac{c^2}{4}$$

since $x_u(0), x_v(l_A(u,v)) \ge c$. Therefore, the expected fraction of satisfied edges is at least

$$\mathbb{E}_{(u,v)\in A}[\frac{c(u,v)}{4}+\frac{c(u,v)^2}{4}] \ge \frac{\mathsf{Opt}}{4}+\frac{\mathsf{Opt}^2}{4} \ge \frac{\mathsf{Opt}}{4}+\frac{\mathsf{Opt}}{16T}$$

since $\mathsf{Opt} \ge \frac{1}{4T}$ (focusing on the label with the most edges and finding the maximum dicut with respect to the edges with this label guarantees to satisfy $\frac{1}{4T}$ fraction of edges).

7.7 Proofs of Lemmas about Gaussians

Let $\phi(x)$ and $\Phi(x)$ be the probability density function (PDF) and the cumulative distribution function (CDF) of the standard Gaussian, respectively. Let $\tilde{\Phi}(x) = 1-\Phi(x)$. We begin with the following simple fact about the tail of Φ.

Lemma 7.7.1 ([CMM06]). *For any $t>0$, $\frac{t}{\sqrt{2\pi}(t^2+1)}e^{-\frac{t^2}{2}} < \tilde{\Phi}(t) < \frac{1}{\sqrt{2\pi}t}e^{-\frac{t^2}{2}}$.*

Lemma 7.7.2 (Restatement of Lemma 7.5.16)**.** *Let $g_1, \ldots, g_n$ ($n \geq 2$) be independent standard Gaussian random variables and $0 < \epsilon < 1$. If $x \geq \sqrt{2\log\frac{n}{\epsilon}}$,*

$$\Pr[\max_j[g_j] \leq x] \geq 1 - \epsilon.$$

Proof. Note that $x \geq \sqrt{2\log 2}$, so $\frac{1}{\sqrt{2\pi}x} \leq 1$.

$$\begin{aligned} & x \geq \sqrt{2\log\frac{n}{\epsilon}} \\ \Rightarrow\quad & \frac{1}{\sqrt{2\pi}x}\exp(-\frac{x^2}{2}) \leq \frac{\epsilon}{n} \\ \Rightarrow\quad & 1 - \Phi(x) \leq \frac{\epsilon}{n}, \end{aligned}$$

where the last inequality follows from Lemma 7.7.1. We can conclude that

$$\Pr[\max_j[C(j)] \leq x] = \Phi(x)^n \geq (1 - \frac{\epsilon}{n})^n \geq 1 - \epsilon.$$

$\square$

Lemma 7.7.3 (Restatement of Lemma 7.5.17)**.** *Let $g_1, \ldots, g_n$ ($n \geq 2$) be independent standard Gaussian random variables and $0 < \epsilon < 1/4$. If $x \leq \frac{\epsilon}{2\sqrt{\log\frac{n}{\epsilon}}}$,*

$$\Pr[\max_j[g_j] - \max_j 2[g_j] \geq x] \geq (1 - 2\epsilon).$$

Proof.

$$\begin{aligned} \Pr[\max_j[g_j] - \max_j 2[g_j] \geq x] &\geq n\int_{-\infty}^{\infty}\Phi[y-x]^{n-1}\phi(y)dy \\ &\geq n\int_{-\infty}^{b}\Phi[y-x]^{n-1}\phi(y)dy \quad \text{for some } b \text{ fixed later} \\ &= n\int_{-\infty}^{b}\Phi[y-x]^{n-1}\phi(y-x)\frac{\phi(y)}{\phi(y-x)}dy \\ &\geq (\inf_{y\in[-\infty,b]}\frac{\phi(y)}{\phi(y-x)})\int_{-\infty}^{b}n\Phi[y-x]^{n-1}\phi(y-x)dy \\ &= (\inf_{y\in[-\infty,b]}\frac{\phi(y)}{\phi(y-x)})\int_{-\infty}^{b}(\Phi[y-x]^n)'dy \\ &= (\inf_{y\in[-\infty,b]}\frac{\phi(y)}{\phi(y-x)})\Phi[b-x]^n \end{aligned}$$

Let $b = x + \sqrt{2 \log \frac{n}{\epsilon}}$. By the same argument with Lemma 7.5.16, we have

$$\begin{aligned} & 1 - \Phi[b - x] \leq \frac{\epsilon}{n} \\ \Rightarrow \quad & \Phi[b - x] \geq 1 - \frac{\epsilon}{n} \\ \Rightarrow \quad & \Phi[b - x] \geq (1 - \epsilon)^{1/n} \end{aligned}$$

Now we bound

$$\begin{aligned} \inf_{y \in [-\infty, b]} \frac{\phi(y)}{\phi(y - x)} &= \inf_{y \in [-\infty, b]} \exp(-\frac{y^2}{2} + \frac{(y - x)^2}{2}) = \inf_{y \in [-\infty, b]} \exp(\frac{-2xy + x^2}{2}) \\ &= \exp(\frac{-2bx + x^2}{2}) \end{aligned}$$

where the last inequality holds since it is monotonically decreasing in y. $x \leq \frac{\epsilon}{2\sqrt{\log \frac{n}{\epsilon}}}$ implies

$$\begin{aligned} & x(x + \sqrt{2 \log \frac{n}{\epsilon}}) \leq \epsilon \\ \Rightarrow \quad & bx \leq \epsilon \\ \Rightarrow \quad & \frac{-2bx + x^2}{2} \geq -\epsilon \\ \Rightarrow \quad & \exp(\frac{-2bx + x^2}{2}) \geq \exp(-\epsilon) \geq 1 - \epsilon \end{aligned}$$

Since both $\inf_{y \in [-\infty, b]} \frac{\phi(y)}{\phi(y-x)}$ and $\Phi[b - x]^n$ are at least $1 - \epsilon$, the lemma follows. $\square$

Lemma 7.7.4 (Restatement of Lemma 7.4.6)**.** *Fix* $\rho, \alpha \in (0, 1)$. *The function* $f(x) := \overline{\Gamma}_\rho(\alpha, x)$ *is concave.*

Proof. Let Y, Z be independent Gaussians and $X := \rho Y + \sqrt{1 - \rho^2} Z$. Fix $0 \leq a \leq b$. We will show that $f(a) + f(b) \geq f(a + b)$. Let $x = \tilde{\Phi}^{-1}(a + b), y = \tilde{\Phi}^{-1}(b), z = \tilde{\Phi}^{-1}(a), w =$

$\tilde{\Phi}^{-1}(\alpha)$. Note that $x \le y \le z$.

$$\begin{aligned}
&f(a)+f(b)-f(a+b)\\
=&\Pr[Y \ge y \text{ and } X \ge w]+\Pr[Y \ge z \text{ and } X \ge w]-\Pr[Y \ge x \text{ and } X \ge w]\\
=&\Pr[Y \ge z \text{ and } X \ge w]-\Pr[x \le Y \le y \text{ and } X \ge w]\\
\ge&\Pr[Y \ge z \text{ and } Z \ge \frac{w-\rho z}{\sqrt{1-\rho^2}}]-\Pr[x \le Y \le y \text{ and } Z \ge \frac{w-\rho y}{\sqrt{1-\rho^2}}]\\
=&a(\Pr[Z \ge \frac{w-\rho z}{\sqrt{1-\rho^2}}]-\Pr[Z \ge \frac{w-\rho y}{\sqrt{1-\rho^2}}])\\
\ge&0
\end{aligned}$$

□

Lemma 7.7.5 (Restatement of Lemma 7.4.7)**.** *For large enough T and $\delta = \frac{1}{T^{1/4}}$, the following holds. For any $a \in [0,1], b \in [0, \frac{1}{T}]$ and $\rho \in (0, \sqrt{\frac{2}{T\delta}})$, $\overline{\Gamma}_\rho(a,b) \le ab + \frac{2}{T^{5/4}}$.*

Proof. Let Y, Z be independent Gaussians and $X := \rho Y + \sqrt{1-\rho^2}Z$. Let $x = \tilde{\Phi}^{-1}(a)$ and $y = \tilde{\Phi}^{-1}(b)$. By taking $T > 2$, we can assume $b < \frac{1}{2}$ and $y > 0$, while we do not put any assumption on a and x.

$$\begin{aligned}
\overline{\Gamma}_\rho(a,b) &= \Pr[X \ge x \text{ and } Y \ge y]\\
&\le \Pr[Z \ge \frac{x-2\rho y}{\sqrt{1-\rho^2}} \text{ and } y \le Y \le 2y] + \Pr[Y \ge 2y]\\
&\le \Pr[Z \ge \frac{x-2\rho y}{\sqrt{1-\rho^2}} \text{ and } Y \ge y] + \Pr[Y \ge 2y]\\
&\le b \cdot \tilde{\Phi}(\frac{x-2\rho y}{\sqrt{1-\rho^2}}) + \tilde{\Phi}(2y). \qquad (7.1)
\end{aligned}$$

By Lemma 7.7.1, $\tilde{\Phi}(2y) < \frac{1}{2\sqrt{2\pi}y}\exp(-2y^2) < b^3 < \frac{1}{T^{5/4}}$.

- $a \ge 1 - \frac{1}{T^{1/4}}$: (7.1) is bounded by $b + \frac{1}{T^{5/4}} \le (a + \frac{1}{T^{1/4}})b + \frac{1}{T^{5/4}} \le ab + \frac{2}{T^{5/4}}$.
- $b \le \frac{1}{T^{5/4}}$: (7.1) is bounded by $b + \frac{1}{T^{5/4}} \le \frac{2}{T^{5/4}}$.

- $a \le 1 - \frac{1}{T^{1/4}}$ and $b \ge \frac{1}{T^{5/4}}$: Note that $x \ge -10\sqrt{\log T}$ and $y \le 10\sqrt{\log T}$. Since $\rho \le \sqrt{\frac{2}{T\delta}} = \frac{\sqrt{2}}{T^{3/8}}$,

$$(x-2\rho y)-\sqrt{1-\rho^2}(x-\frac{1}{T^{1/4}}) \ge \begin{cases} -2\rho y + \frac{1}{2T^{1/4}} \ge 0 & \text{if } x \ge 0 \\ \rho^2 x - 2\rho y + \frac{1}{2T^{1/4}} \ge 0 & \text{if } -10\sqrt{\log T} \le x \le 0, \end{cases}$$

which shows that $\frac{x-2\rho y}{\sqrt{1-\rho^2}} \ge x - \frac{1}{T^{1/4}}$. Therefore,

$$(7.1) \le b \cdot \tilde{\Phi}(x - \frac{1}{T^{1/4}}) + \frac{1}{T^{5/4}} \le b(a + \frac{1}{T^{1/4}}) + \frac{1}{T^{5/4}} \le ab + \frac{2}{T^{5/4}},$$

where the second inequality follows from $\phi(x) \le 1$ for all $x \in \mathbb{R}$.

□

Chapter 8

LDPC Decoding

8.1 Introduction

Low-density parity-check (LDPC) codes are a class of linear error correcting codes originally introduced by Gallager [Gal62] and that have been extensively studied in the last decades. A (d_v, d_c)-LDPC code of block length n is described by a parity-check matrix $H \in \mathbb{F}_2^{m \times n}$ (with $m \leq n$) having d_v ones in each column and d_c ones in each row. It can be also represented by its bipartite *parity-check graph* $(L \cup R, E)$ where L corresponds to the columns of H, R corresponds to the rows of H, and $(u, v) \in E$ if and only if $H_{v,u} = 1$. For a comprehensive treatment of LDPC codes, we refer the reader to the book of Richardson and Urbanke [RU08]. In many studies of LDPC codes, *random LDPC codes* have been considered. For instance, Gallager studied in his thesis the distance and decoding-error probability of an ensemble of random (d_v, d_c)-LDPC codes. Random (d_v, d_c)-LDPC codes were further studied in several works (e.g., [SS94, Mac99, RU01, MB01, DPT$^+$02, LS02, KRU12]). The reasons why random (d_v, d_c)-LDPC codes have been of significant interest are their nice properties, their tendency to simplify the analysis of the decoding algorithms and the potential lack of known explicit constructions for properties satisfied by random codes.

One such nice property that is exhibited by random (d_v, d_c)-LDPC codes is the *expansion* of the underlying parity-check graph. Sipser and Spielman [SS94] exploited this expansion in order to give a linear-time decoding algorithm correcting a constant fraction of errors (for $d_v, d_c = O(1)$). More precisely, they showed that if the underlying graph has the property that every subset of at most δn variable nodes expands by at least a factor of $3d_v/4$, then their decoding algorithm can correct an $\Omega(\delta)$ fraction of errors in linear-time. Since, with high probability, a random (d_v, d_c)-LDPC code satisfies this expansion property for some $\delta = \Omega(1/d_c)$, this implies that the linear-time decoding algorithm of Sipser-Spileman corrects an $\Omega(1/d_c)$ fraction of errors on a random (d_v, d_c)-LDPC code. A few years after the work of Sipser-Spielman, Feldman, Karger and Wainwright [FWK05, Fel03] introduced a decoding algorithm that is based on a simple linear programming (LP) relaxation, and a later paper by Feldman, Malkin, Servedio, Stein and Wainwright [FMS$^+$07] showed that when the underlying parity-check graph has the property that every subset of at most δn variable nodes expands by a factor of at least $2d_v/3 + \Omega(1)$, the linear program of Feldman-Karger-Wainwright corrects an $\Omega(\delta)$ fraction of errors. Again, since with high probability, a random (d_v, d_c)-LDPC code satisfies this expansion property for some $\delta = \Omega(1/d_c)$, this means that the LP of [FWK05] corrects an $\Omega(1/d_c)$ fraction of errors on a random (d_v, d_c)-LDPC code.

However, the fraction of errors that is corrected by the Sipser-Spielman algorithm and the LP relaxation of [FWK05] (which is $O(1/d_c)$) can be much smaller than the best possible: in fact, [Gal62] (as well as [MB01]) showed that for a random (d_v, d_c)-LDPC code, the exponential-time nearest-neighbor Maximum Likelihood (ML) algorithm corrects close to a $H_b^{-1}(d_v/d_c)$ fraction of probabilistic errors, which by Shannon's channel coding theorem is the best possible[1]. Note that, for example, if we set the ratio d_v/d_c to be a small constant and let d_c grow, then the fraction of errors that is corrected by the Sipser-Spielman algorithm and the LP relaxation of Feldman et al. decays to 0 with increasing d_c, whereas the maximum information-theoretically possible fraction is a fixed absolute constant![2] The belief propagation (BP) algorithm also suffers from the same limitation [BM02, KRU12]. In fact, there is no known polynomial-time algorithm that approaches the information-theoretic limit for random (d_v, d_c)-regular LDPC codes. [3]

In the areas of combinatorial optimization and approximation algorithms, hierarchies of linear and semidefinite programs such as the Sherali-Adams [SA90] and the Lasserre [Las01] hierarchies recently gained significant interest[4]. Given a base LP relaxation, such

[1]More precisely, the fraction of errors corrected by the ML decoder is bounded below $H_b^{-1}(d_v/d_c)$ for fixed d_c but gets arbitrarily close to $H_b^{-1}(d_v/d_c)$ as d_c gets larger.

[2]In fact, not only is the fraction of *probabilistic* errors that is corrected by the ML decoder an absolute constant, but so is the fraction of *adversarial* errors [Gal62, BM04]. More precisely, for say $d_v = 0.1d_c$, Theorem 11 of [BM04] implies that the *minimum distance* of a random (d_v, d_c)-regular LDPC code is at least an absolute constant and it approaches the Gilbert-Varshamov bound for rate $R = 1 - d_v/d_c = 0.9$ as d_c gets larger.

[3]We point out that for some ensembles of *irregular* LPDC codes [RSU01] as well as for the recently studied *spatially-coupled* codes [KRU12], belief propagation is known to have better properties. In this paper, our treatment is focused on random *regular* LDPC codes.

[4]We point out that the Lasserre hierarchy is also referred to as the "Sum of Squares" hierarchy in the literature.

hierarchies tighten it into sequences of convex programs where the convex program corresponding to the rth round in the sequence can be solved in time $n^{O(r)}$ and yields a solution that is "at least as good" as those obtained from previous rounds in the sequence. For an introduction and comparison of those LP and SDP hierarchies, we refer the reader to the work of Laurent [Lau03] where it is also shown that the Lasserre hierarchy is at least as strong as the Sherali-Adams hierarchy.

Inspired by the Sherali-Adams hierarchy, Arora, Daskalakis and Steurer [ADS12] improved the best known fraction of correctable probabilistic errors by the LP decoder (which was previously achieved by Daskalakis et al. [DDKW08]) for some range of values of d_v and d_c. Both Arora et al. [ADS12] and the original work of Feldman et al. [FWK05, Fel03] asked whether tightening the base LP using linear or semidefinite hierarchies can improve its performance, potentially approching the information-theoretic limit. More precisely, in all previous work on LP decoding of error-correcting codes, the base LP decoder of Feldman et al. succeeds in the decoding task if and only if the transmitted codeword is the unique optimum of the relaxed polytope with the objective function being the (normalized) l_1 distance between the received vector and a point in the polytope. On the other hand, the decoder is considered to fail whenever there is an optimal non-integral vector[5]. The hope is that adding linear and semidefinite constraints will help "prune" non-integral optima, thereby improving the fraction of probabilistic errors that can be corrected.

In this paper, we prove the first lower bounds on the performance of the Sherali-Adams and Lasserre hierarchies when applied to the problem of decoding random (d_v, d_c)-LDPC codes. Throughout this paper, by a random (d_v, d_c)-LDPC code, we mean one whose parity-check graph is drawn from the following ensemble that was studied in numerous previous works (e.g., [SS94, RU01, MB01, LS02, BM04, KRU12]) and is very close to the ensemble that was originally suggested by Gallager [Gal62]. Set $M := nd_v = md_c$ where n is the block length and m is the number of constraints. Assign d_v (resp. d_c) sockets to each of n (resp. m) vertices on the left (resp. right) and number them $1, \dots, M$ on each side. Sample a permutation $\pi : \{1, \dots, M\} \to \{1, \dots, M\}$ uniformly at random, and connect the i-th socket on the left to the $\pi(i)$-th socket on the right. Place an edge betwen variable i and constraint j if and only if there is an odd number of edges between the sockets corresponding to i and those corresponding to j. Our main results can be stated as follows:

Theorem 8.1.1 (Lower bounds in the Sherali-Adams hierarchy)**.** *For any d_v and $d_c \geq 5$, there exists $\eta > 0$ (depending on d_c) such that a random (d_v, d_c)-LDPC code satisfies the following with high probability: for any received vector, there is a fractional solution to the ηn rounds of the Sherali-Adams hierarchy of value $1/(d_c - 3)$ (for odd d_c) or $1/(d_c - 4)$*

[5] Such an optimal non-integral vector is called a "pseudocodeword" in the LP-decoding literature.

(for even d_c). Consequently, ηn rounds cannot decode more than a $\approx 1/d_c$ fraction of errors.

Theorem 8.1.2 (Lower bounds in the Lasserre hierarchy)**.** *For any d_v and $d_c = 3 \cdot 2^i + 3$ with $i \geq 1$, there exists $\eta > 0$ (depending on d_c) such that a random (d_v, d_c)-LDPC code satisfies the following with high probability: for any received vector, there is a fractional solution to the ηn rounds of the Lasserre hierarchy of value $3/(d_c - 3)$. Consequently, ηn rounds cannot decode more than a $\approx 3/d_c$ fraction of errors.*

We note that Theorems 8.1.1 and 8.1.2 hold, in particular, for random errors. We point out that as in all previous work on LP decoding of error-correcting codes, Theorems 8.1.1 and 8.1.2 assume that a decoder based on a particular convex relaxation succeeds in the decoding task if and only if the transmitted codeword is the unique optimum of the convex relaxation. Thus, if an ϵ fraction of errors occurs, then any fractional solution of value less than ϵ results in a decoding error. Note that the decoder based on the LP (resp. SDP) corresponding to n rounds of the Sherali-Adams (resp. Lasserre) hierarchy is the nearest-neighbor maximum likelihood (ML) decoder.

We note that our LP/SDP hierarchy $O(1/d_c)$ lower bounds for random LDPC codes hold, in particular, for any check-regular code with good check-to-variable expansion. Moreover, the fact that the base LP corrects $\Omega(1/d_c)$ errors follows from the (variable-to-check) expansion of random LDPC codes[6]. In that respect, it is intriguing that expansion constitutes both the *strength* and the *weakness* of random LDPC codes.

Some of our techniques are more generally applicable to a large class of Boolean Constraint Satisfaction Problems (CSPs) called MIN ONES where the goal is to satisfy each of a collection of constraints while minimizing the number of variables that are set to 1. In particular, we obtain improved integrality gaps in the Lasserre hierarchy for the k-HYPERGRAPH VERTEX COVER problem. The k-HYPERGRAPH VERTEX COVER is known to be NP-hard to approximate within a factor of $k - 1 - \epsilon$ [DGKR05]. This reduction would give the same integrality gap only for some *sublinear* number of rounds of the Lasserre hierarchy, whereas the best integrality gap for a linear number of rounds remains at $2 - \epsilon$ [Sch08]. We prove that an integrality gap of $k - 1 - \epsilon$ still holds after a *linear* number of rounds, for any $k = q + 1$ with q an arbitrary prime power.

[6]We note that Feldman et al. [FMS$^+$07] first proved that LP decoding corrects $\Omega(1/d_c)$ on expanding graphs. Their proof was recently simplified by Viderman [Vid13] who also slightly relaxed the expansion requirements. Both works assumed that all variable nodes have the same degree but the proof readily extends to the case where variable nodes can have degree either d_v or $d_v - 2$, which is the typical case for random (d_v, d_c)-LDPC codes.

Theorem 8.1.3. *Let $k = q + 1$ where q is any prime power. For any $\epsilon > 0$, there exist $\beta, \eta > 0$ (depending on k) such that a random k-uniform hypergraph with n vertices and $m = \beta n$ edges, simultaneously satisfies the following two conditions with high probability.*

- *The integral optimum of k-*HYPERGRAPH VERTEX COVER *is at least* $(1 - \epsilon)n$.
- *There is a solution to the ηn rounds of the Lasserre hierarchy of value $\frac{1}{k-1}n$.*

8.1.1 Proof Techniques

The LP of Feldman et al. [FWK05, Fel03] is a relaxation of the NEAREST CODEWORD problem, where given a binary linear code (represented by its parity-check matrix or graph) and a received vector, the goal is to find the codeword that is closest to it in Hamming distance. The NEAREST CODEWORD problem can be viewed as a particular case of a variant of Constraint Satisfaction Problems (CSPs) called MIN ONES, where the goal is to find an assignment that satisfies all constraints while minimizing the number of ones in the assignment (see [KSTW01] for more on MIN ONES problems). In this MIN ONES view, each codeword bit corresponds to a binary variable that the decoder should decide whether to flip or not.

Recently, there has been a significant progress in understanding the limitations of LP and SDP hierarchies for CSPs (e.g., [GMT09, Sch08, Tul09, Cha13]); in these works, the focus was on a different variant of CSPs called MAX CSP, where the goal is to find an assignment maximizing the number of satisfied constraints. These results construct fractional solutions satisfying all constraints and that are typically *balanced* in that any coordinate of the assignment is set to 1 with probability $1/2$ in the case of a binary alphabet. Therefore, they yield a fractional solution where half the variables are fractionally flipped.

In order to construct a fractional solution with a smaller number of (fractionally) flipped variables, we introduce the technique of *stretching* and *collapsing* the domain. Given an instance of the NEAREST CODEWORD problem, we stretch the domain into a finite set G via a map $\phi : G \to \{0, 1\}$. The new CSP instance has the same set V of variables but each variable now takes values in G (as opposed to $\{0, 1\}$). A constraint in the new instance on variables $(v_1, \ldots, v_k)$ is satisfied by an assignment $f : V \to G$ if and only if it is satisfied in the original instance by the assignment $\phi \circ f : V \to \{0, 1\}$. Assume that the map ϕ satisfies $|\phi^{-1}(1)| = 1$ and that the previous results for MAX CSP yield a fractional solution over alphabet G such that each variable v takes any particular value $g \in G$ with probability $1/|G|$. If we can transform this fractional solution into one for the original instance by collapsing $\phi^{-1}(i)$ back to i for every $i \in \{0,1\}$, we would get a fractional solution to the original (binary) instance of the NEAREST CODEWORD problem with value $1/|G|$. In Section 8.3, we show that this stretching and collapsing idea indeed works. This technique can be generalized to any MIN ONES problem (e.g., k-HYPERGRAPH VERTEX COVER).

To apply the known constructions for MAX CSP between our stretching and collapsing steps, we need to construct special structures that are required by those results. For the Sherali-Adams hierarchy in the case of the NEAREST CODEWORD problem, we need to construct two *balanced pairwise independent distributions* on G^k: one supported only on vectors with an *even* number of 0 coordinates and the other supported only on vectors with an *odd* number of 0 coordinates.[7] For the Lasserre hierarchy, we need to construct two *cosets of balanced pairwise independent subgroups*: one supported only on vectors with an even number of 0 coordinates and the other supported only on vectors with an odd number of 0 coordinates.

Constructing the desired balanced pairwise independent distributions in the Sherali-Adams hierarchy can be done by setting up systems of linear equations (one variable for each allowed vector $(x_1, \ldots, x_k)$ modulo symmetry) and checking that the resulting solution yields a valid probability distribution (see Section 8.4.1 for more details). Constructing the desired cosets of balanced pairwise independent subgroups in the Lasserre hierarchy is more involved and our algebraic construction is based on designing sets of points in $\mathbb{F}_q^d$ (for q any power of two and $d = 2, 3$) with special hyperplane-incidence properties. One example is the construction (for every power q of 2) of a subset E of $q+2$ points in $\mathbb{F}_q^2$ containing the origin and such that every line in the $\mathbb{F}_q^2$-plane contains either 0 or 2 points in E. See Section 8.4.2 for more details.

Finally, random (d_v, d_c)-LDPC codes typically have check nodes with slightly different degrees whereas in the CSP literature, it is common to assume that all the constraints contain the same number of variables. Since our algebraic constructions of cosets of balanced pairwise independent subgroups for Lasserre hold only for specific arity values, we need an additional technique to obtain the required predicates for both arity d_c and arity $d_c - 2$ (which are with high probability the two possible check-degrees in a random (d_v, d_c)-LDPC code). We construct such predicates by taking the *direct-sums* of pairs and triples of previously constructed cosets, at the expense of multiplying the value of the fractional solution by an absolute constant.

[7] Here, we are assuming WLOG that $0 \in G$. In fact, we can consider any fixed element of the set G.

8.1.2 Organization

Section 8.2 provides background on the problems and hierarchies that we study in this paper. Section 8.3 introduces the *stretching* and *collapsing* technique and shows how to leverage previous results for MAX CSP to reduce our problem to the construction of special distributions and cosets. This general result holds for any MIN ONES problem. Section 8.4 provides the desired constructions for the problem of decoding random (d_v, d_c)-LDPC codes, proving Theorem 8.1.1 in Section 8.4.1 and Theorem 8.1.2 in Section 8.4.2. The proof of Theorem 8.1.3 about k-HYPERGRAPH VERTEX COVER can be found in Section 8.5.

8.2 Preliminaries

Constraint Satisfaction Problems (CSPs) and MIN ONES. Fix a finite set G. Let $\mathcal{P} = \{P_1, \dots, P_l\}$ be such that each P_i is a subset of G^{k_i} where k_i is called the *arity* of P_i. Note that unlike the usual definition of CSPs, we do not allow *shifts*, namely: for $b_1, \dots, b_{k_i} \in G$, $P_i + (b_1, \dots, b_{k_i})$ is not necessarily in $\mathcal{P}$. Furthermore, predicates are allowed to have different arities. Let $k_{max} := \max_i k_i$ and $k_{min} := \min_i k_i$. An instance of CSP($\mathcal{P}$) is denoted by $(V, \mathcal{C})$ where V is a set of n variables taking values in G. $\mathcal{C} = \{C_1, \dots, C_m\}$ is a set of m constraints such that each C_i is defined by its *type* $t_i \in \{1, 2, \dots, l\}$ (which represents the predicate corresponding to this constraint) and a tuple of k_{t_i} variables $E_i = (e_{i,1}, \dots, e_{i,k_{t_i}}) \in V^{k_{t_i}}$. In all instances in the paper, each variable appears at most once in each constraint. We sometimes abuse notation and regard E_i as a subset of V with cardinality k_{t_i}. We say that $(V, \mathcal{C})$ is (s, α)-*expanding* if for any set of $s' \leq s$ constraints $\{C_{i_1}, \dots, C_{i_{s'}}\} \subseteq \mathcal{C}$, $|\bigcup_{1 \leq j \leq s'} E_{i_j}| \geq (\sum_{1 \leq j \leq s'} |E_{i_j}|) - \alpha \cdot s'$. It is said to be (s, α)-*boundary expanding* if for any set of $s' \leq s$ constraints $\{C_{i_1}, \dots, C_{i_{s'}}\} \subseteq \mathcal{C}$, the number of variables appearing in exactly one constraint is at least $(\sum_{1 \leq j \leq s'} |E_{i_j}|) - \alpha \cdot s'$. Note that in both definitions, a smaller value of α corresponds to a better expansion. It is easy to see that (s, α)-expansion implies $(s, 2\alpha)$-boundary expansion. An assignment $f : V \to G$ satisfies constraint C_i if and only if $(f(e_{i,1}), \dots, f(e_{i,k_{t_i}})) \in P_{t_i}$. When $G = \{0, 1\}$, any instance of CSP($\mathcal{P}$) is an instance of MIN ONES($\mathcal{P}$), where the goal is to find an assignment f that satisfies every constraint and minimizes $|f^{-1}(1)|$. k-HYPERGRAPH VERTEX COVER problem corresponds to MIN ONES($\{P_\vee\}$) where $P_\vee(x_1, \dots, x_k) = 1$ if and only if there is at least one $1 \leq i \leq k$ with $x_i = 1$.

Balanced Pairwise Independent Subsets and Distributions. Let G be a finite set with $|G| = q$ and k be a positive integer. Let P be a subset of G^k and μ be a distribution supported on P. The distribution μ is said to be *balanced* if for all $i = 1, 2, \ldots, k$ and $g \in G$, $\Pr_{(x_1,\ldots,x_k)\sim\mu}[x_i = g] = \frac{1}{q}$. It is called *balanced pairwise independent* if for all $i \neq j$ and $g, g' \in G$, $\Pr_{(x_1,\ldots,x_k)\sim\mu}[x_i = g \text{ and } x_j = g'] = \frac{1}{q^2}$. The predicate P is called balanced (resp. balanced pairwise independent) if the uniform distribution on P induces a balanced (resp. balanced pairwise independent) distribution on P^k.

NEAREST CODEWORD. Fix the domain to be $\{0, 1\}$. The NEAREST CODEWORD problem is defined as MIN ONES($\{P_{odd}, P_{even}\}$), where $x = (x_1, \ldots, x_k) \in \{0, 1\}^k$ belongs to P_{odd} (resp. P_{even}) if and only if $|\{i \in [k] : x_i = 1\}|$ is an odd (resp. even) integer. We slightly abuse the notation and let P_{odd} (resp. P_{even}) represent the odd (resp. even) predicates for all values of k. Let $B = (L \cup R, E_B)$ be the parity-check graph of some binary linear code with $|L| = n$ and $|R| = m$. Let $s \in \{0, 1\}^n$ be the received vector (i.e., the codeword which is corrupted by the noisy channel). Denote $R := \{1, \ldots, m\}$. The instance of the NEAREST CODEWORD problem given s is given by $V = L$ and for each $1 \leq i \leq m$, $E_i = \{v \in L : (v, i) \in E_B\}$, and $t_i = odd$ if $\sum_{v:(v,i)\in E_B} s_v = 1$ (summation over $\mathbb{F}_2$) and $t_i = even$ otherwise. In an integral assignment $f : L \to \{0, 1\}$, $f(v) = 1$ means that the v-th bit is flipped. So if all the constraints are satisfied, $(s_v + f(v))_{v\in L}$ is a valid codeword and $|f^{-1}(1)|$ is its Hamming distance to s. We say that B is (s, α)-expanding or (s, α)-boundary expanding if the corresponding NEAREST CODEWORD instance is so.

Sherali-Adams Hierarchy. Given an instance $(V, \mathcal{C})$ of CSP($\mathcal{P}$) and a positive integer $t \leq |V|$, we define a *t-local distribution* to be a collection $\{X_S(\alpha) \in [0, 1]\}_{S\subseteq V,|S|\leq t,\alpha:S\to G}$ satisfying $X_\emptyset = 1$ and for any $S \subseteq T \subseteq V$ with $|T| \leq t$ and for any $\alpha : S \to G$

$$\sum_{\beta:T\setminus S\to G} X_T(\alpha \circ \beta) = X_S(\alpha),$$

where $\alpha \circ \beta$ denotes an assignment $T \to G$ whose projections on S and $T \setminus S$ are α and β respectively. Given $t \geq k_{max}$, a solution to the t rounds of the Sherali-Adams hierarchy is a t-local distribution. It is said to *satisfy* a constraint C_i if for any $\alpha : E_i \to G$, $(\alpha(e_{i,1}), \ldots, \alpha(e_{i,k_{t_i}})) \notin P_i$ implies that $X_{E_i}(\alpha) = 0$ (i.e., the local distribution is only supported on the satisfying partial assignments). The solution is *balanced* if for any $v \in V$ and $g \in G$, $X_v(g) := X_{\{v\}}(v \mapsto g) = \frac{1}{|G|}$. If $G = \{0, 1\}$, we say that the solution is p-*biased* if for any $v \in V$, $X_v(1) = p$.

Given an LDPC code, let d_c^{max} be the largest degree of any check node. Fix a code represented by its parity-check graph $G = ([n] \cup [m], E)$, and let $N(j)$ be the set of all neighbors of check node j. The LP relaxation of Feldman et al. is given by:

$$\min \frac{1}{n} \sum_{i=1}^{n} f_i$$

subject to:

$$\forall j \in [m], \; \sum_{S \in E_j} w_{j,S} = 1$$

$$\forall (i,j) \in E, \; \sum_{S \in E_j, S \ni i} w_{j,S} = f_i$$

$$\forall i \in [n], \; 0 \le f_i \le 1$$

$$\forall j \in [m], \; \forall S \in E_j, \; w_{j,S} \ge 0$$

where E_j is the set of all subsets of $N(j)$ of even (resp. odd) cardinality depending on whether the received vector has an even (resp. odd) number of 1's in $N(j)$. The following claim shows that a small number of rounds of the Sherali-Adams hierarchy is at least as strong as the basic LP of Feldman et al.

Claim 8.2.1. *The LP corresponding to d_c^{max} rounds of the Sherali-Adams hierarchy is at least as strong as the LP of Feldman et al.*

Proof. To prove this claim, it is enough to map any feasible solution to the LP corresponding to d_c^{max} rounds of the Sherali-Adams hierarchy into a feasible solution to the LP of Feldman et al. with the same objective value. The map is the following:

- For every $i \in [n]$, let $f_i = X_{\{i\}}(1)$.
- For every $j \in [m]$ and every $S \subseteq N(j)$, let $w_{j,S} = X_{N(j)}(\alpha^S)$ where $\alpha^S \in \{0,1\}^{N(j)}$ is the partial assignment defined by $\alpha_i^S = 1$ if $i \in S$ and $\alpha_i^S = 0$ if $i \in N(j) \setminus S$.

$\square$

Lasserre Hierarchy. Given an instance $(V, \mathcal{C})$ of CSP($\mathcal{P}$) and an integer $t \leq |V|$, a solution to the t rounds of the Lasserre hierarchy is a set of vectors $\{V_S(\alpha)\}_{S \subseteq V, |S| \leq t, \alpha : S \to G}$, such that there exists a $2t$-local distribution $\{X_S(\alpha)\}$ with the property: for any $S, T \subseteq V$ with $|S|, |T| \leq t$ and any $\alpha : S \to G$ and $\beta : T \to G$, we have that

$$\langle V_S(\alpha), V_T(\alpha) \rangle = X_{S \cup T}(\alpha \circ \beta),$$

if α and β are consistent on $S \cap T$, and $\langle V_S(\alpha), V_T(\alpha) \rangle = 0$ otherwise. The solution satisfies a constraint or is balanced if the corresponding local distribution is so.

8.3 Solutions from Desired Structures

In this section, we show how to construct solutions to the Sherali-Adams / Lasserre hierarchy for MIN ONES($\mathcal{P}$) from desired structures. Given an instance of MIN ONES($\mathcal{P}$) where $\mathcal{P} = \{P_1, \ldots, P_l\}$ is a collection of predicates with $P_i \subseteq \{0, 1\}^{k_i}$, we want to construct a solution to the Sherali-Adams / Lasserre hierarchy with small bias. However, in order to obtain a solution to the Sherali-Adams / Lasserre hierarchy for general CSPs, most current techniques [Sch08, GMT09, Tul09, Cha13] need a balanced pairwise independent distribution, and the resulting solution is typically *balanced* as well. Since the domain G is fixed to $\{0, 1\}$, a $\frac{1}{2}$-biased solution seems to be the best we can hope for; in fact, this is what Schoenebeck [Sch08] does for k-HYPERGRAPH VERTEX COVER in the Lasserre hierarchy thereby proving a gap of 2 (for any $k \geq 3$).

To bypass this barrier, we introduce the technique of *stretching* and *collapsing* the domain. Let G' be a new domain with $|G'| = q$ and fix a mapping $\phi : G' \to \{0, 1\}$ (in every stretching in this paper, $|\phi^{-1}(1)| = 1$). For each predicate P_i, let P'_i be the corresponding new predicate $P'_i := \{(g_1, \ldots, g_{k_i}) \in (G')^{k_i} : (\phi(g_1), \ldots, \phi(g_{k_i})) \in P_i\}$. Let $\mathcal{P}' = \{P'_1, \ldots, P'_l\}$. Any instance $(V, \mathcal{C})$ of MIN ONES($\mathcal{P}$) can be transformed to the instance $(V, \mathcal{C}')$ of CSP($\mathcal{P}'$) where variables in V can take a value from G' and each predicate P_i is replaced by the predicate P'_i. The next lemma shows that any solution to the Sherali-Adams / Lasserre hierarchy for the new instance can be transformed to a solution for the old instance by *collapsing back* the domain. For $\beta : S \to \{0, 1\}$, let $\phi^{-1}(\beta)$ be $\{\alpha : S \to G', \phi(\alpha(v)) = \beta(v) \text{ for all } v \in S\}$.

Lemma 8.3.1. *Suppose that* $\{X'_S(\alpha)\}$ *(resp.* $\{V'_S(\alpha)\}$*) is a solution to the LP (resp. SDP) corresponding to* t *rounds of the Sherali-Adams (resp. Lasserre) hiearchy for* $(V, \mathcal{C}')$ *and that satisfies every constraint. Then,* $\{X_S(\beta)\}_{|S| \leq t, \beta : S \to \{0,1\}}$ *(resp.* $\{V_S(\beta)\}_{|S| \leq t, \beta : S \to \{0,1\}}$*)*

defined by

$$X_S(\beta) = \sum_{\alpha \in \phi^{-1}(\beta)} X'_S(\alpha)$$
$$\left(\textit{resp. } V_S(\beta) = \sum_{\alpha \in \phi^{-1}(\beta)} V'_S(\alpha)\right)$$

is a valid solution to the t rounds of the Sherali-Adams (resp. Lasserre) hiearchy for $(V, \mathcal{C})$ that satisfies every constraint. Furthermore, if the solution to the new instance is balanced, the obtained solution to the old instance is $\frac{1}{q}$-biased.

Proof. First, we prove the statment for the Sherali-Adams hierarchy.

Sherali-Adams. By definition, we have that $X_\emptyset = X'_\emptyset = 1$, and $X_S(\alpha) \geq 0$. Moreover, for any $S \subseteq T \subseteq V$ with $|T| \leq t$ and for any $\beta : S \to \{0, 1\}$, we have that

$$\begin{aligned}
&\sum_{\gamma : T \setminus S \to \{0,1\}} X_T(\beta \circ \gamma) \\
&= \sum_{\gamma : T \setminus S \to \{0,1\}} \sum_{\alpha \in \phi^{-1}(\beta \circ \gamma)} X'_T(\alpha) \\
&= \sum_{\beta' \in \phi^{-1}(\beta)} \sum_{\gamma : T \setminus S \to \{0,1\}} \sum_{\gamma' \in \phi^{-1}(\gamma)} X'_T(\beta' \circ \gamma') \\
&= \sum_{\beta' \in \phi^{-1}(\beta)} \sum_{\gamma : T \setminus S \to G'} X'_T(\beta' \circ \gamma') \\
&= \sum_{\beta' \in \phi^{-1}(\beta)} X'_S(\beta') \\
&= X_S(\beta).
\end{aligned}$$

Furthermore, if $\{X'_S(\alpha)\}$ is balanced, then for any v, $X_v(1) = \sum_{g \in \phi^{-1}(1)} X'_v(g) = \frac{|\phi^{-1}(1)|}{q} = \frac{1}{q}$. This concludes the proof for the Sherali-Adams hierarchy.

Lasserre. Given a solution $\{V'_S(\alpha)\}_{|S| \leq t, \alpha : S \to G}$ to the t rounds of the Lasserre hierarchy, let $\{X'_S(\alpha)\}_{|S| \leq 2t, \alpha : S \to G}$ be the $2t$-local distribution associated with $\{V'_S(\alpha)\}$. Let the $2t$-local distribution $\{X_S(\beta)\}_{|S| \leq 2t, \beta : S \to \{0,1\}}$ be obtained from $\{X'_S(\alpha)\}$ as as done above for the Sherali-Adams hierarchy. It is a valid $2t$-local distribution. We claim that $\{X_S(\beta)\}$ is the local distribution associated with $\{V_S(\beta)\}$. Fix S, T such that $|S|, |T| \leq t$, $\beta : S \to \{0,1\}$ and $\gamma : T \to \{0,1\}$. By the definition of $V_S(\beta)$ and $V_T(\gamma)$, we have that

$$\begin{aligned}\langle V_S(\beta), V_T(\gamma)\rangle &= \langle \sum_{\beta' \in \phi^{-1}(\beta)} V'_S(\beta'), \sum_{\gamma' \in \phi^{-1}(\gamma)} V'_T(\gamma')\rangle \\ &= \sum_{\beta' \in \phi^{-1}(\beta)} \sum_{\gamma' \in \phi^{-1}(\gamma)} \langle V'_S(\beta'), V'_T(\gamma')\rangle.\end{aligned}$$

If β and γ are inconsistent, then any $\beta' \in \phi^{-1}(\beta)$ and $\gamma' \in \phi^{-1}(\gamma)$ are inconsistent, and hence the RHS is 0 as desired. If they are consistent, then the RHS is equal to

$$\begin{aligned}&\sum_{\beta' \in \phi^{-1}(\beta), \gamma' \in \phi^{-1}(\gamma) \text{ consistent}} \langle V'_S(\beta'), V'_T(\gamma')\rangle \\ &= \sum_{\alpha' \in \phi^{-1}(\beta \circ \gamma)} X'_{S \cup T}(\alpha') = X_{S \cup T}(\beta \circ \gamma).\end{aligned}$$

If $\{V'_S(\alpha)\}$ is balanced, by definition $\{X'_S(\alpha)\}$ is balanced, so the same proof for the Sherali-Adams hierarchy shows that $\{V_S(\alpha)\}$ and $\{X_S(\alpha)\}$ are $\frac{1}{q}$-biased. $\square$

By Lemma 8.3.1 above, it suffices to construct a solution to the stretched instance. Theorems 8.3.1 and 8.3.2 below show that if the predicates $P_1, \ldots, P_l$ satisfy certain desired properties and the instance is sufficiently expanding, there exists a balanced solution to the Sherali-Adams / Lasserre hierarchy. The proof is close to [GMT09] for the Sherali-Adams hierarchy and to [Sch08, Tul09, Cha13] for the Lasserre hierarchy. Compared to their proofs for MAX CSP, we have to deal with 2 more issues. The first is that unlike usual CSPs, our definition of MIN ONES($\mathcal{P}$) allows to use more than one predicate, and predicates can have different arities. The second is that for our purposes, the solution needs to be balanced (i.e., $X_v(g) = \frac{1}{|G|}$ for all v, g). We handle those differences by natural extensions of their techniques.

Theorem 8.3.1. *Let G be a finite set, $k_{min} \geq 3$, and $\mathcal{P} = \{P_1, \ldots, P_l\}$ be a collection of predicates such that each $P_i \subseteq G^{k_i}$ supports a balanced pairwise independent distribution μ_i. Let $(V, \mathcal{C})$ be an instance of CSP($\mathcal{P}$) such that $\mathcal{C}$ is $(s, 2+\delta)$-boundary expanding for some $0 < \delta \leq \frac{1}{4}$. Then, there exists a balanced solution to the $\frac{\delta s}{6k_{max}}$ rounds of the Sherali-Adams hierarchy that satisfies every constraint in $\mathcal{C}$.*

We point out that the updated version [BGMT12] of [GMT09] shows that their construction also works in the Sherali-Adams SDP hierarchy which is stronger than the original Sherali-Adams hierarchy but weaker than Lasserre. Both Theorems 8.3.1 and 8.1.1

hold for the Sherali-Adams SDP hierarchy as well. In the proofs of Theorems 8.3.1 and 8.1.1, we focus on the original Sherali-Adams hierarchy to make the presentations simple.

Proof. The proof closely follows Theorem 4.3 of Georgiou, Magen, and Tulsiani [GMT09]. Their result, as a black-box, gives a solution to the Sherali-Adams hierarchy that satisfies all the constraints. There are two additional things that we need to check:

- More than one predicate: Unlike usual CSPs, our definition of MIN ONES($\mathcal{P}$) allows to use more than one predicate, and predicates can have different arities.
- Balanced solution: For our purposes, we need the solution to be balanced (i.e., $X_v(g) = \frac{1}{|G|}$ for all v and g).

The main part of their proof (Lemma 3.2) is robust to the two issues described above. As many technical parts of the proof can be used as a black-box, we sketch the high-level ideas of the proof and highlight the reason why it is robust to the two issues discussed above. We give the following additional definitions for a CSP-instance after removing some variables: Given an instance $(V, \mathcal{C})$ of CSP($\mathcal{P}$) and a subset $S \subseteq V$, let $\mathcal{C}(S)$ denote the set of all constraints that are entirely contained in S, namely: $\mathcal{C}(S) := \{C_i : E_i \subseteq S\}$). Let $(V \setminus S, \mathcal{C} \setminus \mathcal{C}(S))$ be the *instance after removing* S, namely: for each $C_i \in \mathcal{C} \setminus \mathcal{C}(S)$, the set E_i is replaced by $E_i \cap (V \setminus S)$ and its predicate becomes the corresponding projection of P_{t_i} on $G^{|E_i \cap (V \setminus S)|}$.

Expansion Correction. Let S be a subset of V and $\mathcal{C}(S) = \{C_i = (E_i, t_i)\}_{i=1,\ldots,m_S}$ be the constraints induced by S. Each predicate P_{t_i} is associated with a balanced pairwise independent distribution μ_{t_i}. Perhaps the most natural way to combine these distributions to define a local distribution on the assignments $\{\alpha : S \to G\}$ is to take the (normalized) product of all the distributions, i.e.,

$$\Pr_S[\alpha] = \left(\prod_{i=1}^{m_S} \mu_{t_i}(\alpha(e_{i,1}), \ldots, \alpha(e_{i,k_{t_i}}))\right) / Z_S,$$

$$Z_S = \sum_{\alpha : S \to G} \prod_{i=1}^{m_S} \mu_{t_i}(\alpha(e_{i,1}), \ldots, \alpha(e_{i,k_{t_i}})).$$

Call this distribution *canonical* for S. Clearly, any assignment α that has a positive probability will satisfy all constraints in $\mathcal{C}(S)$.

For any subset S, we can define the canonical local distribution. But generally the distributions will not be consistent (i.e., for some $S \subseteq S'$, the canonical distribution on S might be different from the marginal distribution on S obtained from the canonical distribution on S'). Since the canonical distribution on S' induces a local distribution on any $S \subseteq S'$, it might be possible that the canonical distributions of carefully chosen sets are consistent and induce a local distribution for every set we are interested in.

Georgiou et al. [GMT09] define the canonical distribution on some family $\bar{\mathcal{S}}$ of sets that satisfies the following conditions:

- Any $\bar{S} \in \bar{\mathcal{S}}$ satisfies $|\bar{S}| \leq \frac{s}{4}$.
- For any set $S \subseteq V$ with $|S| \leq \delta s/(6k_{max})$, there is an $\bar{S} \in \bar{\mathcal{S}}$ such that $S \subseteq \bar{S}$.
- For any $\bar{S} \in \bar{\mathcal{S}}$, the instance $(V \setminus \bar{S}, \mathcal{C} \setminus \mathcal{C}(\bar{S}))$, obtained by removing $\bar{S}$ and its induced constraints, is $(\frac{3}{4}s, \frac{8}{3}+\delta)$-boundary expanding. Recall that $(V \setminus \bar{S}, \mathcal{C} \setminus \mathcal{C}(\bar{S}))$ is different from the induced instance $(V \setminus \bar{S}, \mathcal{C}(V \setminus \bar{S}))$.

The existence of such an $\bar{\mathcal{S}}$ is shown in Theorem 3.1 of [BGMT12].[8]

Consistent Distributions. The final local distributions $\{X_S(\alpha)\}$ are defined as follows: for each S, find $\bar{S} \in \bar{\mathcal{S}}$ that contains S, and use the canonical distribution defined on $\bar{S}$. It only remains to show that for any $\bar{S}, \bar{S}' \in \bar{\mathcal{S}}$, their canonical distributions are consistent. The following lemma is the crucial part of [GMT09].

Lemma 8.3.2. *[Lemma 3.2 of [GMT09]] Let $(V, \mathcal{C})$ be a CSP-instance as above and $S_1 \subseteq S_2$ be two sets of variables such that both $(V, \mathcal{C})$ and $(V \setminus S_1, \mathcal{C} \setminus \mathcal{C}(S_1))$ are $(t, 2+\delta)$-boundary expanding for some $\delta \in (0,1)$ and $|\mathcal{C}(S_2)| \leq t$. Then for any $\alpha_1 \in G^{S_1}$,*

$$\sum_{\alpha_2 \in G^{S_2}, \alpha_2(S_2) = \alpha_1} \Pr_{S_2}[\alpha_2] = \Pr_{S_1}[\alpha_1].$$

Applying Lemma 8.3.2 two times (once with $(S_1, S_2) \leftarrow (\bar{S}, \bar{S} \cup \bar{S}')$ and once with $(S_1, S_2) \leftarrow (\bar{S}', \bar{S} \cup \bar{S}')$), we conclude that both $\Pr_{\bar{S}}$ and $\Pr_{\bar{S}'}$ are marginal distributions of $\Pr_{\bar{S} \cup \bar{S}'}$, and hence should be consistent.

We check the two issues which are not explicitly dealt in their paper. First, we note that $\Pr_S$ is defined as long as we have a distribution μ_i for each predicate P_i. The proof

[8] The corresponding theorem in the original version [GMT09] seems to have a minor error, so we here follow the final version of their work.

of Lemma 8.3.2 only depends on the fact that each μ_i is balanced pairwise independent and not on any further structure of the predicates. Furthermore, predicates having different arities are naturally handled as long as we have $(t, 2+\delta)$-boundary expansion and pairwise independent distributions. Therefore, having more than one predicate with different arities does not affect the statement. Finally, we check that the resulting local distribution is balanced. Fix any variable $v \in V$ and let $\bar{S} \in \bar{\mathcal{S}}$ be a set containing v. Applying Lemma 8.3.2 with $S_1 \leftarrow \{v\}$ and $S_2 \leftarrow \bar{S}$ ($\Pr_{\{v\}}$ is the uniform distribution on G since $\{v\}$ does not contain any constraint), we get that the canonical distribution on $\bar{S}$ induces the uniform distribution on G for v. $\square$

Theorem 8.3.2. *Let G be a finite abelian group, $k_{min} \geq 3$ and $\mathcal{P} = \{P_1, \ldots, P_l\}$ be a collection of predicates such that each P_i is a coset of a balanced pairwise independent subgroup of G^{k_i}. Let $(V, \mathcal{C})$ be an instance of CSP($\mathcal{P}$) such that $\mathcal{C}$ is $(s, 1+\delta)$-expanding for $\delta \leq \frac{1}{4}$. Then, there exists a balanced solution to the $\frac{s}{16}$ rounds of the Lasserre hierarchy that satisfies every constraint in $\mathcal{C}$.*

Proof. The proof closely follows Theorem D.9 of Chan [Cha13], which generalizes the work of Schoenebeck [Sch08] and Tulsiani [Tul09]. His result, as a black-box, gives a solution to the Lasserre hierarchy that satisfies all the constraints. There are two additional things that we need to check:

- More than one predicate: Unlike usual CSPs, our definition of MIN ONES($\mathcal{P}$) allows to use more than one predicate, and predicates can have different arities.
- Balanced solution: For our purposes, we need the solution to be balanced (i.e., $||V_v(g)||_2^2 = \frac{1}{|G|}$ for all v and g).

Since these are immediate consequences of the previous results, instead of proving them in details, we describe the high-level ideas of the construction while focusing on the points that we need to check.

Describing Each Predicate by Linear Equations. Let $\mathbb{T}$ be the unit circle in the complex plane. Given a finite abelian group G, let $\hat{G}$ be the set of characters (homomorphisms from G to $\mathbb{T}$). $\hat{G}$ is again an abelian group (under pointwise multiplication) with the same cardinality as G. The identity is the all-ones function $\mathbf{1}$, and the inverse of χ is $\frac{1}{\chi} = \bar{\chi}$, where $\bar{\cdot}$ indicates the complex conjugate.

Consider $\widehat{G^V}$ which is isomorphic to $\hat{G}^V$. A character $\chi = (\chi_v)_{v \in V} \in \hat{G}^V$ is said to be v-relevant if $\chi_v \in \hat{G}$ is not the trivial character. The support of a character χ is defined to be $\mathrm{supp}(\chi) := \{v \in V : \chi \text{ is } v\text{-relevant}\}$, and the weight of χ is $|\chi| := |\mathrm{supp}(\chi)|$.

A linear equation is a pair $(\chi, z) \in \hat{G}^V \times \mathbb{T}$, and an assignment $f : V \to G$ satisfies (χ, z) if and only if $\chi(f) := \prod_v \chi_v(f(v)) = z$. Given a constraint $C_i = (E_i, t_i)$ where the predicate P_{t_i} is a coset of a subgroup of G^{k_i}, there is a set of linear equations L_i such that an assignment f satisfies C_i if and only if it satisfies all the linear equations in L_i. See Section D.1 of Chan [Cha13] for technical details. Since each predicate is equivalently formulated by a set of linear equations, having different predicates will not matter, as long as the linear equations have the desired properties.

Resolution Complexity. Given an instance of MIN ONES $(V, \mathcal{C})$ and the set $\mathcal{L} := \cup_i L_i$ of linear equations describing all the predicates, its width-t resolution $\mathcal{L}_t$ is the smallest set satisfying the following:

- $\mathcal{L} \subseteq \mathcal{L}_t$.
- $(\chi, z), (\psi, y) \in \mathcal{L}_t$ and $|\chi\bar{\psi}| \leq t \Rightarrow (\chi\bar{\psi}, z\bar{y}) \in \mathcal{L}_t$. Say $(\chi\bar{\psi}, z\bar{y})$ is *derived* from (χ, z) and (ψ, y).

$\mathcal{L}_t$ is said to *refute* $\mathcal{L}$ if $(\mathbf{1}, z) \in \mathcal{L}_t$ with $z \neq 1$, and $\mathcal{L}_t$ is said to *fix* $v \in V$ if there exists $(\chi, z) \in \mathcal{L}_t$ with $\mathrm{supp}(\chi) = \{v\}$.

Lemma 8.3.3. *If $(V, \mathcal{C})$ is $(s, 1+\delta)$-expanding for $\delta \leq 1/4$ and each predicate is a coset of a balanced pairwise independent subgroup, then $\mathcal{L}_{s/8}$ can neither refute $\mathcal{L}$ nor fix a variable.*

Proof. The proof is identical to that of Theorem 4.3 of Tulsiani, which Theorem D.8 of Chan follows, except that they only prove the lemma for refutation. We give the high-level ideas of the proof, pointing out that fixing a variable is also impossible.

Assume towards contradiction that $\mathcal{L}_t$ refutes $\mathcal{L}$ or fixes a variable, and let $(\chi^*, z^*) \in \mathcal{L}_t$ with $|\chi^*| \in \{0, 1\}$. Without loss of generality, we can assume that (χ^*, z^*) is derived from $\{(\chi_i, z_i) | 1 \leq i \leq m\}$, where each (χ_i, z_i) is derived only from L_i. Let $S^* := \{i : \chi_i \neq \mathbf{1}\}$ and $s^* := |S^*|$. The crucial property they use is that χ_i with $i \in S^*$ has weight at least 3, which follows from the condition on predicates: Tulsiani requires a predicate to be a linear code of dual distance at least 3, and Chan requires it to be a balanced pairwise independent subgroup, which are indeed equivalent when G is a finite field.

If $s^* \leq s$, since the instance is $(s, 1+\delta)$-expanding, out of $\sum_{i \in S^*} |E_i|$ constraint-variable pairs $(i, e_{i,j})_{i \in S^*, 1 \leq j \leq k_{t_i}}$, at most $(2+2\delta)s^*$ pairs have another pair with the same variable. Since each χ_i with $i \in S^*$ has $|\chi_i| \geq 3$ and contributes 3 such pairs, at least $3s^* - (2+2\delta)s^* = (1-2\delta)s^*$ variables are covered exactly once by $\{\mathrm{supp}(\chi_i)\}_{i \in S^*}$, making it impossible to derive any (χ, z) with $|\chi| < (1-2\delta)s^*$. It shows that $s^* > s$. The original argument (Claim 4.4 of [Tul09]) assumed that every predicate is of the same arity, but the above argument naturally adapted it to irregular arities.

Backtracking the derivations, we must have $(\chi^*, z^*) \in \mathcal{L}_{s/8}$, which is derived from $\frac{s}{2} \le s^* \le s$ nontrivial characters from L_i's (Claim 4.5 of Tulsiani). Similar expansion / minimum weight arguments again ensure that $|\chi^*| > \frac{s}{8}$, which results in a contradiction. $\square$

Solution and Balance. Given that $\mathcal{L}_{s/8}$ does not refute $\mathcal{L}$, Theorem D.5 of [Cha13] ensures that there exists a solution $\{V_S(\alpha)\}_{|S| \le s/16, \alpha: S \to G}$ to the $s/16$ rounds of the Lasserre hierarchy that satisfies every constraint. Furthermore, one of his lemmas also proves that for every $v \in V$ and $g \in G$, $||V_v(g)||_2^2 = \frac{1}{|G|}$ using the fact that $\mathcal{L}_{s/8}$ does not fix any variable.

Lemma 8.3.4 (Proposition D.7 of [Cha13]). *For $S \subseteq V$ with $|S| \le s/16$, let*

$$H_S := \Big\{ \beta | \beta : S \to G \textit{ and } \beta \textit{ satisfies every}$$
$$(\chi, z) \in \mathcal{L}_{s/8} \textit{ with } \mathrm{supp}(\chi) \subseteq S \Big\}.$$

For any $\alpha : S \to G$,

$$||V_S(\alpha)||_2^2 = \frac{\mathbb{I}[\alpha \in H_S]}{|H_S|},$$

where $\mathbb{I}[\cdot]$ is the indicator function.

Combining all three parts above, we have a balanced solution to the $\frac{s}{16}$ rounds of the Lasserre hierarchy that satisfies every constraint. $\square$

8.4 Decoding Random (d_v, d_c)-LDPC Codes

In this section, we apply Theorems 8.3.1 and 8.3.2 to random (d_v, d_c)-LDPC codes. In Section 8.4.1, we construct balanced pairwise independent distributions supported on even and odd predicates for different arity values and complete the proof of Theorem 8.1.1 for Sherali-Adams. In Section 8.4.2, we show that both even and odd predicates contain cosets of balanced pairwise independent subgroups and introduce an additional technique based on taking the direct-sum of cosets of subgroups to conclude the proof of Theorem 8.1.2

for Lasserre. We will need the next two lemmas which show that with high probability, a random (d_v, d_c)-LDPC code is almost regular and expanding.

Lemma 8.4.1. *Consider the parity-check graph of a random (d_v, d_c)-LDPC code. With high probability, every vertex on the left (resp. right) will have degree either d_v or $d_v - 2$ (resp. d_c or $d_c - 2$).*

Proof. Let $M := nd_v = md_c$. Fix a vertex v on the left. In order to have at most $d_v - 2$ neighbors, v needs to either have a neighbor with triple edges or two neighbors with double edges. The probability of the first event is at most by $m \cdot \binom{d_v}{3} \cdot \binom{d_c}{3} \cdot 3! \cdot \frac{1}{M(M-1)(M-2)} = O(\frac{1}{n^2})$. The probability of the second event is at most by $m^2 \cdot \binom{d_v}{4} \cdot (\binom{d_c}{2})^2 \cdot 4! \cdot \frac{1}{M(M-1)(M-2)(M-3)} = O(\frac{1}{n^2})$. By taking a union bound over all v, the probability that there exists a vertex with at most $d_v - 2$ different neighbors is $O(\frac{1}{n})$. The proof for the right side is similar. □

Lemma 8.4.2. *Given any $0 < \delta < 1/2$, there exists $\eta > 0$ (depending on d_c) such that the parity-check graph of a random (d_v, d_c)-LDPC code is $(\eta n, 1 + \delta)$-expanding with high probability.*

Proof. Let $k := d_c$. Fix a set S of $s \leq \eta m$ vertices on the right for some $\eta > 0$ chosen later. Suppose that the degree of each vertex in S is given. By the above lemma, with high probability, each degree is either k or $k-2$. Let $\bar{k}$ be the average degree of these s vertices, and $\bar{c} = \bar{k} - 1 - \delta$. Fix a set Γ of $\bar{c}s$ vertices on the left.

For a vertex $v \in S$ with degree k', the probability that it has all k' neighbors from Γ is at most $(\frac{2\bar{c}s}{n})^{k'}$. If we condition that other vertices in S have neighbors in Γ, this estimate only decreases. Therefore, the probability that the vertices in S have neighbors only from the Γ is at most $(\frac{2\bar{c}s}{n})^{\bar{k}s}$. Taking a union bound over $\binom{n}{\bar{c}s} \leq (\frac{ne}{\bar{c}s})^{\bar{c}s}$ choices of Γ, conditioned on any degrees of S, the probability of the bad event conditioned on any sequence of degrees of S is at most

$$(\frac{2\bar{c}s}{n})^{\bar{k}s} \cdot (\frac{ne}{\bar{c}s})^{\bar{c}s} \leq n^{(-1-\delta)s}(ks)^{(1+\delta)s}(2e)^{ks}.$$

Taking a union bound over $\binom{m}{s} \leq \binom{n}{s} \leq (\frac{en}{s})^s$ choices for S, the probability that some set S of size s becomes bad is at most $(\frac{s}{n})^{\delta s}(k^{1+\delta}(2e)^{k+1})^s$. Let $\beta = k^{1+\delta}(2e)^{k+1}$ so that the above quantity becomes $(\frac{s}{n})^{\delta s}\beta^s = (\frac{s\beta^{1/\delta}}{n})^s$. When we sum this probability over all

$s \le \eta n$, we have

$$\sum_{s=1}^{\eta n} (\frac{s\beta^{1/\delta}}{n})^{\delta s} = \sum_{s=1}^{\ln^2 n} (\frac{s\beta^{1/\delta}}{n})^{\delta s} + \sum_{s=\ln^2 n+1} (\frac{s\beta^{1/\delta}}{n})^{\delta s}$$
$$\le O(\frac{\beta^1}{n^\delta} \ln^2 n) + O((\eta \cdot \beta^{1/\delta})^{\delta \ln^2 n}).$$

The first term is $o(1)$ for large n. The second term is also $o(1)$ for $\eta < 1/(\beta^{1/\delta})$. $\square$

8.4.1 Distributions for Sherali-Adams

To construct a solution for the Sherali-Adams hierarchy using Theorem 8.3.1, we need each $P'_i \subseteq (G')^{k_i}$ to support a balanced pairwise independent distribution. For any $q \ge 2$ and $k = q + 1$, let $G' := \{0, 1, \ldots, q - 1\}$ and $\phi : G' \to \{0, 1\}$ be defined by $\phi(0) = 1$ and $\phi(g) = 0$ for every $g \ne 0$. The *odd* and *even* predicates P'_{odd} and P'_{even} are defined by: $y \in P'_{odd}$ (resp. P'_{even}) if and only if $|\{i \in [k] : y_i = 0\}|$ is an odd (resp. even) integer. The choice of $k = q + 1$ is optimal since, as shown in the following lemma, if $k = q$, there is no balanced pairwise independent distribution that is supported on the even larger predicate $\{y \in (G')^k : y_i = 0 \text{ for some } i\}$ which contains P'_{odd}.

Lemma 8.4.3. *Let $G = \{0, \ldots, k - 1\}$ be a finite set. There is no balanced pairwise independent distribution ν on G^k where every atom $(x_1, \ldots, x_k)$ in the support has at least one 0 coordinate.*

Proof. Given $x = (x_1, \ldots, x_k) \in G^k$, let $|x|$ be the number of 0's among $x_1, \ldots, x_k$. The fact that μ is balanced implies $\mathbb{E}_{x\sim\mu}[|x|] = 1$, but the other requirement implies $|x| \ge 1$ for any x in the support. Therefore, any x in the support satisfies $|x| = 1$. Fix any $i \ne j$. If $x_i = 0$, x_j cannot be 0 and x_i and x_j are not pairwise independent. $\square$

Set $p := 1/q$. To construct a distribution on $y \in (G')^k$, we will show how to sample $x \in \{0, 1\}^k$. Given x, each y_i is set to 0 if $x_i = 0$ and uniformly sampled from $\{1, \ldots, q - 1\}$ otherwise. It is easy to see that when this distribution on x is $(1 - p)$-biased (i.e. $\Pr[x_i = 0] = p$ for all i) and pairwise independent (i.e. $\Pr[x_i = x_j = 0] = p^2$) for all $i \ne j$), y becomes balanced pairwise independent. Furthermore, x and y have the same number of 0's. Therefore, it suffices to show how to sample a $(1 - p)$-biased pairwise independent vector x.

Odd predicate, Odd $k \geq 3$, $q = k-1$. Let $\mathbf{0} := (0, \ldots, 0)$, $\mathbf{1} = (1, \ldots, 1)$ and e_i be the i-th unit vector. Sample $x \in (G')^k$ from the distribution with probability mass function: $\Pr[x = \mathbf{0}] = p^2$ and $\Pr[x = \mathbf{1} - e_i] = \frac{1-p^2}{k}$ for each i. Each support-vector has an odd number of 0's. For any i, $\Pr[x_i = 0] = \Pr[x = \mathbf{1} - e_i] + \Pr[x = \mathbf{0}] = \frac{1-p^2}{k} + p^2 = p$. For any $i \neq j$, $\Pr[x_i = x_j = 0] = \Pr[x = \mathbf{0}] = p^2$. This simple construction is optimal: If $k = q+1$ is even, the following lemma shows that there is no such balanced pairwise independent distribution supported in P'_{odd}.

Lemma 8.4.4. *Let $G = \{0, \ldots, k-2\}$ be a finite set for even k. There is no balanced pairwise independent distribution ν on G^k where every atom $(x_1, \ldots, x_k)$ in the support has an odd number of zeros.*

Proof. Assume for contradiction that such a μ exists. For odd $1 \leq i \leq k-1$, let a_i be the probability that the $(x_1, \ldots, x_k)$ sampled from μ has exactly i zeros. From balanced pairwise independence, they should satisfy the following set of inequalities:

- Valid probability distribution:

$$\sum_{1 \leq i \leq k-1, i \text{ odd}} a_i = 1.$$

- Balance:

$$\sum_{1 \leq i \leq k-1, i \text{ odd}} a_i \cdot \frac{i}{k} = \frac{1}{k-1}$$
$$\iff \sum_{1 \leq i \leq k-1, i \text{ odd}} i a_i = \frac{k}{k-1}.$$

- Pairwise independence:

$$\sum_{3 \leq i \leq k-1, i \text{ odd}} a_i \cdot \frac{i(i-1)}{k(k-1)} = \frac{1}{(k-1)^2}$$
$$\iff \sum_{3 \leq i \leq k-1, i \text{ odd}} i(i-1) a_i = \frac{k}{k-1}.$$

Subtracting the first equation from the second, we get that

$$\sum_{3 \leq i \leq k-1, i \text{ odd}} (i-1) a_i = \frac{1}{k-1}.$$

Subtracting k times this equation from the third equation above, we get that

$$\sum_{3 \le i \le k-1, i \text{ odd}} (i-1)(i-k)a_i = 0,$$

which is contradiction since all $a_i \ge 0$. $\square$

Even Predicate, $k \ge 3$, $q = k - 1$. Sample $x \in (G')^k$ from the distribution with probability mass function: $\Pr[x = \mathbf{1} - e_i - e_j] = p^2$ for each $i \neq j$ and $\Pr[x = \mathbf{1}] = 1 - p^2\binom{k}{2} = \frac{1-p}{2}$. Each support-vector has an even number of 0's. For any i, $\Pr[x_i = 0] = \Pr[\exists j \neq i : x = \mathbf{1} - e_i - e_j] = p^2(k-1) = p$. For $i \neq j$, $\Pr[x_i = x_j = 0] = \Pr[x = \mathbf{1} - e_i - e_j] = p^2$.

Other values of k and q. If $k \ge 4$ is an even integer, By Lemma , there is no balanced pairwise independent distribution that is supported in the odd predicate when $q = k - 1$. However, it is still possible to have such a distribution when $q = k - 2$ for both odd and even predicates. In Lemma 8.4.5 below, we prove the existence of pairwise independent distributions supported in the odd and even predicates for slightly smaller values of q (in terms of k). These distributions will be used to handle instances where the constraints have different arities.

Lemma 8.4.5. *Let $G = \{0, 1, \dots, q-1\}$ be a finite set. For the following combinations of arity values k and alphabet size values q, each of the odd predicate and the even predicate supports a balanced pairwise independent distribution on G^k: (i) Any even integer $k \ge 4$ with $q = k - 2$, (ii) Any odd integer $k \ge 5$ with $q = k - 3$ and (iii) Any even integer $k \ge 6$ with $q = k - 4$.*

Proof. We again construct each distribution by sampling $x \in \{0, 1\}^k$ first. $y = (y_1, \dots, y_k) \in G^k$ is given

- For each i, if $x_i = 0$, $y_i \leftarrow 0$.
- If $x_i \neq 0$, y_i is chosen uniformly from $\{1, \dots, q-1\}$. independently.

If x is $\frac{q-1}{q}$-biased and pairwise independent on $\{0, 1\}^k$, it is easy to check that y is balanced pairwise indepedent on G^k. From now on, we show how to sample the vector x and prove that it satisfies the desired properties.

Even $k \geq 4$, $q = k-2$. We first deal with the odd predicate. Our strategy to sample x is the following. Sample $r \in \{1, 3, k-1\}$ with probabilty a_1, a_3, a_{k-1} respectively. Sample a set R uniformly from $\binom{\{1,2,\ldots,k\}}{r}$ and fix $x_i = 1$ if and only if $i \in R$. The probabilities a_1, a_3, a_{k-1} should satisfy the following three equations.

- Valid probability distribution:

$$a_1 + a_3 + a_{k-1} = 1.$$

- $\left(\frac{q-1}{q}\right)$-biased:

$$\frac{1}{k}a_1 + \frac{\binom{k-1}{2}}{\binom{k}{3}}a_3 + \frac{k-1}{k}a_{k-1} = \frac{1}{k-2}$$
$$\iff a_1 + 3a_3 + (k-1)a_{k-1} = \frac{k}{k-2}.$$

- Pairwise Independence:

$$\frac{\binom{k-2}{1}}{\binom{k}{3}}a_3 + \frac{k-2}{k}a_{k-1} = (\frac{1}{k-2})^2$$
$$\iff 6a_3 + (k-1)(k-2)a_{k-1} = \frac{k(k-1)}{(k-2)^2}.$$

We have that

$$a_1 = \frac{2k^3 - 13k^2 + 25k - 12}{2k^3 - 12k^2 + 24k - 16},$$
$$a_3 = \frac{k-1}{2k^2 - 8k + 8},$$

and

$$a_{k-1} = \frac{k-3}{k^3 - 6k^2 + 12k - 8}$$

is the solution to the above system. They are well-defined and nonnegative for $k \geq 4$.

For the even predicate, we can choose x as above, using $r \in \{0, 2, 4\}$.

- Valid probability distribution:

$$a_0 + a_2 + a_4 = 1.$$

- $\left(\frac{q-1}{q}\right)$-biased:

$$\frac{\binom{k-1}{1}}{\binom{k}{2}}a_2 + \frac{\binom{k-1}{3}}{\binom{k}{4}}a_4 = \frac{1}{k-2}$$
$$\iff 2a_2 + 4a_4 = \frac{k}{k-2}.$$

- Pairwise Independence:

$$\frac{1}{\binom{k}{2}}a_2 + \frac{\binom{k-2}{2}}{\binom{k}{4}}a_4 = (\frac{1}{k-2})^2$$
$$\iff 2a_2 + 12a_4 = \frac{k(k-1)}{(k-2)^2}.$$

We have that

$$a_0 = \frac{4k^2 - 23k + 32}{8k^2 - 32k + 32},$$

$$a_2 = \frac{2k^2 - 5k}{4k^2 - 16k + 16},$$

and

$$a_4 = \frac{k}{8k^2 - 32k + 32}$$

is the solution to the above system. They are well-defined and nonnegative for $k \geq 4$.

Odd $k \geq 5$, $q = k - 3$. We can use the same framework as above, except that in every equation, the denominator of the RHS is changed from $k - 2$ to $k - 3$.

For the even predicate,

$$a_0 = \frac{2k^2 - 17k + 36}{4k^2 - 24k + 36},$$

$$a_2 = \frac{k^2 - 4k}{2k^2 - 12k + 18},$$

and

$$a_4 = \frac{k}{4k^2 - 24k + 36}$$

is the solution to

$$
\begin{aligned}
a_0 + a_2 + a_4 &= 1 \\
2a_2 + 4a_4 &= \frac{k}{k-3} \\
2a_2 + 12a_4 &= \frac{k(k-1)}{(k-3)^2}.
\end{aligned}
$$

They are well-defined and nonnegative for $k \geq 5$.

For the odd predicate, we have that

$$
a_1 = \frac{k^3 - 8k^2 + 16k}{k^3 - 7k^2 + 15k - 9},
$$

$$
a_3 = \frac{k^2 - 4k}{k^3 - 9k^2 + 27k - 27},
$$

and

$$
a_k = \frac{k^2 - 10k + 27}{k^4 - 10k^3 + 36k^2 - 54k + 27}
$$

is the solution to

$$
\begin{aligned}
a_1 + a_3 + a_k &= 1 \\
a_1 + 3a_3 + ka_k &= \frac{k}{k-3} \\
6a_3 + k(k-1)a_k &= \frac{k(k-1)}{(k-3)^2}.
\end{aligned}
$$

They are well-defined and nonnegative for $k \geq 5$.

Even $k \geq 6$, $q = k - 4$. We can use the same framework as above, except that in every equation, the denominator of the RHS is changed from $k-3$ to $k-4$.

For the even predicate,

$$
a_0 = \frac{4k^2 - 45k + 128}{8k^2 - 64k + 128},
$$

$$
a_2 = \frac{2k^2 - 11k}{4k^2 - 32k + 64},
$$

and

$$
a_4 = \frac{3k}{8k^2 - 64k + 128}
$$

is the solution to

$$\begin{aligned} a_0 + a_2 + a_4 &= 1 \\ 2a_2 + 4a_4 &= \frac{k}{k-4} \\ 2a_2 + 12a_4 &= \frac{k(k-1)}{(k-4)^2}. \end{aligned}$$

They are well-defined and nonnegative for $k \geq 6$.

For the odd predicate, we have that

$$a_1 = \frac{2k^3 - 23k^2 + 75k - 48}{2k^3 - 20k^2 + 64k - 64},$$

$$a_3 = \frac{3k^2 - 19k + 16}{2k^3 - 24k^2 + 96k - 128},$$

and

$$a_{k-1} = \frac{k^2 - 13k + 48}{k^4 - 14k^3 + 72k^2 - 160k + 128}$$

is the solution to

$$\begin{aligned} a_1 + a_3 + a_k &= 1 \\ a_1 + 3a_3 + (k-1)a_k &= \frac{k}{k-4} \\ 6a_3 + (k-1)(k-2)a_k &= \frac{k(k-1)}{(k-4)^2}. \end{aligned}$$

They are well-defined and nonnegative for $k \geq 6$. $\square$

The constructed distributions for the Sherali-Adams hierarchy are summarized in Table 8.1.

Proof of Theorem 8.1.1. Consider a random (d_v, d_c)-LDPC code and fix $\delta = 1/8$. Lemma 8.4.1 and Lemma 8.4.2 ensure that with high probability, the degree of each check node is either d_c or $d_c - 2$ and there exists $\eta > 0$ such that the code is $(\eta n, 1+\delta)$-expanding, and hence $(\eta n, 2+2\delta)$-boundary expanding. For any received vector, let $(V, \mathcal{C})$ be the corresponding instance of NEAREST CODEWORD. Let $q = d_c - 3$ (resp. $d_c - 4$) if d_c is odd (resp. even). Stretch the domain from $\{0, 1\}$ to $G' := \{0, 1, \ldots q-1\}$. The above constructions show that for any $k \in \{d_c, d_c - 2\}$ and $type \in \{even, odd\}$, $P_{type} \subseteq (G')^k$ supports

Type \ Arity	d_c odd ($q = d_c - 3$)		d_c even ($q = d_c - 4$)	
	$k = d_c$	$k = d_c - 2$	$k = d_c$	$k = d_c - 2$
Odd	Lemma 8.4.5 (ii)	Section 8.4.1	Lemma 8.4.5 (iii)	Lemma 8.4.5 (i)
Even	Lemma 8.4.5 (ii)	Section 8.4.1	Lemma 8.4.5 (iii)	Lemma 8.4.5 (i)

Table 8.1: Distributions for Sherali-Adams

a balanced pairwise independent distribution. Theorem 8.3.1 gives a balanced solution to the $\frac{2\delta\eta n}{6d_c} = \frac{\eta n}{24d_c}$ rounds of the Sherali-Adams hierarchy that satisfies every constraint in the stretched instance. Lemma 8.3.1 transforms this solution to a $\frac{1}{q}$-biased solution to the same number of rounds for the original NEAREST CODEWORD instance.

8.4.2 Subgroups for Lasserre

As in the Sherali-Adams hierarchy, to find a good solution in the Lasserre hierarchy, it suffices to construct a stretched instance. To construct a solution in the Lasserre hierarchy via Theorem 8.3.2, we need the stretched domain G' to be a finite *abelian group* and each stretched predicate P'_i to be a coset of a balanced pairwise independent *subgroup* of $(G')^k$. We will first construct such predicates for q being any power of 2 and $k = q+1$. For such q and k, let $G' := \mathbb{F}_q$ and $\phi : G' \to \{0,1\}$ be defined by $\phi(0) = 1$ and $\phi(g) = 0$ for every $g \neq 0$. As for Sherali-Adams, the predicates P'_{odd} and P'_{even} are defined in the natural way, namely: $(x_1, \dots, x_k) \in P'_{odd}$ (resp. P'_{even}) if and only if $|\{i \in [k] : x_i = 0\}|$ is an odd (resp. even) integer. We show that each of P'_{odd} and P'_{even} contains a coset of a balanced pairwise independent subgroup of $(G')^k$.

Odd Predicate, $k = 2^i + 1$, $q = k - 1$. For the odd predicate P'_{odd}, we actually show that it contains a balanced pairwise independent subgroup of $(G')^k$. Let $\{\alpha x + \beta y\}_{\alpha,\beta\in\mathbb{F}_q}$ be the set of all q^2 bivariate linear functions over $\mathbb{F}_q$. Let $E := \{(0,1)\} \cup \{(1,a)\}_{a\in\mathbb{F}_q}$ be the set of $q+1 = k$ evaluation points. Our subgroup is defined by $H' := \{(\alpha x + \beta y)_{(x,y)\in E}\}_{\alpha,\beta\in\mathbb{F}_q}$. Note that H' is a subgroup of $(G')^k$. In general, there are $q+1$ distinct lines passing through the origin in the $\mathbb{F}_q^2$-plane; our set E contains exactly one point from each of those lines. The balanced pairwise independence of H' follows from Lemma 8.4.6.

Lemma 8.4.6. *Let $d \in \mathbb{N}$ and $E \subseteq \mathbb{F}_q^d \setminus \{0\}$ contain at most one point from each line passing the origin. Then, the subgroup $\{(\sum_{i=1}^d \alpha_i x_i)_{(x_1,\dots,x_d)\in E}\}_{\alpha_1,\dots,\alpha_d\in\mathbb{F}_q}$ is balanced pairwise independent.*

Proof. Let $(b_1, \dots, b_d) \neq (c_1, \dots, c_d) \in E$ be two points not on the same line passing through the origin. For balanced pairwise independence, we need $(\sum_i \alpha_i b_i, \sum_i \alpha_i c_i)_{\alpha_1, \dots, \alpha_d \in \mathbb{F}_q}$ to be the uniform distribution on $\mathbb{F}_q^2$. Since there are exactly q^d choices for the tuple $(\alpha_1, \dots, \alpha_d)$, for any $\beta, \gamma \in \mathbb{F}_q$, it suffices to show that there exists q^{d-2} choices of the tuple $(\alpha_1, \dots, \alpha_d) \in \mathbb{F}_q$ such that $\sum_i \alpha_i b_i = \beta, \sum_i \alpha_i c_i = \gamma$. Since the two points are not on the same line through the origin, there must be two indices $i \neq j$ such that $b_i c_j \neq b_j c_i$. Without loss of generality, assume that $i = 1$ and $j = 2$. For any choice of $(\alpha_3, \dots, \alpha_d)$, there is exactly one solution (α_1, α_2) to the system:

$$\alpha_1 b_1 + \alpha_2 b_2 = \beta - \sum_{i=3}^{d} \alpha_i b_i$$

$$\alpha_1 c_1 + \alpha_2 c_2 = \gamma - \sum_{i=3}^{d} \alpha_i c_i$$

□

The next lemma concludes the analysis of the odd predicate.

Lemma 8.4.7. *Each element of H' has an odd number of 0 coordinates.*

Proof. Recall that $k = q + 1$ with q a power of 2 and $G' := \mathbb{F}_q$. Our set of evaluation points is defined by

$$E := \{(0,1)\} \cup \{(1,a)\}_{a \in \mathbb{F}_q}$$

and our subgroup H' of $(G')^k$ is defined by

$$H' := \{(\alpha x + \beta y)_{(x,y) \in E}\}_{\alpha, \beta \in F_q}$$

Let $h_{\alpha,\beta} := (\alpha x + \beta y)_{(x,y) \in E}$ be any element of H' (where $\alpha, \beta \in \mathbb{F}_q$). The fact that $h_{\alpha,\beta}$ has an odd number of 0 coordinates can be seen by distinguishing the following three cases:

- For $\alpha = \beta = 0$: $h_{\alpha,\beta} = (0, 0, \dots, 0)$, which has k 0 coordinates, and k is set to be an odd integer.
- For $\beta = 0$ and $\alpha \neq 0$: $(0, 1)$ is the unique zero of the function $\alpha x + \beta y$ in E.
- For $\beta \neq 0$: $(1, \alpha/\beta)$ is the unique zero of the function $\alpha x + \beta y$ in E. □

Even Predicate, $k = 2^i + 1$**,** $q = k - 1$**.** Dealing with P'_{even} is more difficult, since P'_{even} will not contain any subgroup: this can be seen by observing that the zero element $(0, 0, \dots, 0) \in (G')^k$ has an odd number of 0 coordinates and should be in any subgroup. Instead, we show that P'_{even} will contain a *coset* of a balanced pairwise independent subgroup. As in the above case of the odd predicate, our subgroup H' will be of the form $\{(\alpha x + \beta y)_{(x,y)\in E'}\}_{\alpha,\beta\in\mathbb{F}_q}$, for some subset $E' \subseteq \mathbb{F}_q^2$ of $q + 1 = k$ evaluation points. As before, the set E' will contain exactly one non-zero point on each line passing through the origin and hence balanced pairwise independence will follow from Lemma 8.4.6. Moreover, the set E' will have the property that $H' - (1, 1, \dots, 1) \subseteq P'_{even}$; i.e, for any $\alpha, \beta \in \mathbb{F}_q$, there is an even number of points $(x, y) \in E'$ satisfying the equation $\alpha x + \beta y = 1$. For example, if $\alpha = \beta = 0$, no point satisfies this equation. If at least one of α, β is nonzero, then $\{\alpha x + \beta y = 1\}_{(\alpha,\beta)\in\mathbb{F}_q^2\setminus\{(0,0)\}}$ consists of all $(q^2 - 1)$ distinct lines not passing through the origin. Thus, we set $E' := E \setminus \{0\}$ where E is the set which is guaranteed to exist by Lemma 8.4.8.

Lemma 8.4.8. *For every q that is a power of 2, there is a subset $E \subseteq \mathbb{F}_q^2$ containing the origin $(0, 0)$ such that $|E| = q + 2$ and every line in the $\mathbb{F}_q^2$-plane contains either 0 or 2 points in E.*

Proof. Consider the map $h : \mathbb{F}_q \to \mathbb{F}_q$ given by $h(a) = a^2 + a$. Since $h(a) = h(a + 1)$ for all $a \in \mathbb{F}_q$, we can see that h is two-to-one. Hence, there exists $\eta \in \mathbb{F}_q$ such that the polynomial $g(a) = a^2 + a + \eta$ has no roots in $\mathbb{F}_q$. Fix such an η. Define the map $f : \mathbb{F}_q \to \mathbb{F}_q$ by $f(a) = (g(a))^{-1}$ for all $a \in \mathbb{F}_q$. Note that since g has no roots in $\mathbb{F}_q$, f is well defined and non-zero on $\mathbb{F}_q$. Now let $E := \{(0, 0)\} \cup \{(0, 1)\} \cup \{f(a)(1, a) : a \in \mathbb{F}_q\}$. We next argue that every line l in $\mathbb{F}_q^2$ contains either 0 or 2 points in E. We distingish several cases:

- l contains the origin $(0, 0)$: If l is a vertical line, then it has the form $l : (x = 0)$ and $(0, 1)$ is the only other point of E that lies on l. Henceforth, assume that l is non-vertical. Then, it has the form $l : (y = \alpha x)$ for some $\alpha \in \mathbb{F}_q$. In this case, the unique other point of E that lies on l is $f(\alpha)(1, \alpha)$.
- l doesn't contain $(0, 0)$ but contains $(0, 1)$: Thus, it is of the form $l : (y = \alpha x + 1)$ for some $\alpha \in \mathbb{F}_q$. Then, a point $f(a)(1, a)$ lies on l if and only if $af(a) = \alpha f(a) + 1$ which is equivalent to $a = \alpha + g(a)$. This means that a is a root of the polynomial $g(a) + \alpha - a = a^2 + \eta + \alpha$. By Lemma 8.4.9 below, this polynomial has a unique root (of multiplicity 2) in $\mathbb{F}_q$. So l contains exactly 2 points in E.
- l contains neither $(0, 0)$ nor $(0, 1)$: If l is a vertical line, then it has the form $l :$ $(x = \beta)$ for some $\beta \in \mathbb{F}_q \setminus \{0\}$. Then, a point $f(a)(1, a)$ lies on l if and only if $f(a) = \beta$, which is equivalent to $g(a) = \beta^{-1}$ (since $\beta \neq 0$). This means that

a is a root of the polynomial $g(a) - \beta^{-1} = a^2 + a + \eta - \beta^{-1}$. By Lemma 8.4.9 below, this polynomial has either 0 or 2 roots in $\mathbb{F}_q$. Hence, l contains either 0 or 2 points in E. Henceforth, assume that l is non-vertical. Then, it has the form $l : (y = \alpha x + \beta)$ for some $\alpha \in \mathbb{F}_q$ and $\beta \in \mathbb{F}_q \setminus \{0, 1\}$. Then, a point $f(a)(1, a)$ lies on l if and only if $af(a) = \alpha f(a) + \beta$, which is equivalent to $a = \alpha + \beta g(a)$. This is equivalent to $g(a) = a/\beta - \alpha/\beta$. This means that a is a root of the polynomial $g(a) - a/\beta + \alpha/\beta = a^2 + a(1 - 1/\beta) + \eta + \alpha/\beta$. By Lemma 8.4.9 below and since $\beta \neq 1$, this polynomial has either 0 or 2 roots in $\mathbb{F}_q$. So l contains either 0 or 2 points in E. $\square$

Lemma 8.4.9. *Let q be a power of 2. Then, a quadratic polynomial $p(a) = a^2 + c_1 a + c_0$ over $\mathbb{F}_q$ has a unique root (of multiplicity 2) if and only if $c_1 = 0$.*

Proof. If $p(a)$ has a unique root $\lambda \in \mathbb{F}_q$, then $(a - \lambda)$ divides $p(a)$ and hence $p(a) = (a - \lambda)^2 = a^2 - 2\lambda a + \lambda^2$. Since $\mathbb{F}_q$ has characteristic 2, we get that $p(a) = a^2 + \lambda^2$ and we conclude that $c_1 = 0$. Conversely, assume that $p(a) = a^2 + c_0$ for some $c_0 \in \mathbb{F}_q$. Since $\mathbb{F}_q$ has characteristic 2, the map $\kappa : \mathbb{F}_q \to \mathbb{F}_q$ given by $\kappa(a) = a^2$ is a bijection. Hence, there exists $\lambda \in \mathbb{F}_q$ such that $\kappa(\lambda) = \lambda^2 = c_0$. Using again the fact that $\mathbb{F}_q$ has characteristic 2 , we conclude that $p(a) = a^2 - \lambda^2 = (a - \lambda)^2$ and hence $p(a)$ has a unique root (of multiplicity 2) in $\mathbb{F}_q$. $\square$

Even Predicate, $q = 2^i$, $k = 2q$. Since a check node in a random (d_v, d_c)-LDPC code has degree d_c or $d_c - 2$, we need to construct even and odd predicates for both arities d_c and $d_c - 2$ and over the same alphabet. We first construct an additional even predicate with arity $k = 2q$ based on *trivariate* linear forms.

Lemma 8.4.10. *Let q be a power of 2 and $k = 2q$. There exists a subgroup of $\mathbb{F}_q^k$ such that every element in the subgroup contains an even number of 0 coordinates.*

Proof. Our subgroup H' will be of the form $\{(\alpha x + \beta y + \gamma z)_{(x,y,z) \in E}\}_{\alpha, \beta, \gamma \in F_q}$, for some subset $E \subseteq \mathbb{F}_q^3$ of $2q = k$ evaluation points. The set $E \subseteq \mathbb{F}_q^3$ is given by

$$E := \{(1, a, a) : a \in \mathbb{F}_q\} \cup \{(0, b, b + 1) : b \in \mathbb{F}_q\}.$$

Clearly, $|E| = 2q$. The lemma follows from Claim 8.4.11 and Claim 8.4.12 below. $\square$

Claim 8.4.11. *Every trivariate linear form $(\alpha x + \beta y + \gamma z)$ has either 0, 2, q or $2q$ roots in E (which are all even integers).*

Proof. Let $\psi_{\alpha,\beta,\gamma}$ be a fixed trivariate $\mathbb{F}_q$-linear form, for some $\alpha, \beta, \gamma \in \mathbb{F}_q$. Let $E_1 := \{(1, a, a) : a \in \mathbb{F}_q\}$ and $E_2 := \{(0, b, b+1) : b \in \mathbb{F}_q\}$. We distinguish two cases:

- Case 1: $\beta + \gamma \neq 0$ in $\mathbb{F}_q$. Then, $\psi_{\alpha,\beta,\gamma}(1, a, a) = 0$ if and only if $a(\beta + \gamma) = -\alpha$, which is equivalent to $a = -(\beta + \gamma)^{-1}\alpha$. Hence, $\psi_{\alpha,\beta,\gamma}$ has exactly one root in E_1. Moreover, $\psi_{\alpha,\beta,\gamma}(0, b, b+1) = 0$ if and only if $b(\beta + \gamma) = -\gamma$, which is equivalent to $b = -(\beta + \gamma)^{-1}\alpha_3$. Hence, $\psi_{\alpha,\beta,\gamma}$ has exactly one root in E_2. So we conclude that in this case $\psi_{\alpha,\beta,\gamma}$ has exactly 2 roots in $E = E_1 \cup E_2$.

- Case 2: $\beta + \gamma = 0$ in $\mathbb{F}_q$. Then, $\psi_{\alpha,\beta,\gamma}(1, a, a) = 0$ if and only if $a(\beta + \gamma) = -\alpha$, which is equivalent to $\alpha_1 = 0$. Hence, $\psi_{\alpha,\beta,\gamma}$ has either 0 roots in E_1 (if $\alpha \neq 0$) or q roots in E_1 (if $\alpha = 0$). Moreover, $\psi_{\alpha,\beta,\gamma}(0, b, b+1) = 0$ if and only if $b(\beta+\gamma) = -\gamma$, which is equivalent to $\gamma = 0$. Hence, $\psi_{\alpha,\beta,\gamma}$ has either 0 roots in E_2 (if $\gamma \neq 0$) or q roots in E_2 (if $\gamma = 0$). So we conclude that in this case $\psi_{\alpha,\beta,\gamma}$ has either 0, q or $2q$ roots in $E = E_1 \cup E_2$.

$\square$

Claim 8.4.12. *H' is a balanced pairwise independent subgroup of $\mathbb{F}_q^k$.*

Proof. Applying Lemma 8.4.6 with $d = 3$, it is enough to show that any two distinct vectors in E are linearly-independent over $\mathbb{F}_q$. To show this, assume for the sake of contradiction that there exist $v_1 \neq v_2 \in \mathbb{F}_q$ and a scalar $\beta \in \mathbb{F}_q$ such that $v_2 = \beta v_1$. We distinguish three cases:

- $v_1, v_2 \in E_1$. Then, $v_1 = (1, a_1, a_1 + 1)$ and $v_2 = (1, a_2, a_2 + 1)$ for some $a_1 \neq a_2 \in \mathbb{F}_q$. Then, $v_2 = \beta v_1$ implies that $\beta = 1$ and hence $a_2 = a_1$, a contradiction.

- $v_1, v_2 \in E_2$. Then, $v_1 = (0, b_1, b_1+1)$ and $v_2 = (0, b_2, b_2+1)$ for some $b_1 \neq b_2 \in \mathbb{F}_q$. Then, $v_2 = \beta v_1$ implies that $\beta = 1$ and $b_1 = b_2$, a contradiction.

- $v_1 \in E_1$ and $v_2 \in E_2$. Then, $v_1 = (1, a, a)$ and $v_2 = (0, b, b + 1)$ for some $a, b \in \mathbb{F}_q$. Then, $v_2 = \beta v_1$ implies that $\beta = 0$ and hence that both $b = 0$ and $b + 1 = 0$, a contradiction.

$\square$

Type \ Arity	$q+1$	$2q$	$d_c - 2 = 3q+1$	$d_c = 3q+3$
Odd	Lem. 8.4.7 (H_1)		$H_1 \oplus H_3$	$H_1 \oplus H_1 \oplus H_1$
Even	Lem. 8.4.8 (H_2)	Lem. 8.4.10 (H_3)	$H_2 \oplus H_3$	$H_1 \oplus H_1 \oplus H_2$

Table 8.2: Subgroups for Lasserre

Direct sums of cosets of subgroups. For any $q = 2^i$, we constructed 3 cosets of subgroups: $H_1 \subseteq \mathbb{F}_q^{q+1}$ contained in the odd predicate, $H_2 \subseteq \mathbb{F}_q^{q+1}$ contained in the even predicate, $H_3 \subseteq \mathbb{F}_q^{2q}$ contained in the even predicate. Any direct sum of them gives a coset of a subgroup of $\mathbb{F}_q^k$ with k being the sum of the individual arities. If we add one coset contained in the even predicate and one contained in the odd predicate, the direct sum will be contained in the odd predicate. On the other hand, if we add two cosets that are contained in the same (even or odd) predicate, the direct sum will be contained in the even predicate. For $d_c = 3q+3$, we use such direct sums to construct the desired even and odd predicates for arities d_c and $d_c - 2$ as follows:

- $H_1 \oplus H_1 \oplus H_1$: A coset of a subgroup of $\mathbb{F}_q^{3q+3}$, contained in the odd predicate.
- $H_1 \oplus H_1 \oplus H_2$: A coset of a subgroup of $\mathbb{F}_q^{3q+3}$, contained in the even predicate.
- $H_1 \oplus H_3$: A coset of a subgroup of $\mathbb{F}_q^{3q+1}$, contained in the odd predicate.
- $H_2 \oplus H_3$: A coset of a subgroup of $\mathbb{F}_q^{3q+1}$, contained in the even predicate.

The constructed subgroups for the Lasserre hierarchy are summarized in Table 8.2.

Proof of Theorem 8.1.2. Consider a random (d_v, d_c)-LDPC code when $d_c = 3 \cdot 2^i + 3$ and fix $\delta = 1/8$, $q = 2^i = \frac{d_c - 3}{3}$. Lemmas 8.4.1 and 8.4.2 ensure that with high probability, each check-degree is either d_c or $d_c - 2$ and the code is $(\eta n, 1+\delta)$-expanding for some $\eta > 0$. For any received vector, let $(V, \mathcal{C})$ be the corresponding instance of NEAREST CODEWORD. Stretch the domain from $\{0,1\}$ to $G' := \mathbb{F}_q$. The above constructions show that for any $k \in \{d_c, d_c - 2\}$ and $type \in \{even, odd\}$, $P_{type} \subseteq (G')^k$ is a coset of a balanced pairwise independent subgroup. Theorem 8.3.2 gives a balanced solution to the $\frac{\eta n}{16}$ rounds of the Lasserre hierarchy that satisfies every constraint in the stretched instance. Lemma 8.3.1 transforms this solution to a $\frac{1}{q}$-biased solution to the same number of rounds for the original NEAREST CODEWORD instance.

8.5 Hypergraph Vertex Cover

The result for k-HYPERGRAPH VERTEX COVER will follow from the machinery and predicates that we constructed in Sections 8.3 and 8.4. We first restate Theorem 8.1.3.

Theorem 8.5.1 (Restatement of Theorem 8.1.3)**.** *Let $k = q + 1$ where q is any prime power. For any $\epsilon > 0$, there exist $\beta, \eta > 0$ (depending on k) such that a random k-uniform hypergraph with n vertices and $m = \beta n$ edges, simultaneously satisfies the following two conditions with high probability.*

- *The integral optimum of k-HYPERGRAPH VERTEX COVER is at least $(1 - \epsilon)n$.*
- *There is a solution to the ηn rounds of the Lasserre hierarchy of value $\frac{1}{k-1}n$.*

In the rest of this section, we prove Theorem 8.5.1. Fix k such that $q = k-1$ is a prime power. Given an instance of k-HYPERGRAPH VERTEX COVER, which is an instance of MIN ONES($\{P_\vee\}$), we stretch the domain from $\{0, 1\}$ to $\mathbb{F}_q$ by the map $\phi : \mathbb{F}_q \to \{0, 1\}$ with $\phi(0) = 1$, $\phi(g) = 0$ for $g \neq 0$. Then the corresponding predicate $P'_\vee \subseteq \mathbb{F}_q^k$ is a tuple of k elements from $\mathbb{F}_q^k$ that has at least one zero. We show that $P'_\vee$ contains a pairwise independent subgroup H' of $\mathbb{F}_q^k$. Indeed, we use the same H' that was used for the odd predicate for random LDPC codes, i.e., $H' := \{(\alpha x + \beta y)_{(x,y)\in E}\}_{\alpha,\beta\in F_q}$ where $E := \{(0, 1)\} \cup \{(1, a)\}_{a\in\mathbb{F}_q}$. In Section 8.4.2, we proved that H' is balanced pairwise independent and always has an odd number of zeros when k is odd. Here we allow k to be even so this is not true, but we still have that any element of H' has at least one zero (indeed, the only element in H that does not have exactly one zero is $(0, 0, \ldots, 0)$, which has k zeros). This constructs the desired predicate for $P'_\vee$. Given this predicate, the same technique of stretching the domain, constructing a Lasserre solution by Theorem 8.3.2, and collapsing back the domain using Lemma 8.3.1 gives a solution to the Lasserre hierarchy that is $\frac{1}{k-1}$-biased. Lemma 8.5.1 below, which ensures that random k-uniform hypergraphs have a large integral optimum and are highly expanding for some fixed number of hyperedges, concludes the proof of Theorem 8.5.1.

Lemma 8.5.1. *Let $k \geq 3$ be a positive integer and $\epsilon, \delta > 0$. There exists $\eta \leq \beta$ (depending on k) such that a random k-uniform hypergraph (V, E) with βn edges, where each edge e_i is sampled from $\binom{V}{k}$ with replacement, has the following properties with high probability.*

- *It is $(\eta n, k - 1 - \delta)$-expanding.*
- *Every subset of ϵn vertices contains a hyperedge. Therefore, the optimum of k-HYPERGRAPH VERTEX COVER is at least $(1 - \epsilon)n$.*

Proof. The proof uses standard probabilistic arguments and can be found in previous works [ACG$^+$10, Tul09]. Fix a subset $S \subseteq V$ of size ϵn. The probability that one hyperedge is contained in S is

$$\frac{\binom{\epsilon n}{k}}{\binom{n}{k}} \geq \frac{(\epsilon n/k)^k}{(en/k)^k} = (\epsilon/e)^k.$$

The probability that S does not contain any edge is at most

$$(1 - (\epsilon/e)^k)^{\beta n} \leq \exp(-(\epsilon/e)^k \beta n).$$

Since there are $\binom{n}{\epsilon n} \leq (e/\epsilon)^{\epsilon n} = \exp(\epsilon n(1 + \log(1/\epsilon)))$ choices for S, if $\beta > (e/\epsilon)^k$, with high probability, every subset of ϵn vertices contains a hyperedge.

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

$$\binom{n}{cs} \cdot \binom{\binom{cs}{k}}{s} \cdot s! \binom{\beta n}{s} \cdot \binom{n}{k}^{-s},$$

($\binom{n}{cs}$ for fixing variables to be covered, $\binom{\binom{cs}{k}}{s}$ for assigning them to s hyperedges, $s!\binom{\beta n}{s}$ for a set of s hyperedges) which is at most

$$(s/n)^{\delta s}(e^{2k+1-\delta}k^{1+\delta}\beta)^s \leq (s/n)^{\delta s}\beta^{5s} = (\frac{s\beta^{5/\delta}}{n})^{\delta s}.$$

By summing the probability over $s = 1, \ldots, \eta n$, the probability that it is not $(\eta n, k-1-\delta)$-expanding is

$$\begin{aligned}\sum_{s=1}^{\eta n}(\frac{s\beta^{5/\delta}}{n})^{\delta s} &= \sum_{s=1}^{\ln^2 n}(\frac{s\beta^{5/\delta}}{n})^{\delta s} + \sum_{s=\ln^2 n+1}(\frac{s\beta^{5/\delta}}{n})^{\delta s} \\ &\leq O(\frac{\beta^5}{n^\delta}\ln^2 n) + O((\eta \cdot \beta^{5/\delta})^{\delta \ln^2 n}).\end{aligned}$$

The first term is $o(1)$ for large n. The second term is also $o(1)$ for $\eta < 1/(\beta^{5/\delta})$. $\square$

8.6 Discussion

In this chapter, we showed that fairly powerful extensions of LP decoding, based on the Sherali-Adams and Lasserre hierarchies, fail to correct much more errors than the basic LP-decoder.

It would be interesting to prove analogous lower bound for LDPC codes whose adjacency graph has *large girth*. Such codes have been studied in some previous works on LP decoding, e.g., [Fel03] and [ADS12]. Note that the adjacency graph of random LDPC codes has small girth with high probability.

Finally, it would be very interesting to understand whether LP/SDP hierarchies can come close to capacity on *irregular* ensembles [RSU01] or on *spatially-coupled* codes [KRU12]. For the latter, some limitations of the base LP have been shown in [BGU14], but the performance of LP/SDP hierarchies remains open.

Part III

Coloring

Chapter 9

Coloring Overview

9.1 Introduction

Coloring (hyper)graphs is one of the most important and well-studied tasks in discrete mathematics and theoretical computer science. A K-uniform hypergraph $G = (V, E)$ is said to be χ-colorable if there exists a coloring $c : V \mapsto \{1, \ldots, \chi\}$ such that no hyperedge is monochromatic, and such a coloring c is referred to as a proper χ-coloring. Graph and hypergraph coloring has been the focus of active research in both fields, and has served as the benchmark for new research paradigms such as the probabilistic method (Lovász local lemma [EL75]) and semidefinite programming (Lovász theta function [Lov79]).

While such structural results are targeted towards special classes of hypergraphs, given a general χ-colorable K-uniform hypergraph, the problem of reconstructing a χ-coloring is known to be a hard task. Even assuming 2-colorability, reconstructing a proper 2-coloring is a classic NP-hard problem for $K \geq 3$. Given the intractability of proper 2-coloring, two notions of *approximate coloring* of 2-colorable hypergraphs have been studied in the literature of approximation algorithms. The first notion, called MIN COLORING, is to minimize the number of colors while still requiring that every hyperedge be non-monochromatic. The second notion, called MAX 2-COLORING allows only 2 colors, but the objective is to maximize the number of non-monochromatic hyperedges.[1]

Even with these relaxed objectives, the promise that the input hypergraph is 2-colorable seems grossly inadequate for polynomial time algorithms to exploit in a significant way. For MIN COLORING, given a 2-colorable K-uniform hypergraph, the best known algo-

[1]The maximization version is also known as MAX-SET-SPLITTING, or more specifically MAX k-SET-SPLITTING when considering K-uniform hypegraphs, in the literature.

rithm uses $O(n^{1-\frac{1}{K}})$ colors [CF96, AKMH96], which tends to the trivial upper bound n as K increases. This problem has been actively studied from the hardness side, motivating many new developments in constructions of probabilistically checkable proofs. Coloring 2-colorable hypergraphs with $O(1)$ colors was shown to be NP-hard for $K \geq 4$ in [GHS02] and $K = 3$ in [DRS05]. An exciting body of recent work has pushed the hardness beyond poly-logarithmic colors [DG13, GHH$^+$14, KS14b, Hua15]. In particular, [KS14b] shows quasi-NP-hardness of $2^{(\log n)^{\Omega(1)}}$-coloring a 2-colorable hypergraph (very recently the exponent was shown to approach $1/10$ in [Hua15]). This situation contrasts with graphs ($K = 2$) where it is not known to be hard to color 3-colorable graphs with just 5 colors unless we assume much stronger conjectures [DMR09].

The hardness results for MAX 2-COLORING show an even more pessimistic picture, wherein the naive *random assignment* (randomly give one of two colors to each vertex independently to leave a $(\frac{1}{2})^{K-1}$ fraction of hyperedges monochromatic in expectation), is shown to have the best guarantee for a polynomial time algorithm when $K \geq 4$ (see [Hås01]).

Given these strong intractability results, it is natural to consider what further relaxations of the objectives could lead to efficient algorithms. For maximization versions, Austrin and Håstad [AH13] proved that (almost[2]) 2-colorability is useless (in a formal sense that they define) for any Constraint Satisfaction Problem (CSP) that is a relaxation of 2-coloring [Wen14]. Therefore, it seems more natural to find a stronger promise on the hypergraph than mere 2-colorability that can be significantly exploited by polynomial time coloring algorithms for the objectives of MIN COLORING and MAX 2-COLORING. This motivates our main question *"how strong a promise on the input hypergraph is required for polynomial time algorithms to perform significantly better than naive algorithms for* MIN COLORING *and* MAX 2-COLORING*?"*

There is a very strong promise on K-uniform hypergraphs which makes the task of proper 2-coloring easy. If a hypergraph is K-partite (i.e., there is a K-coloring such that each hyperedge has each color exactly once), then one can properly 2-color the hypergraph in polynomial time. The same algorithm can be generalized to hypergraphs which admit a c-balanced coloring (i.e., c divides K and there is a K-coloring such that each hyperedge has each color exactly $\frac{K}{c}$ times). This can be seen by random hyperplane rounding of a simple SDP, or even simpler by solving a homogeneous linear system and iterating [Alo14], or by a random recoloring method analyzed using random walks [McD93]. In fact, a proper 2-coloring can be efficiently achieved assuming that the hypergraph admits a fair partial 2-coloring, namely a pair of disjoint subsets A and B of the vertices such

[2]We say a hypergraph is *almost* χ-colorable for a small constant $\epsilon > 0$, there is a χ-coloring that leaves at most ϵ fraction of hyperedges monochromatic.

that for every hyperedge e, $|e \cap A| = |e \cap B| > 0$ [McD93].

The promises on structured colorings that we consider in this part are natural relaxations of the above strong promise of a perfectly balanced (partial) coloring.

- A hypergraph is said to have *discrepancy* ℓ when there is a 2-coloring such that in each hyperedge, the difference between the number of vertices of each color is at most ℓ.
- A χ-coloring $(\chi \leq K)$ is called *rainbow* if every hyperedge contains each color at least once.
- A χ-coloring $(\chi \geq K)$ is called *strong* if every hyperedge contains K different colors.

These three notions are interesting in their own right, and have been independently studied. Discrepancy minimization has recently seen different algorithmic ideas [Ban10, LM12, Rot14] to give constructive proofs of the classic six standard deviations result of Spencer [Spe85]. Rainbow coloring admits a natural interpretation as a partition of V into the maximum number of disjoint vertex covers, and has been actively studied for geometric hypergraphs due to its applications in sensor networks [BPRS13]. Strong coloring is closely related to graph coloring by definition, and is known to capture various other notions of coloring [AH05]. It is easy to see that ℓ-discrepancy ($\ell < K$), χ-rainbow colorability ($2 \leq \chi \leq K$), and χ-strong colorability $(K \leq \chi \leq 2K - 2)$ all imply 2-colorability. For odd K, both $(K+1)$-strong colorability and $(K-1)$-rainbow colorability imply discrepancy-1, so strong colorability and rainbow colorability seem stronger than low discrepancy.

9.2 Our Results

We study both algorithmic and hardness results for MIN COLORING and MAX 2-COLORING under these strong promises.

9.2.1 MIN COLORING

For MIN COLORING, we show that all three promises lead to a better coloring.

Theorem 9.2.1. *Consider any K-uniform hypergraph $H = (V, E)$ with n vertices and m edges. For any $\ell < O(\sqrt{K})$, If H has discrepancy-ℓ, $(K - \ell)$-rainbow colorable, or $(K + \ell)$-strong colorable, one can color H with $\tilde{O}((\frac{m}{n})^{\frac{\ell^2}{K^2}}) \leq \tilde{O}(n^{\frac{\ell^2}{K}})$ colors.*

These results significantly improve the current best $\tilde{O}(n^{1-\frac{1}{K}})$ colors that assumes only 2-colorability. Our techniques give slightly better results depending on the promise — see Theorem 11.3.1.

We also show strong hardness results. Given a hypergraph $H = (V, E)$ and a subset $I \subseteq V$, we say that I *induces* a hyperedge e when $e \subseteq I$. The hypergraph induced by I is (I, E_I) where E_I is the set of hyperedges induced by I. The following is our main theorem. Note that in any result in this section that guarantees a coloring with some desired properties in the completeness case, each color contains the same fraction of vertices.

Theorem 9.2.1. *For any $\epsilon > 0$ and $Q, k \geq 2$, given a Qk-uniform hypergraph $H = (V, E)$, it is NP-hard to distinguish the following cases.*

- *Completeness: There is a k-coloring $c : V \to [k]$ such that for every hyperedge $e \in E$ and color $i \in [k]$, e has at least $Q - 1$ vertices of color i.*
- *Soundness: Every $I \subseteq V$ of measure ϵ induces at least a fraction $\epsilon^{O_{Q,k}(1)}$ of hyperedges. In particular, there is no independent set of measure ϵ, and every $\lfloor \frac{1}{\epsilon} \rfloor$-coloring of H induces a monochromatic hyperedge.*

Fixing $Q = 2$ gives a hardness of rainbow coloring with K optimized to be $2k$.

Corollary 9.2.2. *For all integers $c, k \geq 2$, given a $2k$-uniform hypergraph H, it is NP-hard to distinguish whether H is rainbow k-colorable or is not even c-colorable.*

On the other hand, fixing $k = 2$ gives a strong hardness result of discrepancy minimization (with 2 colors). A coloring is said to have discrepancy Δ when in each hyperedge, the difference between the maximum and the minimum number of occurrences of a single color is at most Δ.

Corollary 9.2.3. *For any $c, Q \geq 2$, given a $2Q$-uniform hypergraph $H = (V, E)$, it is NP-hard to distinguish whether H is 2-colorable with discrepancy 2 or is not even c-colorable.*

The above result strengthens the result of Austrin et al [AGH14] that shows hardness of 2-coloring in the soundness case. However, their result also holds in $(2Q + 1)$-uniform hypergraphs with discrepancy 1, which is not covered by the results in this thesis.

The algorithmic and hardness results of highly structured hypergraphs are summarized in Table 9.1. It is worth emphasizing that prior to this work, even hardness of 2-coloring a rainbow 3-colorable hypergraph was not known. Indeed such a result seemed out of reach of the sort of Fourier-based PCP techniques used for hardness of hypergraph coloring in [GHS02] and follow-ups. In this part we leverage invariance principle based techniques to analyze test distributions that ensure balanced rainbow colorability.One of our contributions is to distill a general recipe for combining test distributions with suitable outer PCPs (various forms of smooth LABEL COVER) to establish such inapproximability results. This makes our approach quite flexible and can also be readily applied to several other problems as described in Section 9.2.1.

Promised Coloring Structure	Algorithm	Hardness
K-partite	2-colorable	Not rainbow K-colorable (Almost/UG) Not weak $O(1)$-colorable
Rainbow $(K-1)$-colorable	$\tilde{O}(n^{1/K})$**-colorable**	**(Almost) Not weak** $O(1)$**-colorable**
Rainbow $\frac{K}{2}$-colorable		**Not weak** $O(1)$**-colorable**
2-colorable with perfect balance	2-colorable	
2-colorable with discrepancy 1	$\tilde{O}(n^{1/K})$**-colorable**	Not 2-colorable
2-colorable with discrepancy 2	$\tilde{O}(n^{4/K})$**-colorable**	**Not weak** $O(1)$**-colorable**

Table 9.1: Summary of algorithmic and hardness results for MIN COLORING a highly structured K-uniform hypergraph. Almost means that $\epsilon > 0$ fraction of vertices and incident hyperedges must be deleted to have the structure. UG indicates that the result is based on the Unique Games Conjecture. The results of this thesis are in boldface.

HYPERGRAPH VERTEX COVER. Rainbow k-coloring has a tight connection to HYPERGRAPH VERTEX COVER, because it partitions the set of vertices into k disjoint vertex covers. In particular, Corollary 9.2.2 implies that K-HYPERGRAPH VERTEX COVER is NP-hard to approximate within a factor of $(\frac{K}{2} - \epsilon)$, but the better inapproximability factor of $(K - 1 - \epsilon)$ is already established by the classical result of Dinur et al [DGKR05]. We give the first analytic proof of the same theorem, with two slight improvements: the size of the minimum vertex cover in the completeness case is improved to $\frac{1}{K-1}$ from $(\frac{1}{K-1}+\epsilon)$, and in the soundness case every set of measure ϵ induces $\epsilon^{O_K(1)}$ fraction of hyperedges.

Theorem 9.2.4. *For any $\epsilon > 0$ and $K \geq 3$, given a K-uniform hypergraph $H = (V, E)$, it is NP-hard to distinguish the following cases.*

- *Completeness: There is a vertex cover of measure $\frac{1}{K-1}$.*

- *Soundness: Every $I \subseteq V$ of measure ϵ induces at least a fraction $\epsilon^{O_K(1)}$ of hyperedges.*

Bansal and Khot [BK10] and Sachdeva and Saket [SS13] focused on *almost* rainbow k-colorable hypergraphs (where one is allowed to remove a small fraction of vertices and all incident hyperedges to ensure rainbow colorability) to show hardness of scheduling problems. This notion allows us to prove the following more structured hardness as well as $(K-1-\epsilon)$-inapproximability for HYPERGRAPH VERTEX COVER. It improves [SS13] in the number of colors used, and almost matches [BK10] which is based on the Unique Games Conjecture.

Theorem 9.2.5. *For any $\epsilon > 0$ and $K \geq 3$, given a K-uniform hypergraph $H = (V, E)$, it is NP-hard to distinguish the following cases.*

- *Completeness: There exists $V^* \subseteq V$ of measure ϵ and a coloring $c : [V \setminus V^*] \to [K-1]$ such that for every hyperedge of the induced hypergraph on $V \setminus V^*$, $K-2$ colors appear once and the other color twice. Therefore, H has a vertex cover of size at most $\frac{1}{K-1} + \epsilon$.*

- *Soundness: There is no independent set of measure ϵ.*

MAX Q-OUT-OF-$(2Q+1)$-SAT. MAX Q-OUT-OF-$(2Q+1)$-SAT refers to the problem of finding a satisfying assignment in a $(2Q+1)$-CNF formula, given the promise that some assignment makes each clause have at least Q true literals. We give an analytic proof following our recipe of the following result, which was first established based on simpler combinatorial techniques in Austrin et al [AGH14].

Theorem 9.2.6. *For $Q \geq 2$, there exists $\epsilon > 0$ depending on Q such that given a $(2Q+1)$-CNF formula, it is NP-hard to distinguish the following cases.*

- *Completeness: There is an assignment such that each clause has at least Q true literals.*

- *Soundness: No assignment can satisfy more than a fraction $(1-\epsilon)$ of clauses.*[3]

[3]An explicit value of ϵ as a function of Q in the soundness is $\exp(-O(Q \log Q))$, which is better than the value $\exp(-O(Q^c))$ for some large absolute constant c implicit in the proof of [AGH14].

9.2.2 MAX 2-COLORING

Our algorithmic result for MAX 2-COLORING proves that our three promises, unlike mere 2-colorability, give enough structure for polynomial time algorithms to perform significantly better than naive algorithms. We also study these promises from a hardness perspective to understand the asymptotic threshold at which beating naive algorithms goes from easy to UG/NP-Hard. In particular assuming the UGC, for MAX 2-COLORING under ℓ-discrepancy or $(K-\ell)$-rainbow colorability, this threshold is $\ell = \Theta(\sqrt{K})$.

Theorem 9.2.2. *There is a randomized polynomial time algorithm that produces a 2-coloring of a K-uniform hypergraph H with the following guarantee. For any $0 < \epsilon < \frac{1}{2}$ (let $\ell = K^{\epsilon}$), there exists a constant $\eta > 0$ such that if H is $(K-\ell)$-rainbow colorable or $(K+\ell)$-strong colorable, the fraction of monochromatic edges in the produced 2-coloring is $O((\frac{1}{K})^{\eta K})$ in expectation.*

Our results indeed show that this algorithm significantly outperforms the random assignment even when ℓ approaches $\sqrt{K}$ asymptotically. See Theorem 11.2.2 and Theorem 11.2.3 for the precise statements.

For the ℓ-discrepancy case, note that our results on SYMMETRIC CSP (Theorem 5.1.1 and (5.1)) show that when $\ell < \sqrt{K}$, there exists an approximation algorithm that marginally (by an additive factor much less than 2^{-K}) outperforms the random assignment.

The following hardness results suggest that this gap between low-discrepancy and rainbow/strong colorability might be intrinsic.

Theorem 9.2.3. *For sufficiently large odd K, given a K-uniform hypergraph which admits a 2-coloring with at most a $(\frac{1}{2})^{6K}$ fraction of edges of discrepancy larger than 1, it is UG-hard to find a 2-coloring with a $(\frac{1}{2})^{5K}$ fraction of monochromatic edges.*

Theorem 9.2.4. *For even $K \geq 4$, given a K-uniform hypergraph which admits a 2-coloring with no edge of discrepancy larger than 2, it is NP-hard to find a 2-coloring with a $K^{-O(K)}$ fraction of monochromatic edges.*

Theorem 9.2.5. *For K sufficiently large, given a K-uniform hypergraph which admits a 2-coloring with no edge of discrepancy larger than $O(\log K)$, it is NP-hard to find a 2-coloring with a $2^{-O(K)}$ fraction of monochromatic edges.*

Theorem 9.2.6. *For K such that $\chi := K - \sqrt{K}$ is an integer greater than 1, and any $\epsilon > 0$, given a K-uniform hypergraph which admits a χ-coloring with at most ϵ fraction of non-rainbow edges, it is UG-hard to find a 2-coloring with a $(\frac{1}{2})^{K-1}$ fraction of monochromatic edges.*

Table 9.2 summarizes our results for MAX 2-COLORING.

Promises	Algorithm	Hardness
ℓ-Discrepancy	$1-(1/2)^{K-1}+\delta, \quad \ell<\sqrt{K}$	UG: $1-(1/2)^{5K}, \quad \ell=1$ NP: $1-(1/K)^{O(K)}, \quad \ell=2$ NP: $1-(1/2)^{O(K)}, \quad \ell=\Omega(\log K)$ UG: $1-(1/2)^{K-1}, \quad \ell\geq\sqrt{K}$
$(K-\ell)$-Rainbow	$1-(1/K)^{\Omega(K)}, \quad \ell=o(K)$	UG: $1-(1/2)^{K-1}, \quad \ell\geq\sqrt{K}$
$(K+\ell)$-Strong	$1-(1/K)^{\Omega(K)}, \quad \ell=o(K).$	

Table 9.2: Summary of our algorithmic and hardness results for MAX 2-COLORING with valid ranges of ℓ.

9.3 Related Work

The work of Austrin et al. [AGH14] shows NP-hardness of finding a proper 2-coloring under the discrepancy-1 promise. The work of Bansal and Khot [BK10] shows hardness of $O(1)$-coloring even when the input hypergraph is promised to be almost K-partite (under the Unique Games Conjecture). Our work is inspired by recent developments concerning the inapproximability of HYPERGRAPH VERTEX COVER and the Constraint Satisfaction Problem (CSP). At a high level, Theorem 9.2.1 looks similar to the result of Sachdeva and Saket [SS13] who proved almost the same statement *without perfect completeness* — we need to delete $\epsilon > 0$ fraction of vertices and all incident hyperedges to have a similar guarantee in the completeness case. Achieving perfect completeness is a nontrivial task, as manifested in K-CSP — approximating a $(1-\epsilon)$-satisfiable instance of K-CSP is NP-hard within a factor of $\frac{2K}{2^K}$ [Cha13], while the best inapproximability factor for perfectly satisfiable K-CSP is $\frac{2^{O(K^{1/3})}}{2^K}$ [Hua13].

In CSP, significant research efforts have been made for proving every predicate strictly dominating parity is *approximation resistant* (i.e., no efficient algorithm can beat the ratio achieved by simply picking a random assignment) even on satisfiable instances. O'donnell and Wu [OW09] proved this assuming the d-to-1 conjecture for $K = 3$, and recently this was proven to be true assuming only $\mathbf{P} \neq \mathbf{NP}$ by Håstad ($K = 3$, [Hås14]) and Wenner ($K \geq 4$, [Wen13]). Many of these works are based on invariance principle based techniques, and it is natural to ask whether they let us to achieve perfect completeness in Hypergraph Coloring as well. To the best of our knowledge, our work is the first to apply invariance based techniques to prove NP-hardness of Hypergraph Coloring / Vertex Cover problems (Khot and Saket [KS14a] used them to prove hardness of finding an independent set in 2-colorable 3-uniform hypergraphs, assuming the d-to-1 conjecture).

Fourier-analytic proofs of harndess of K-HYPERGRAPH VERTEX COVER are known for small K [GHS02, Hol02, Kho02a, Sak14]. Even though they cannot be easily generalized to large K, the recent work of Saket [Sak14] for $K = 4$ uses general *reverse hypercontractivity* studied by Mossel et al. [Mos10], and we extend his result to present a framework to study general K-uniform hypergraphs. This generalized reverse hypercontractivity might have other applications in hardness of approximation or in other areas of theoretical computer science. In the rest of the section, for simplicity of illustration we fix $Q = k = 2$ (so that the test distribution becomes that of [Sak14]) and give a high level glimpse into our proof strategy.

9.4 Organization

Chapter 10 presents our hardness results for MIN COLORING and MAX 2-COLORING. Section 10.1 introduces our *recipe* and hardness techniques in a simpler setting. Section 10.2 contains mathematical tools in that chapter, including Fourier analysis for *blocked functions* and variants of LABEL COVER. Our main technical tool, generalized reverse hypercontractivity, is introduced in Section 10.3. Section 10.4 proves the main Theorem 9.2.1. In Section 10.6, we show the versatility of our approach by proving Theorem 9.2.4, and 9.2.6, using the same procedure. Section 10.7 proves Theorem 9.2.3, Theorem 9.2.4, Theorem 9.2.5, and Theorem 9.2.6 for MAX 2-COLORING.

Chapter 11 presents our algorithmic results for MIN COLORING and MAX 2-COLORING. Section 11.1 introduces our techniques, and Section 11.2 and Section 11.3 study MAX 2-COLORING and MIN COLORING respectively.

Chapter 10

Hardness of Coloring

10.1 Techniques

We illustrate our main ideas and general recipe in a simple setting.

10.1.1 Techniques

We reduce LABEL COVER to 4-uniform hypergraph coloring. Given a LABEL COVER instance based on a bipartite graph $G = (U \cup V, E)$ with projections $\pi_e : [R] \to [L]$ (see Section 10.2 for the formal definition), let U be the *small side* and V be the *big side*. Let $\Omega = \{1, 2\}$. Our hypergraph $H = (V', E')$ is defined by $V' := V \times \Omega^R$, and E' is described by the following procedure to sample a hyperedge.

- Sample $u \in U$ and its neighbors $v, w \in V$.
- Sample $x_1, x_2, y_1, y_2 \in \Omega^R$ as the following: for $1 \le i \le L$,
  - With probability half, $(x_1)_{\pi^{-1}_{(u,v)}(i)}, (x_2)_{\pi^{-1}_{(u,v)}(i)}, (y_1)_{\pi^{-1}_{(u,w)}(i)}$ are sampled i.i.d., but $(y_2)_j = 3 - (y_1)_j$ for every $j \in \pi^{-1}_{(u,w)}(i)$.
  - With probability half, $(y_1)_{\pi^{-1}_{(u,w)}(i)}, (y_2)_{\pi^{-1}_{(u,w)}(i)}, (x_1)_{\pi^{-1}_{(u,v)}(i)}$ are sampled i.i.d., but $(x_2)_j = 3 - (x_1)_j$ for every $j \in \pi^{-1}_{(u,v)}(i)$.
- Output a hyperedge containing four vertices $(v, x_1), (v, x_2), (w, y_1), (w, y_2)$.

Completeness is obvious from the above distribution. For each *block* that corresponds to $\pi^{-1}_{(u,v)}(i)$ or $\pi^{-1}_{(u,w)}(i)$, one of (x_1, x_2) and (y_1, y_2) is allowed to be sampled independently, but the other pair has to satisfy that two points are different in every coordinate in that block.

For soundness, let I be a large independent set, let $f_v : \Omega^R \to \{0, 1\}$ be the indicator function of $I \cap (\{v\} \times [k]^R)$. Then I satisfies the following two properties.

$$\mathbb{E}_{v,x_1}[f_v(x_1)] \gg 0, \qquad \mathbb{E}_{u,v,w}\mathbb{E}_{x_1,x_2,y_1,y_2}[f_v(x_1)f_v(x_2)f_w(y_1)f_w(y_2)] = 0.$$

These two properties seem to be contrary for randomly chosen I, so I with the above two properties should exploit some structure of the reduction. We prove that the existence of such I leads to a good decoding strategy to the LABEL COVER instance. This implies that there is no large independent set if the LABEL COVER does not admit a good labeling.

Dealing with noise and influences. Before proceeding to the analysis, we discuss two issues that highlight technical difficulties in proving NP-hardness (as opposed to Unique Games-hardness) of coloring with perfect completeness (as opposed to imperfect completeness) in terms of noise.

Strong vs Weak Noise. Given a function $f : \Omega^R \to [0, 1]$, consider the *noise operator* $T_{1-\gamma}$ defined by $T_{1-\gamma}f(x) = \mathbb{E}_y[f(y)|x]$ where y resamples each coordinate of x with probability γ. It is central to most decoding strategies that we actually analyze noised functions $T_{1-\gamma}f_v$ and $T_{1-\gamma}f_w$ instead of the original functions. We call the step of passing from the original functions to the noised functions *strong noise*. The easiest way to give strong noise is to explicitly include it in the test distribution, independently for all points. However, such explicit and strong noise breaks perfect completeness, since all points might be noised together and we cannot control the behavior.

To deal with this issue, we call *weak noise* to be a property inherent in the test distribution, bounding the correlation between the points we sample. In the test distribution we gave above, it refers to sampling exactly one of (x_1, x_2) or (y_1, y_2) completely independently (for each block). The fact that only one pair is noised is not strong enough to be directly applicable to decoding, but the bounded correlation allows us to apply the result of Mossel [Mos10] to show that the expected value of the product does not change much we replace each f by the noised version only for the sake of analysis. This idea of *smoothing a function in the analysis* allows us maintain perfect completeness.

Block Noise, Block Influence. Consider the projections $\pi_{(u,v)}, \pi_{(u,w)} : [R] \to [L]$. Let $d > 1$ be the degree of the projections. d coordinates of x_1, x_2 and d coordinates of y_1, y_2 must be treated in the same *block* which is often regarded as one coordinate.

The aforementioned result of Mossel in fact shows that we can replace f by $\overline{T}_{1-\gamma}f$, where $\overline{T}_{1-\gamma}$ is the *block noise* operator when we view each block as one coordinate. This is not strong enough for our decoding strategy, but the idea of Wenner [Wen13] lets us to replace $\overline{T}_{1-\gamma}f$ by the *individually noised* function $T_{1-\gamma}f$ if f almost depends on only *shattered* parts (roughly, shattered parts of a function under a projection do not distinguish whether the projection is 1-to-1 or not). This shattering behavior can be achieved by Smooth LABEL COVER defined by Khot [Kho02a].

At the end of analysis, our invariance principle will show that

$$\sum_{1 \le i \le L} \overline{\mathsf{Inf}}_i[T_{1-\gamma}f_v]\overline{\mathsf{Inf}}_i[T_{1-\gamma}f_w]$$

is large where $\overline{\mathsf{Inf}}$ indicates the influence when we view each block as one coordinate. It turns out to suffice to deal with these *block noises*, since they appear only in the analysis of the decoding; our decoding procedure itself does not depend on the projections, and the goal of the decoding is to have two vertices output the coordinates in the same block. To summarize, we put an effort to pass from block noise to individual noise in the beginning of our analysis, but we keep block influence to the end of analysis where it is naturally integrated with the decoding.

Recipe. We briefly discuss the five main steps in the soundness analysis and how they relate to each other. We view distilling and clearly articulating this recipe and highlighting its versatility also as one of the contributions of this thesis.

1. Fixing a good pair: Given an independent set I of measure ϵ, using smoothness of LABEL COVER, we show that in the original instance of LABEL COVER, there is a large fraction $u \in U$ and its neighbors $v, w \in V$ with the following properties. $\mathbb{E}[f_v], \mathbb{E}[f_w] \ge \frac{\epsilon}{2}$, and they almost depend on shattered parts. In the subsequent steps, we fix such u, v, w and analyze the probability that either (u, v) or (u, w) is satisfied by our decoding strategy.

2. Lower bounding in each hypercube: In Theorem 10.3.8, we show

$$\mathbb{E}[f_v(x_1)f_v(x_2)], \mathbb{E}[f_w(y_1)f_w(y_2)] \ge \zeta(\epsilon) > 0.$$

It uses *reverse hypercontractivity* [MOR$^+$06, MOS13], which is discussed in Section 10.3. Roughly, it says the noise operator T_ρ increases q-norm $\|T_\rho f\|_q$ when $q < 1$, so that $\|Tf\|_q \ge \|f\|_p$ for some $q < p < 1$ depending on ρ (note that $\|f\|_q \le \|f\|_p$). The case $k = 2$ follows directly from the previous result, but for larger k we generalize the reverse hypercontractivity to more general operators, even between different spaces. This step does not depend on noise or the degree of projections (e.g. the same ζ works for $T_{1-\gamma}f$ and $\overline{T}_{1-\gamma}f$).

3. Smoothing functions (based on 1.): Based on the bounded correlation of the test distribution, we use the result of Mossel [Mos10] to pass from f to $\overline{T}_{1-\gamma}f$. The fact that f_v, f_w almost depend on shattered parts allows us to use Theorem 10.4.5 to pass from $\overline{T}_{1-\gamma}f$ to $T_{1-\gamma}f$. Therefore we have

$$\begin{aligned}&\mathbb{E}_{x_1,x_2,y_1,y_2}[f_v(x_1)f_v(x_2)f_w(y_1)f_w(y_2)]\\ \approx\ &\mathbb{E}_{x_1,x_2,y_1,y_2}[T_{1-\gamma}f_v(x_1)T_{1-\gamma}f_v(x_2)T_{1-\gamma}f_w(y_1)T_{1-\gamma}f_w(y_2)].\end{aligned}$$

For simplicity, let $f' = T_{1-\gamma}f$.

4. Invariance (based on 2. and 3.): Since I is independent, the above results imply

$$\begin{aligned}0 &\approx \mathbb{E}_{x_1,x_2,y_1,y_2}[f'_v(x_1)f'_v(x_2)f'_w(y_1)f'_w(y_2)] \ll \zeta^2\\ &\leq \mathbb{E}_{x_1,x_2}[f'_v(x_1)f'_v(x_2)]\mathbb{E}_{y_1,y_2}[f'_w(y_1)f'_w(y_2)].\end{aligned}$$

In Theorem 10.4.7, we use an invariance principle inspired by that of Wenner [Wen13] and Chan [Cha13] to conclude that $\sum_{1\leq i\leq L}\overline{\mathsf{Inf}}_i[f'_v]\overline{\mathsf{Inf}}_i[f'_w] \geq \tau$, which implies that there are matching (blocks of) influential coordinates. The crucial property we used is that x_i is independent of (y_1, y_2) — one point is independent of the joint distribution of the points not in the same hypercube.

5. Decoding Strategy (based on 3. and 4.): The standard decoding strategy based on Fourier coefficients of f shows that either (u, v) or (u, w) will be satisfied with good probability. As previously discussed, $\sum_{1\leq i\leq L}\overline{\mathsf{Inf}}_i[f'_v]\overline{\mathsf{Inf}}_i[f'_w] \geq \tau$ gives large common block influences of individually noised functions, and they are sufficient for the decoding.

10.2 Preliminaries

For a positive integer k, let $[k] := \{1, 2, \ldots, k\}$. Let $\mathbb{S}_k$ be the set of k-permutations — $(x_1, \ldots, x_k) \in [k]^k$ such that $x_i \neq x_j$ for all $i \neq j$. For a vector $x \in \mathbb{R}^m$ and $S \subseteq [m]$, x_S denotes the projection of x onto the coordinates in S. Definitions and simple properties introduced from Section 10.2.1 to Section 10.2.4 are from Mossel [Mos10].

10.2.1 Correlated Spaces

Given a probability space (Ω, μ) (we always consider finite probability spaces), let $\mathcal{L}(\Omega)$ be the set of functions $\{f : \Omega \to \mathbb{R}\}$ and for an interval $I \subseteq \mathbb{R}$, $\mathcal{L}_I(\Omega)$ be the set of functions

$\{f : \Omega \to I\}$. A collection of probability spaces are said to be correlated if there is a joint probability distribution on them. We will denote k correlated spaces $\Omega_1, \ldots, \Omega_k$ with a joint distribution μ as $(\Omega_1 \times \cdots \times \Omega_k; \mu)$. Note that the definition of correlated spaces includes the joint distribution. Two instantiations of correlated spaces, even though they are defined on the same underlying sets, are considered different when their distributions are not the same.

Given two correlated spaces $(\Omega_1 \times \Omega_2, \mu)$, we define the correlation between Ω_1 and Ω_2 by

$$\rho(\Omega_1, \Omega_2; \mu) := \sup \left\{ \mathsf{Cov}[f, g] : f \in \mathcal{L}(\Omega_1), g \in \mathcal{L}(\Omega_2), \mathsf{Var}[f] = \mathsf{Var}[g] = 1 \right\}.$$

The following lemma of Wenner [Wen13] gives a convenient way to bound the correlation.

Lemma 10.2.1 (Corollary 2.18 of [Wen13])**.** *Let $(\Omega_1 \times \Omega_2, \mu)$ and $(\Omega_1 \times \Omega_2, \mu')$ be two distinct instantiations of correlated spaces such that the marginal distribution of at least one of Ω_1 and Ω_2 is identical on μ and μ'. For any $0 \le \delta \le 1$, consider another correlated spaces $(\Omega_1 \times \Omega_2, \delta\mu + (1-\delta)\mu')$. Then,*

$$\rho(\Omega_1, \Omega_2; \delta\mu + (1-\delta)\mu') \le \sqrt{\delta \cdot \rho(\Omega_1, \Omega_2; \mu)^2 + (1-\delta) \cdot \rho(\Omega_1, \Omega_2; \mu')^2}.$$

Given k correlated spaces $(\Omega_1 \times \cdots \times \Omega_k, \mu)$, we define the correlation of these spaces by

$$\rho(\Omega_1, \ldots, \Omega_k; \mu) := \max_{1 \le i \le k} \rho\Big(\prod_{1 \le j \le i-1} \Omega_j \times \prod_{i+1 \le j \le k} \Omega_j, \Omega_i; \mu\Big).$$

10.2.2 Operators

Let $(\Omega_1 \times \Omega_2, \mu)$ be two correlated spaces. The *Markov operator* associated with them is the operator mapping $f \in \mathcal{L}(\Omega_1)$ to $Tf \in \mathcal{L}(\Omega_2)$ by

$$(Tf)(y') = \mathbb{E}_{(x,y)\sim\mu}[f(x) | y = y'].$$

The *noise operator* or *Bonami-Beckner operator* T_ρ $(0 \le \rho \le 1)$ associated with a single probability space (Ω, μ) is the Markov operator associated with $(\Omega \times \Omega, \nu)$, where $\nu(x, y) = (1-\rho)\mu(x)\mu(y) + \rho\mathbb{I}[x = y]\mu(x)$ and $\mathbb{I}[\cdot]$ is the indicator function — ν samples (x, y) independently with probability $1 - \rho$, and samples $x = y$ with probability ρ. Note that $T_\rho f(y) = \rho f(y) + (1-\rho)\mathbb{E}_\mu[f(x)]$.

10.2.3 Functions and Influences

Let (Ω, μ) be a probability space. Given a function $f \in \mathcal{L}(\Omega)$ and $p \in \mathbb{R}$, let $\|f\|_p := \mathbb{E}_{x \sim \mu}[|f(x)|^p]^{1/p}$. We also use $\|f\|_{p,\mu}$ for the same quantity if it is instructive to emphasize μ. We note that $\|f\|_p$ for $p < 0$ is also used throughout the paper, but in this case we ensure that $f > 0$. For $f, g \in \mathcal{L}(\Omega)$, $\langle f, g \rangle := \mathbb{E}_{x \sim \mu}[f(x)g(x)]$.

Consider a product space $(\Omega^R, \mu^{\otimes R})$ and $f \in \mathcal{L}(\Omega^R)$. The *Efron-Stein decomposition* of f is given by

$$f(x_1, \dots, x_R) = \sum_{S \subseteq [R]} f_S(x_S)$$

where (1) f_S depends only on x_S and (2) for all $S \not\subseteq S'$ and all $x_{S'}$, $\mathbb{E}_{x' \sim \mu^{\otimes R}}[f_S(x') | x'_{S'} = x_{S'}] = 0$.

The *influence* of the jth coordinate on f is defined by

$$\mathsf{Inf}_j[f] := \mathbb{E}_{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_R}[\mathsf{Var}_{x_j}[f(x_1, \dots, x_R)]].$$

Given the noise operator T_ρ for (Ω, μ), we let $T_\rho^{\otimes R}$ be the noise operator for $(\Omega^R, \mu^{\otimes R})$ (i.e. noising each coordinate independently) and call it T_ρ. The noise operator and the influence has a convenient expression in terms of the Efron-Stein decomposition.

$$T_\rho[f] = \sum_S \rho^{|S|} f_S \,; \qquad \mathsf{Inf}_j[f] = \| \sum_{S : j \in S} f_S \|_2^2 = \sum_{S : j \in S} \|f_S\|_2^2$$

The following lemma lets us to reason about the influences of the product of functions.

Lemma 10.2.2 ([ST09]). *Let $(\Omega_1 \times \cdots \times \Omega_k, \mu)$ be k probability spaces and $(\Omega_1^L \times \cdots \times \Omega_k^L, \mu^{\otimes L})$ be the corresponding product spaces. Let $f_i \in \mathcal{L}_{[-1,1]}(\Omega_i^L)$, and $F \in \mathcal{L}_{[-1,1]}(\Omega_1^L \times \cdots \times \Omega_k^L)$ such that $F(x_1, \dots, x_k) = \prod_{1 \le i \le k} f_i(x_i)$. Then for $1 \le j \le L$, $\mathsf{Inf}_j(F) \le k \sum_{i=1}^k \mathsf{Inf}_j(f_i)$.*

Proof. We use $(x_i)_{-j} \in (\Omega_i)^{L-1}$ to denote x_i except the jth coordinate.

$$\begin{aligned}
\mathsf{Inf}_j(F) =& \mathbb{E}_{[(x_1)_{-j},\ldots,(x_k)_{-j}]}\mathbb{E}_{[(x_1)_j,\ldots,(x_k)_j,(x_1')_j,\ldots,(x_k')_j]}[(F(x_1,\ldots,x_k) - F(x_1',\ldots,x_k'))^2] \\
=& \mathbb{E}_{[(x_1)_{-j},\ldots,(x_k)_{-j}]}\mathbb{E}_{[(x_1)_j,\ldots,(x_k)_j,(x_1')_j,\ldots,(x_k')_j]}[(\prod_i f_i(x_i) - \prod_i f_i(x_i'))^2] \\
\le& k \sum_i \mathbb{E}_{[(x_1)_{-j},\ldots,(x_k)_{-j}]}\mathbb{E}_{[(x_1)_j,\ldots,(x_k)_j,(x_1')_j,\ldots,(x_k')_j]}[(f_i(x_i) - f_i(x_i'))^2] \\
=& k \sum_i \mathbb{E}_{[(x_i)_{-j}]}\mathbb{E}_{[(x_i)_j,(x_i')_j]}[(f_i(x_i) - f_i(x_i'))^2] \\
=& k \sum_i \mathsf{Inf}_j(f_i)
\end{aligned}$$

where the inequality follows from the fact that

$$\forall a_1,\ldots,a_k,b_1,\ldots,b_k \in [-1,1] : (\prod_i a_i - \prod_i b_i)^2 \le k \cdot \sum_i (a_i - b_i)^2$$

proven in Lemma 4 of Samorodnitsky and Trevisan [ST09]. □

10.2.4 Blocks

Let R, L, d be positive integers satisfying $R = dL$. Let $(\Omega^R, \mu^{\otimes R})$ be a product space and $\pi : [R] \to [L]$ be a projection such that $|\pi^{-1}(j)| = d$ for $1 \le j \le L$. Define $\overline{\Omega} := \Omega^d$. Given $x \in \Omega^R$, we *block* x to have $\overline{x} \in (\overline{\Omega})^L$ defined by

$$\overline{x}_j := (x_{j'})_{\pi(j')=j}.$$

Given $f \in \mathcal{L}(\Omega^R)$, its *blocked version* $\overline{f} \in \mathcal{L}(\overline{\Omega}^L)$ is defined by $\overline{f}(\overline{x}) := f(x)$ for any $x \in \Omega^R$. These blocked versions of functions and arguments depend on the projection π. For each function f, the associated projection will be clear from the context, and the same projection is used to block its argument x. The influence $\mathsf{Inf}_j[\overline{f}]$ and the noise operator $T_\rho \overline{f}$ are naturally defined. Define

$$\overline{\mathsf{Inf}}_j[f] := \mathsf{Inf}_j[\overline{f}], \quad \forall j \in [L] ; \qquad (\overline{T}_\rho f)(x) := (T_\rho \overline{f})(\overline{x}), \quad \forall x \in \Omega^R ,$$

and call them *block influence* and *block noise operator* respectively. They also have the following nice expressions in terms of f's Efron-Stein decomposition.

$$\overline{T}_\rho f = \sum_S \rho^{|\pi(S)|} f_S ; \qquad \overline{\mathsf{Inf}}_j[f] = \sum_{S: S\cap\pi^{-1}(j)\neq\emptyset} \|f_S\|_2^2 .$$

A subset $S \subseteq [R]$ is said to be *shattered* by π if $|S| = |\pi(S)|$. For a positive integer J, define the *bad* part of f_v under π and J as

$$f^{\mathsf{bad}} = \sum_{S:\text{not shattered and } |\pi(S)|<J} f_S.$$

10.2.5 Q-HYPERGRAPH LABEL COVER

An instance of Q-HYPERGRAPH LABEL COVER is based on a Q-uniform hypergraph $H = (V, E)$. Each hyperedge-vertex pair (e, v) such that $v \in e$ is associated with a projection $\pi_{e,v} : [R] \to [L]$ for some positive integers R and L. A labeling $l : V \to [R]$ *strongly satisfies* $e = \{v_1, \ldots, v_Q\}$ when $\pi_{e,v_1}(l(v_1)) = \cdots = \pi_{e,v_Q}(l(v_Q))$. It *weakly satisfies* e when $\pi_{e,v_i}(l(v_i)) = \pi_{e,v_j}(l(v_j))$ for some $i \neq j$. The following are two desired properties of instances of Q-HYPERGRAPH LABEL COVER.

- ϵ-weakly dense: any subset of V of measure at least $\epsilon' \geq \epsilon$ induces at least $\frac{(\epsilon')^Q}{2^{Q+1}}$ fraction of hyperedges.
- T-smooth: for all $v \in V$ and $i \neq j \in [R]$, $\quad \Pr_{e \in E: e \ni v}[\pi_{e,v}(i) = \pi_{e,v}(j)] \leq \frac{1}{T}$.

The following theorem asserts that it is NP-hard to find a good labeling in such instances.

Theorem 10.2.3. *For any $Q \geq 2, T \geq 1$ and $\eta, \epsilon > 0$, given an instance of Q-*HYPERGRAPH LABEL COVER *that is ϵ-weakly-dense and T-smooth, it is NP-hard to distinguish*

- *Completeness: There exists a labeling l that strongly satisfies every hyperedge.*
- *Soundness: No labeling l can weakly satisfy a fraction η of hyperedges.*

Proof. We reduce from T-smooth LABEL COVER first defined in Khot [Kho02a] to T-smooth Q-HYPERGRAPH LABEL COVER using the technique of Gopalan et al. [GKS10].

An instance of LABEL COVER consists of a biregular bipartite graph $G = (U \cup V, E)$ where each edge $e = (u, v)$ is associated with a projection $\pi_e : [R] \to [L]$ for some positive integers R and L. A labeling $l : U \cup V \to [R]$ satisfies e when $\pi_e(l(v)) = l(u)$. It is called T-smooth when for any $i \neq j$, $\Pr_e[\pi_e(i) = \pi_e(j)] \leq \frac{1}{T}$. The following theorem shows hardness of T-smooth LABEL COVER.

Theorem 10.2.4 ([Kho02a]). *For any $T \geq 1$ and $\eta' > 0$, given an instance of* LABEL COVER *that is T-smooth, it is NP-hard to distinguish*

- *Completeness: There exists a labeling l that satisfies edge.*
- *Soundness: No labeling l can satisfy a fraction η' of hyperedges.*

We first claim that in Theorem [Kho02a], without loss of generality, we can assume that the degree d of $u \in U$ is large enough as a function of Q and ϵ, such that for any $\epsilon' \geq \frac{\epsilon}{2}$,

$$\frac{\binom{d\epsilon'}{Q}}{\binom{d}{Q}} = \prod_{i=0}^{Q-1} \frac{(d\epsilon' - i)}{d - i} \geq \frac{(\epsilon')^Q}{2}. \tag{10.1}$$

This is possible because in the construction of [Kho02a], the operations to increase T and reduce η' both increase the degree, so we can increase the degree while making T and η' even stronger for our purpose.

Given such an instance of LABEL COVER $G = (U_G \cup V_G, E_G)$, the corresponding instance of $H = (V_H, E_H)$ is produced by

- $V_H = V_G$
- For $u \in U_G$ and Q distinct neighbors $v_1, \ldots, v_Q \in V_G$, we add a hyperedge $e = \{v_1, \ldots, v_Q\} \in E_H$ with the associated projections $\pi_{e,v_i} := \pi_{(u,v_i)}$. Say this hyperedge is *formed from* u. We can have the same hyperedges formed from different vertices.

Fix $v \in V_H$ and $i \neq j \in [R]$.

$$\Pr_{e \in E_H : v \in e}[\pi_{e,v}(i) = \pi_{e,v}(j)] = \Pr_{e=(u,v) \in E_G}[\pi_e(i) = \pi_e(j)] \leq \frac{1}{T},$$

so the resulting instance is also T-smooth.

For weak density, fix $I \subseteq V_H$ of measure ϵ, and for $u \in U_G$, let $\epsilon(u)$ be the fraction of neighbors of u contained in I. Biregularity of G implies $\epsilon = \mathbb{E}_u[\epsilon(u)]$. Let $\epsilon'(u) = \epsilon(u)$ if $\epsilon(u) \geq \frac{\epsilon}{2}$ and $\epsilon'(u) = 0$ otherwise. An averaging argument shows that $\mathbb{E}_u[\epsilon'(u)] \geq \frac{\mathbb{E}_u[\epsilon(u)]}{2} = \frac{\epsilon}{2}$. For any $u \in U_G$, whether $\epsilon'(u) = \epsilon(u) \geq \frac{\epsilon}{2}$ or $\epsilon'(u) = 0$, by (10.1), the fraction of hyperedges induced by I, out of the hyperedges formed from u, is at least

$$\frac{\binom{d\epsilon'(u)}{Q}}{\binom{d}{Q}} \geq \frac{(\epsilon'(u))^Q}{2}.$$

Then the fraction of hyperedges induced by I is at least

$$\mathbb{E}_{u \in U_G}[\frac{(\epsilon'(u))^Q}{2}] = \frac{1}{2}\mathbb{E}_{u \in U_G}[(\epsilon'(u))^Q] \geq \frac{1}{2}(\mathbb{E}_{u \in U_G}[\epsilon'(u)])^Q \geq \frac{\epsilon^Q}{2^{Q+1}}.$$

For completeness, given a labeling $l : U_G \cup V_G \to [R]$ that satisfies every edge of G, its projection to $V_G = V_H$ will strongly satisfy every hyperedge of H.

For soundness, let $l : V_H \to [R]$ be a labeling that weakly satisfies η fraction of hyperedges for some $\eta > 0$. Let $\eta(u)$ be the fraction of hyperedges satisfied by l formed from u, out of all hyperedges formed from u. Consider the following randomized strategy for G: V_G is labelled by l, and each $u \in U_G$ independently samples one of its neighbors v and set $l(u) \leftarrow \pi_{(u,v)}(l(v))$. The expected fraction of edges incident on u satisfied by this decoding strategy is (let $N(u)$ be the set of neighbors of u and $({}_{N(u)}P_Q)$ be the set of Q-tuples of the neighbors where Q vertices are pairwise distinct)

$$\begin{aligned}
&\mathbb{E}_{v_1 \in N(u)}[\Pr_{v_2 \in N(u)}[\pi_{(u,v_1)}(l(v_1)) = \pi_{(u,v_2)}(v_2)]] \\
&= \Pr_{(v_1,\ldots,v_Q) \in N(u)^Q}[\pi_{(u,v_1)}(l(v_1)) = \pi_{(u,v_2)}(v_2)] \\
&\geq \Pr_{(v_1,\ldots,v_Q) \in ({}_{N(u)}P_Q)}[\pi_{(u,v_1)}(l(v_1)) = \pi_{(u,v_2)}(v_2)] \\
&\geq \frac{1}{\binom{Q}{2}} \Pr_{(v_1,\ldots,v_Q) \in ({}_{N(u)}P_Q)}[e := \{v_1, \ldots, v_Q\} \text{ is weakly satisfied}] \\
&= \frac{1}{\binom{Q}{2}} \Pr_{\{v_1,\ldots,v_Q\} \in \binom{N(u)}{Q}}[e := \{v_1, \ldots, v_Q\} \text{ is weakly satisfied}] \\
&= \frac{\eta(u)}{\binom{Q}{2}}.
\end{aligned}$$

Overall, the strategy satisfies $\frac{\eta}{\binom{Q}{2}}$ fraction of edges of G in expectation. Setting $\eta' < \frac{\eta}{\binom{Q}{2}}$, we have contradiction, completing the proof of soundness. □

10.2.6 $(Q+1)$-BIPARTITE HYPERGRAPH LABEL COVER

$(Q+1)$-BIPARTITE HYPERGRAPH LABEL COVER is used in Theorem 9.2.6 for Q-out-of-$(2Q+1)$-SAT. An instance of $(Q+1)$-BIPARTITE HYPERGRAPH LABEL COVER is based on a $(Q+1)$-uniform bipartite hypergraph $H = (U \cup V, E)$, where each hyperedge e contains one vertex from U and Q vertices from V. For every hyperedge $e = \{u, v_1, \ldots, v_Q\}$ such that $u \in U$ and $v_q \in V$, each v_q is associated with a projection $\pi_{e,v_q} : [R] \to [L]$

for some positive integers R and L. A labeling $l : U \cup V \to [R]$ *strongly satisfies* $e = \{v_1, \ldots, v_Q\}$ when $l(u) = \pi_{e,v_1}(l(v_1)) = \cdots = \pi_{e,v_Q}(l(v_Q))$ (we can imagine that $\pi_{e,u}$ is also defined as the identity). It *weakly satisfies* e when $\pi_{e,v_i}(l(v_i)) = \pi_{e,v_j}(l(v_j))$ for some $i \neq j$ or $\pi_{e,v_i}(l(v_i)) = l(u)$ for some i. As usual, the instance is T-smooth if for any $v \in V$ and $i \neq j$,

$$\Pr_{e \in E : v \in e}[\pi_{e,v}(i) = \pi_{e,v}(j)] \leq \frac{1}{T}.$$

Note that we do not need weak density for Q-out-of-$(2Q+1)$-SAT.

Theorem 10.2.5. *For any $Q \geq 2, T \geq 1$ and $\eta > 0$, given an instance of $(Q+1)$-*BIPARTITE HYPERGRAPH LABEL COVER *that is T-smooth, it is NP-hard to distinguish*

- *Completeness: There exists a labeling l that strongly satisfies every hyperedge.*
- *Soundness: No labeling l can weakly satisfy a fraction η of hyperedges.*

Proof. As in Theorem 10.2.3, we reduce from T-smooth LABEL COVER.

Given an instance of LABEL COVER $G = (U_G \cup V_G, E_G)$, the corresponding instance of $H = (U_H \cup V_H, E_H)$ is produced by

- $U_H = U_G, V_H = V_G$
- For $u \in U_G$ and Q distinct neighbors $v_1, \ldots, v_Q \in V_G$, we add a hyperedge $e = \{u, v_1, \ldots, v_Q\} \in E_H$ with the associated projections $\pi_{e,v_i} := \pi_{(u,v_i)}$. Say this hyperedge is *formed from* u.

Fix $v \in V_H$ and $i \neq j \in [R]$.

$$\Pr_{e \in E_H : v \in e}[\pi_{e,v}(i) = \pi_{e,v}(j)] = \Pr_{e=(u,v) \in E_G}[\pi_e(i) = \pi_e(j)] \leq \frac{1}{T},$$

so the resulting instance is also T-smooth.

For completeness, given a labeling $l : U_G \cup V_G \to [R]$ that satisfies every edge of G, it is easy to check that the same l will strongly satisfy every hyperedge of H.

For soundness, let $l : V_H \to [R]$ be a labeling that weakly satisfies η fraction of hyperedges for some $\eta > 0$. Let $\eta(u)$ be the fraction of hyperedges satisfied by l formed from u, out of all hyperedges formed from u. Consider the following randomized strategy for G:

- V_G is labeled by l.
- Each $u \in U_G$ is assigned $l(u)$ with probability half. With the remaining $1/2$ probability, it independently samples one of its neighbors v and sets $l(u) \leftarrow \pi_{(u,v)}(l(v))$.

Let $N(u)$ be the set of neighbors of u and $({}_{N(u)}P_Q)$ be the set of Q-tuples of the neighbors where Q vertices are pairwise distinct. The expected fraction of edges incident on u satisfied by this decoding strategy is

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}_{v_1 \in N(u)}[\Pr_{v_2 \in N(u)}[\pi_{(u,v_1)}(l(v_1)) = \pi_{(u,v_2)}(l(v_2))]] + \frac{1}{2}\Pr_{v \in N(u)}[\pi_{(u,v)}(l(v)) = l(u)] \\
&= \frac{1}{2}\Pr_{(v_1,\ldots,v_Q) \in N(u)^Q}[\pi_{(u,v_1)}(l(v_1)) = \pi_{(u,v_2)}(l(v_2)) \text{ or } \pi_{(u,v_1)}(l(v_1)) = l(u)] \\
&\geq \frac{1}{2}\Pr_{(v_1,\ldots,v_Q) \in ({}_{N(u)}P_Q)}[\pi_{(u,v_1)}(l(v_1)) = \pi_{(u,v_2)}(l(v_2)) \text{ or } \pi_{(u,v_1)}(l(v_1)) = l(u)] \\
&\geq \frac{1}{2\binom{Q}{2}}\Pr_{(v_1,\ldots,v_Q) \in ({}_{N(u)}P_Q)}[e := \{v_1, \ldots, v_Q\} \text{ is weakly satisfied}] \\
&= \frac{1}{2\binom{Q}{2}}\Pr_{\{v_1,\ldots,v_Q\} \in \binom{N(u)}{Q}}[e := \{v_1, \ldots, v_Q\} \text{ is weakly satisfied}] \\
&= \frac{\eta(u)}{2\binom{Q}{2}}.
\end{aligned}
$$

Overall, the strategy satisfies $\frac{\eta}{2\binom{Q}{2}}$ fraction of edges of G in expectation. Setting $\eta' < \frac{\eta}{2\binom{Q}{2}}$, we have contradiction, completing the proof of soundness. □

10.2.7 MULTILAYERED LABEL COVER

We reduce MULTILAYERED LABEL COVER defined by Dinur et al. [DGKR05] with the smoothness property to K-HYPERGRAPH VERTEX COVER. An instance of MULTILAYERED LABEL COVER with A layers is based on a graph $G = (V, E)$ where $V = V_1 \cup \cdots \cup V_A$ and $E = \cup_{1 \leq i < j \leq A} E_{i,j}$. Let $[R_i]$ be the label set of the variables in the V_i such that R_i divides R_j for all $i < j$. Any edge $e \in E_{i,j}$ is between $u \in V_i$ and $v \in V_j$, and associated with a projection $\pi_e : [R_j] \rightarrow [R_i]$. Given a labeling $l : V \rightarrow [R_A]$, an edge $e = (u, v)$ with $u \in V_i$ and $v \in V_j$ ($i < j$) is satisfied when $\pi_e(l(v)) = l(u)$. The following are desired properties of an instance. Note that the definition of weak density here is not parameterized by ϵ.

- Weakly dense: for any $\epsilon > 0$ satisfying $\lceil \frac{4}{\epsilon} \rceil \leq A$, given $m = \lceil \frac{4}{\epsilon} \rceil$ layers $i_1 < \cdots < i_m$ and given any sets $I_{i_j} \subseteq V_{i_j}$ with $|I_{i_j}| \geq \epsilon |V_{i_j}|$, there exist $j < j'$ such that at least $\frac{\epsilon^3}{16}$ fraction of the edges between V_{i_j} and $V_{i_{j'}}$ are indeed between I_{i_j} and $I_{i_{j'}}$.

- T-smooth: for any $1 \leq i < j \leq A$, $v \in V_j$ and $a \neq b \in [R_j]$,

$$\Pr_{u \in V_i : (u,v) \in E_{i,j}} [\pi_{u,v}(a) = \pi_{u,v}(b)] \leq \frac{1}{T}.$$

Theorem 10.2.6 ([Kho02a]). *For every $\eta > 0$, $A \geq 2$ and $T \geq 1$, given an instance of* MULTILAYERED LABEL COVER *with A layers that is weakly dense and T-smooth, it is NP-hard to distinguish the following cases:*

- *Completeness: There exists a labeling l that satisfies every edge.*

- *Soundness: No labeling l can satisfy a fraction η of any $E_{i,j}$.*

10.3 Reverse Hypercontractivity

The version of reverse hypercontractivity we use is stated below.

Theorem 10.3.1 ([MOS13]). *Let (Ω, μ) be a probability space. Fix $0 \leq \rho < 1$. There exist $q < 0 < p < 1$ such that for any $f \in \mathcal{L}_{[0,\infty)}(\Omega)$,*

$$\|T_\rho f\|_q \geq \|f\|_p.$$

We now generalize the above reverse hypercontractivity result to more general operators, extending the noise operator T_ρ in two ways.

- Between two difference spaces: while T_ρ is the Markov operator associated with two correlated copies of the same probability space $(\Omega_1 \times \Omega_1, \nu)$, we are interested in the Markov operator T associated with two correlated spaces $(\Omega_1 \times \Omega_2, \nu')$, possibly $\Omega_1 \neq \Omega_2$.

- Arbitrary distribution instead of diagonal distribution: ν samples x, y independently according to the marginal and output (x, x) with probability ρ and (x, y) with probability $1 - \rho$. Since $\Omega_1 \neq \Omega_2$, the former does not make sense. Instead, with probability ρ, ν' samples (x, y) according to another arbitrary distribution ν'', as long as the marginals of x and y are preserved.

This extension is based on simple observation that such an operator T can be expressed as $T = PT_\rho$ for some Markov operator $P : \mathcal{L}(\Omega_1) \to \mathcal{L}(\Omega_2)$ which shares the marginals with T. The idea of decomposition in terms of T_ρ was also used in [MOS13] when analyzing general operators on the same space. The following lemma shows that any Markov operator does not decrease q-norm when $q \le 1$.

Lemma 10.3.2. *Let $(\Omega_1 \times \Omega_2, \mu)$ be two correlated spaces, with the marginal distribution μ_i of Ω_i. Let P be the Markov operator associated with it. For any $q \le 1$ and $f \in \mathcal{L}_{(0,\infty)}(\Omega_1)$,*

$$\|Pf\|_q \ge \|f\|_q.$$

Proof. Since $x \mapsto x^q$ is concave,

$$\begin{aligned}\|Pf\|_q^q &= \mathbb{E}_{y\sim\mu_2}[(Tf(y))^q] = \mathbb{E}_{y\sim\mu_2}[(\mathbb{E}_{x\sim\mu_1}[f(x)|y])^q] \\ &\ge \mathbb{E}_{y\sim\mu_2}[\mathbb{E}_{x\sim\mu_1}[f(x)^q|y])] = \mathbb{E}_{x\sim\mu_1}[f(x)^q] = \|f\|_q^q .\end{aligned}$$

□

The following main lemma says that whenever T_ρ exhibits the reverse hypercontractive behavior for some p, q, the same conclusion holds for Markov operators with the same parameters.

Lemma 10.3.3 (Reverse Hypercontractivity of two correlated spaces)**.** *Let $(\Omega_1 \times \Omega_2, \mu)$ be two correlated spaces, and with the marginal distribution μ_i of Ω_i. Let T be the Markov operator associated with it. Suppose that $T = \rho P + (1-\rho)J_{1,2}$ for $0 \le \rho < 1$, where $J_{1,2}$ is the Markov operator associated with $(\Omega_1 \times \Omega_2, \mu_1 \otimes \mu_2)$ and P is the Markov operator associated with $(\Omega_1 \times \Omega_2, \nu)$ for some ν with the same marginals as μ. Let $q < p < 1$ be such that $\|T_\rho f\|_q \ge \|f\|_p$ for any $f \in \mathcal{L}_{[0,\infty)}(\Omega_1)$. Then,*

$$\|Tf\|_q \ge \|f\|_p.$$

Proof. Note that $T_\rho = \rho I_1 + (1-\rho)J_1$, where I_1 is the identity operator, and J_1 is the Markov operator associated with $(\Omega_1^2, \mu_1^{\otimes 2})$. The following simple relationship holds between T and T_ρ.

$$PT_\rho = \rho P I_1 + (1-\rho)PJ_1 = \rho P + (1-\rho)J_{1,2} = T$$

The fact that $T = PT_\rho$ implies

$$\|Tf\|_q = \|PT_\rho f\|_q \ge \|T_\rho f\|_q \ge \|f\|_p,$$

where the first inequality follows from Lemma 10.3.2. □

Along the way to apply the above result to our setting, we introduce a basic intermediate problem which may be of independent interest.

Question 10.3.4. *Let $(\Omega_1 \times \Omega_2, \mu)$ be two correlated spaces. Given two (biased, not necessarily Boolean) hypercubes Ω_1^L and Ω_2^L, their subsets $S \subseteq \Omega_1^L, S' \subseteq \Omega_2^L$, and two random points $x \in \Omega_1^L, y \in \Omega_2^L$ such that each (x_i, y_i) is sampled from μ independently, what is the probability that $x \in S$ and $y \in S'$?*

To answer this question, we use the following reverse Hölder inequality in a similar way to [MOR+06].

Theorem 10.3.5. *[HLP52] Let f and g be nonnegative functions and supplse $\frac{1}{p} + \frac{1}{p'} = 1$, where $p < 1$. Then*

$$\mathbb{E}[fg] = ||fg||_1 \geq ||f||_p ||g||_{p'}.$$

Using the above inequality and the standard two-function hypercontractivity induction [O'D14], the following lemma shows that as long as μ contains nonzero copy of product distributions (equivalent to $T = \rho P + (1 - \rho) J_{1,2}$ for $\rho < 1$), the above probability is at least a positive number depending only on the measure of S and S', and ρ (but crucially it does not depend on L). Note that when f is an indicator function whose value is either 0 or 1, for any $p > 0$, $||f||_p = (\mathbb{E}_x[f(x)^p])^{1/p} = (\mathbb{E}[f])^{1/p}$.

Lemma 10.3.6. *Let $(\Omega_1, \Omega_2, \mu), \rho, T, P$ be defined as Lemma 10.3.3. There exist $0 < p, q < 1$ such that for any $f \in \mathcal{L}_{[0,\infty)}(\Omega_1^L)$ and $g \in \mathcal{L}_{[0,\infty)}(\Omega_2^L)$,*

$$\mathbb{E}_{(x,y) \sim \mu^{\otimes L}}[f(x)g(y)] = \mathbb{E}_{y \sim \mu_2^{\otimes L}}[g(y) T^{\otimes L} f(y)] \geq \|f\|_p \|g\|_q$$

Proof. The equality holds by definition, so it only remains to prove the inequality. We first prove it for $L = 1$, and do the induction on L. Invoke Theorem 10.3.1 to get $q' < 0 < p < 1$ such that $\|T_\rho f\|_{q'} \geq \|f\|_p$. Let $0 < q < 1$ be such that $\frac{1}{q} + \frac{1}{q'} = 1$. By the reverse Hölder inequality and Lemma 10.3.3,

$$\mathbb{E}_{(x,y) \sim \mu}[f(x)g(y)] = \mathbb{E}_{y \sim \mu_2}[g(y) T f(y)] \geq \|Tf\|_{q'} \|g\|_q \geq \|f\|_p \|g\|_q$$

as desired.

For $L > 1$, we use the notation $x = (x', x_L)$ where $x' = (x_1, \ldots, x_{L-1})$, and similar notation for y. Note that $(x', y') \sim \mu^{\otimes L-1}$ and $(x_L, y_L) \sim \mu$. We also write f_{x_L} for the restriction of f in which the last coordinate is fixed to value x_L, and similarly for g.

$$\begin{aligned}
\mathbb{E}_{(x,y) \sim \mu^{\otimes L}}[f(x)g(y)] &= \mathbb{E}_{(x_L, y_L) \sim \mu} \mathbb{E}_{(x', y') \sim \mu^{\otimes L-1}}[f_{x_L}(x') g_{y_L}(y')] \\
&\geq \mathbb{E}_{(x_L, y_L) \sim \mu}[\|f_{x_L}\|_{p, \mu_1^{\otimes L-1}} \|g_{y_L}\|_{q, \mu_2^{\otimes L-1}}]
\end{aligned}$$

by induction. Let F, G be the function defined by $F(x_L) = \|f_{x_L}\|_p$, $G(y_L) = \|g_{y_L}\|_q$.

$$\mathbb{E}_{(x_L,y_L)\sim\mu}[F(x_L)G(y_L)] \geq \|F\|_{p,\mu_1}\|G\|_{q,\mu_2}$$

by the base case. Finally,

$$\|F\|_{p,\mu_1} = \mathbb{E}_{x_L\sim\mu_1}[|F(x_L)|^p]^{1/p} = (\mathbb{E}_{x_L\sim\mu_1}\mathbb{E}_{x'\sim\mu_1^{\otimes L-1}}[|f_{x_L}|^p])^{1/p} = \|f\|_{p,\mu_1^{\otimes L}}$$

and similarly $\|G\|_{q,\mu_2} = \|g\|_{q,\mu_2^{\otimes L}}$. The induction is complete. □

By another induction on the number of functions, we can extend the answer to the previous question to $k > 2$.

Question 10.3.7. *Let (Ω^k, μ) be k correlated copies of the same space. Given a hypercube Ω^L, its subsets $S \subseteq \Omega^L$, and k random points $x_1, \ldots, x_k \in \Omega^L$ such that each $((x_1)_i, \ldots, (x_k)_i)$ is sampled from μ independently for $i \in [L]$, what is the probability that $x_j \in S$ for all $j \in [k]$?*

Theorem 10.3.8. *Let (Ω^k, ν) be k correlated spaces with the same marginal σ for each copy of Ω. Suppose that ν is described by the following procedure to sample from Ω^k.*

- *With probability ρ $(0 \leq \rho < 1)$, it samples from an arbitrary distribution on Ω^k, which has the marginal σ for each copy of Ω.*
- *With probability $1-\rho$, it samples from $\sigma^{\otimes k}$.*

Let $F_1, \ldots, F_k \in \mathcal{L}_{[0,1]}(\Omega^L)$ such that $\mathbb{E}[F_i] \geq \epsilon > 0$ for all i. Then there exists $\zeta := \zeta(\rho, \epsilon, k) = \epsilon^{O_{\rho,k}(1)} > 0$ (independent of L) such that

$$\mathbb{E}_{x_1,\ldots,x_k}[\prod_{1\leq i\leq k} F_i(x_i)] \geq \zeta$$

where for each $1 \leq j \leq L$, $((x_1)_j, \ldots, (x_k)_j)$ is sampled according to ν.

Proof. We proceed by the induction on k. For $k = 1$, $\zeta = \epsilon$ works.

For $k > 1$, consider two correlated spaces $(\Omega \times \Omega^{k-1}, \nu)$ where the marginal of Ω is σ and the marginal of Ω^{k-1} is ν'. Note that the marginal of ν' on each copy of Ω is still σ. Invoke Lemma 10.3.6 to obtain $0 < p, q < 1$ be such that

$$\mathbb{E}_{(x,y)\sim\nu^{\otimes L}}[F(x)G(y)] \geq \|F\|_{p,\sigma^{\otimes L}}\|G\|_{q,\nu'^{\otimes L}}$$

for any $F \in \mathcal{L}_{[0,\infty)}(\Omega^L)$ and $G \in \mathcal{L}_{[0,\infty)}(\Omega^{k-1})^L$.

$$\mathbb{E}_{x_1,\dots,x_k}[\prod_{1 \le i \le k} F_i(x_i)] \ge \|F_1\|_{p,\sigma^{\otimes L}} \| \prod_{i=2}^{k} F_i(x_i)\|_{q,\nu'^{\otimes L}}$$

Since $F_i \in \mathcal{L}_{[0,1]}(\Omega^L)$, $\|F_i\|_p \ge \epsilon^{1/p}$. Since ν' can be also described by the procedure in the statement of the theorem (except that it is on Ω^{k-1}), we obtain $\zeta(\rho, \epsilon, k-1)$ such that

$$\| \prod_{i=2}^{k} F_i(x_i)\|_{q,\nu'^{\otimes L}} \ge \Big(\mathbb{E}_{x_2,\dots,x_k}[\prod_{i=2}^{k} F_i(x_i)]\Big)^{1/q} \ge \zeta(\rho, \epsilon, k-1)^{1/q}$$

Therefore, $\zeta(\rho, \epsilon, k) = \zeta(\rho, \epsilon, k-1)^{1/q}\epsilon^{1/p}$ completes the induction. Since p, q depend only on ρ, $\zeta(\rho, \epsilon, k) = \epsilon^{O_{\rho,k}(1)}$ in every step of induction. $\square$

Remark 10.3.9. *The same statement holds even when we replace Ω^k by the product of k different spaces $\Omega_1 \times \cdots \times \Omega_k$.*

10.4 Hardness of Rainbow Coloring

Fix $Q, k \ge 2$. In this section, we show a reduction from Q-HYPERGRAPH LABEL COVER to coloring a Qk-uniform hypergraph, proving Theorem 9.2.1.

10.4.1 Distributions

We first define the distribution for each block. Qk points $x_{q,i} \in [k]^d$ for $1 \le q \le Q$ and $1 \le i \le k$ are sampled by the following procedure.

- Sample $q' \in [Q]$ uniformly at random.
- Sample $x_{q',1}, \dots, x_{q',k} \in [k]^d$ i.i.d.
- For $q \neq q'$, $1 \le j \le d$, sample a permutation $((x_{q,1})_j, \dots, (x_{q,k})_j) \in \mathbb{S}_k$ uniformly at random.

There are several distributions involved.

Let $\Omega := [k]$ and ω be the uniform distribution on Ω. For any $1 \le q \le Q$, $1 \le i \le k$ and $1 \le j \le d$, the marginal of $(x_{q,i})_j$ follows (Ω, ω).

For any $1 \le q \le Q$ and $1 \le i \le k$, the marginal of $(x_{q,i})$ follows $(\Omega^d, \omega^{\otimes d})$. Let $\overline{\Omega} := \Omega^d$.

Let (Ω^k, μ) be the marginal distribution of $((x_{q,1})_j, \ldots, (x_{q,k})_j)$, which is the same for all q and i. Note that μ is not uniform — with probability $\frac{1}{Q}$ it is uniform on $[k]^k$, but with probability $\frac{Q-1}{Q}$ it samples from $k!$ permutations.

Let $(\Omega^{dk}, \overline{\mu})$ be the marginal distribution of $(x_{q,1}, \ldots, x_{q,k})$, which is the same for all q.

Finally, let $(\Omega^{Qkd}, \overline{\mu}')$ be the entire distribution of $(x_{q,i})_{q \in [Q], i \in [k]}$.

We first consider $(\Omega^{Qkd}, \overline{\mu}')$ as Qk correlated spaces $(\overline{\Omega}^{Qk}, \overline{\mu}')$, and bound $\rho(\overline{\Omega}^{Qk}; \overline{\mu}')$. Let $\overline{\Omega}_{q,i}$ denote the copy of $\overline{\Omega}$ associated with $x_{q,i}$, and $\overline{\Omega}'_{q,i}$ be the product of the other $Qk - 1$ copies.

Fix some q and i. Note that $\overline{\mu}' = \frac{1}{Q}\alpha_q + \frac{Q-1}{Q}\beta_q$ where α_q denotes the distribution given $q' = q$ (so that each entry of $x_{q,1}, \ldots, x_{q,k}$ is sampled i.i.d.), and β_q denotes the distribution given $q' \neq q$. Since each entry of $x_{q,i}$ is sampled i.i.d. in α_q, $\rho(\overline{\Omega}_{q,i}, \overline{\Omega}'_{q,i}; \alpha_q) = 0$. Observed that, in both α_q and β_q, the marginal of $x_{q,i}$ is $\omega^{\otimes d}$. By Lemma 10.2.1, we conclude that $\rho(\overline{\Omega}_{q,i}, \overline{\Omega}'_{q,i}; \overline{\mu}') \le \sqrt{\frac{Q-1}{Q}}$. Therefore we have

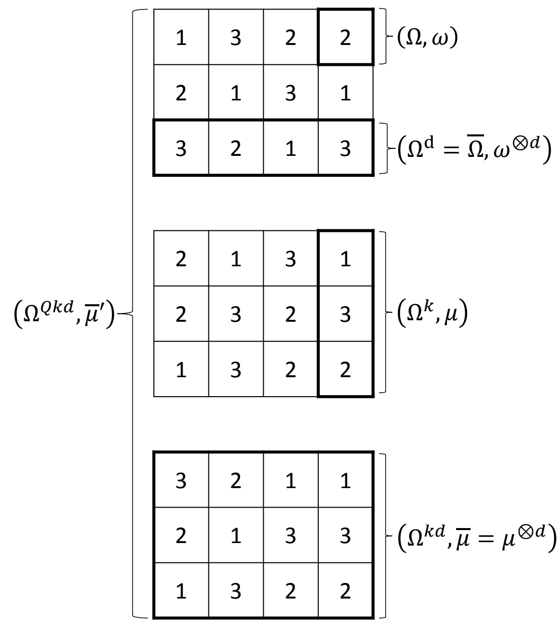

Figure 10.1: An example for $Q = k = 3$, $d = 4$. $q' = 2$ so that all columns of the first and third block are permutations.

$$\rho((\overline{\Omega}_{q,i})_{q,i}; \overline{\mu}') = \max_{q,i} \rho(\overline{\Omega}_{q,i}, \overline{\Omega}'_{q,i}; \overline{\mu}') \le \sqrt{\frac{Q-1}{Q}} .$$

10.4.2 Reduction and Completeness

We now describe the reduction from Q-HYPERGRAPH LABEL COVER. Given a Q-uniform hypergraph $H = (V, E)$ with Q projections from $[R]$ to $[L]$ for each hyperedge (without loss of generality[1], we assume each projection is d-to-1 where $d = R/L$), the resulting instance of Qk-hypergraph coloring is $H' = (V', E')$ where $V' = V \times [k]^R$. Let $\mathsf{Cloud}(v) := \{v\} \times [k]^R$. The set E' consists of hyperedges generated by the following procedure.

- Sample a random hyperedge $e = (v_1, \ldots, v_Q) \in E$ with associated projections $\pi_{e,v_1}, \ldots, \pi_{e,v_Q}$ from E.
- Sample $(x_{q,i})_{1 \le q \le Q, 1 \le i \le k} \in \Omega^R$ in the following way. For each $1 \le j \le L$, independently sample $((x_{q,i})_{\pi^{-1}_{e,v_q}(j)})_{q,i}$ from $(\Omega^{Qkd}, \overline{\mu}')$.
- Add a hyperedge between Qk vertices $\{(v_q, x_{q,i})\}_{q,i}$ to E'. We say this hyperedge is *formed from* $e \in E$.

Given the reduction, completeness is easy to show.

Lemma 10.4.1. *If an instance of* Q-HYPERGRAPH LABEL COVER *admits a labeling that strongly satisfies every hyperedge* $e \in E$*, there is a coloring* $c : V' \to [k]$ *such that every hyperedge* $e' \in E'$ *has at least* $(Q-1)$ *vertices of each color.*

Proof. Let $l : V \to [R]$ be a labeling that strongly satisfies every hyperedge $e \in E$. For any $v \in V, x \in [k]^R$, let $c(v, x) = x_{l(v)}$. For any hyperedge $e' = \{(v_q, x_{q,i})\}_{q,i} \in E'$, $c(v_q, x_{q,i}) = (x_{q,i})_{l(v_q)}$, and all but one q satisfies $\{(x_{q,1})_{l(v_q)}, \ldots, (x_{q,k})_{l(v_q)}\} = [k]$. Therefore, the above strategy ensures that every hyperedge of E' contains at least $(Q-1)$ vertices of each color. □

10.4.3 Soundness

Lemma 10.4.2. *For any* $\epsilon > 0$*, there exists* $\eta := \eta(\epsilon, Q, k)$ *such that if* $I \subseteq V'$ *of measure* ϵ *induces less than* $\epsilon^{O_{Q,k}(1)}$ *fraction of hyperedges, the corresponding instance of* Q-HYPERGRAPH LABEL COVER *admits a labeling that weakly satisfies a fraction* η *of hyperedges.*

As introduced in Section 10.1, the proof of soundness consists of the following five steps.

[1]We can assume that the number of labels from $[R]$ that project to a fixed label in $[L]$ is the same for all projections, since original LABEL COVER is also hard to approximate with this condition as shown in Theorem 1.17 of [Wen13].

STEP 1. **Fixing a Good Hyperedge.** Let $I \subseteq V'$ be of measure ϵ. For each vertex $v \in V$, let $f_v : [k]^R \to \{0,1\}$ be the indicator function of $I \cap \mathsf{Cloud}(v)$. Call a vertex v *heavy* when $\mathbb{E}[f_v] \geq \frac{\epsilon}{2}$. By averaging, at least $\frac{\epsilon}{2}$ fraction of vertices are heavy. By Theorem 10.2.3, we can assume that the original Q-HYPERGRAPH LABEL COVER instance is $\frac{\epsilon}{2}$-weakly-dense. At least $\delta := \frac{(\frac{\epsilon}{2})^Q}{2^{Q+1}}$ fraction of hyperedges are induced by the heavy vertices.

Recall that we can require the original Q-HYPERGRAPH LABEL COVER instance to be T-smooth for T that can be chosen arbitrarily large. Let J be a positive integer. The parameters J and T will be determined later as large constants depending on Q, k, and ϵ.

Fix f_v and $S \subseteq [R]$. Over a random hyperedge e containing v and the associated projection $\pi_{e,v}$, we bound the probability that $|S|$ is not shattered and $|\pi_{e,v}(S)| < J$. If $|S| \leq J$, by union bound over all pairs $i \neq j$, the probability that S is not shattered is at most $\frac{J^2}{T}$. If $|S| > J$, the probability that $|\pi_{e,v}(S)| < J$ is at most the probabilty that a fixed J-subset of S is not shattered, which is at most $\frac{J^2}{T}$. Since $\sum_S \|(f_v)_S\|_2^2 = \|f_v\|_2^2 \leq 1$, we have

$$\mathbb{E}_e[\|f_v^{\mathsf{bad}}\|_2^2] \leq \frac{J^2}{T}.$$

where f_v^{bad} denotes the bad part of f_v under $\pi_{e,v}$ and J (we suppress the dependence on the projection $\pi_{e,v}$ and J for notational convenience). Therefore, $\mathbb{E}_e[\|f_v^{\mathsf{bad}}\|_2] \leq (\frac{J^2}{T})^{1/2}$ and at least $1 - (\frac{J^2}{T})^{1/4}$ fraction of hyperedges containing v satisfy $\|f_v^{\mathsf{bad}}\|_2 \leq (\frac{J^2}{T})^{1/4}$. Call such hyperedges *good* for v.

By union bound, at least $1 - Q(\frac{J^2}{T})^{1/4}$ fraction of hyperedges are good for every vertex they contain. By setting $Q(\frac{J^2}{T})^{1/4} \leq \frac{\delta}{2}$, we can conclude that at least a fraction $\frac{\delta}{2}$ of hyperedges are induced by the heavy vertices and good for every vertex they contain.

Throughout the rest of the section, fix such a hyperedge $e = (v_1, \ldots, v_Q)$ and the associated projections $\pi_{e,v_1}, \ldots, \pi_{e,v_Q}$. For simplicity, let $f_q := f_{v_q}$ and $\pi_q := \pi_{e,v_q}$ for $q \in [Q]$. We now measure the fraction of hyperedges formed from e that are wholly contained within I. The fraction such hyperedges is

$$\mathbb{E}_{x_{q,i}}[\prod_{1 \leq q \leq Q, 1 \leq i \leq k} f_q(x_{q,i})] \,. \tag{10.2}$$

STEP 2. **Lower Bounding in Each Hypercube.** Fix any $q \in [Q]$. We prove that $\mathbb{E}[\prod_{1 \leq i \leq k} T_{1-\gamma} f_q(x_{q,i})] \geq \zeta$ for some $\zeta > 0$ and every $\gamma \in [0,1]$. The main tool in this part is a generalization of reverse hypercontractivity, which is discussed in Section 10.3. The final result is the following.

Theorem 10.4.3 (Restatement of Theorem 10.3.8)**.** *Let (Ω^k, ν) be k correlated spaces with the same marginal σ for each copy of Ω. Suppose that ν is described by the following procedure to sample from Ω^k.*

- *With probability ρ $(0 \leq \rho < 1)$, it samples from an arbitrary distribution on Ω^k, which has the same marginal σ for each copy of Ω.*
- *With probability $1-\rho$, it samples from $\sigma^{\otimes k}$.*

Let $F_1, \dots, F_k \in \mathcal{L}_{[0,1]}(\Omega^L)$ such that $\mathbb{E}[F_i] \geq \epsilon > 0$ for all i. Then there exists $\zeta := \zeta(\rho, \epsilon, k) = \epsilon^{O_{\rho,k}(1)} > 0$ (independent of L) such that

$$\mathbb{E}_{x_1,\dots,x_k}[\prod_{1 \leq i \leq k} F_i(x_i)] \geq \zeta$$

where for each $1 \leq j \leq L$, $((x_1)_j, \dots, (x_k)_j)$ is sampled according to ν.

For each $1 \leq j \leq L$, $((\overline{x_{q,1}})_j, \dots, (\overline{x_{q,k}})_j)$ is sampled according to $(\overline{\Omega}^k, \overline{\mu})$. $\overline{\mu}$ satisfies the requirement of Theorem 10.3.8 — with probability $\frac{1}{Q}$, it samples from $\omega^{\otimes kd}$, and with probability $\frac{Q-1}{Q}$, it samples from d permutations from $\mathbb{S}_k$ independently so that the marginal of each $(\overline{x_{q,i}})_j$ is $\omega^{\otimes d}$ for all i and j.

Therefore, we can apply Theorem 10.3.8 (setting $\Omega \leftarrow \overline{\Omega}$, $k \leftarrow k$, $\sigma \leftarrow \omega^{\otimes d}$, $\nu \leftarrow \overline{\mu}$, $\rho \leftarrow \frac{Q-1}{Q}$, $F_1 = \cdots = F_k \leftarrow \overline{f_q}$, $\epsilon \leftarrow \frac{\epsilon}{2}$) to conclude that there exists $\zeta := \zeta(\frac{Q-1}{Q}, \frac{\epsilon}{2}, k) = \epsilon^{O_{Q,k}(1)} > 0$ such that

$$\mathbb{E}_{x_{q,1},\dots,x_{q,k}}[\prod_{1 \leq i \leq k} f_q(x_{q,i})] = \mathbb{E}_{\overline{x_{q,1}},\dots,\overline{x_{q,k}}}[\prod_{1 \leq i \leq k} \overline{f_q}(\overline{x_{q,i}})] \geq \zeta.$$

The only properties of f_q used were $\mathbb{E}[f_q] \geq \frac{\epsilon}{2}$ and $f_q \in \mathcal{L}_{[0,1]}(L^R)$. For any $0 \leq \gamma \leq 1$, $T_{1-\gamma} f_q$ have the same properties, so we have the following lower bound for every $q \in [Q]$

$$\mathbb{E}[\prod_{1 \leq i \leq k} T_{1-\gamma} f_q(x_{q,i})] \geq \zeta \ . \tag{10.3}$$

STEP 3. **Smoothing Functions.** From unnoised functions to block noised functions, we use the following theorem from Mossel [Mos10].

Theorem 10.4.4 ([Mos10])**.** *Let $(\Omega_1 \times \cdots \times \Omega_K, \nu)$ be K correlated spaces such that $\rho(\Omega_1, \dots, \Omega_K; \nu) \leq \rho < 1$. Consider K product spaces $((\Omega_1)^L \times \cdots \times (\Omega_K)^L, \nu^{\otimes L})$,* *and* $F_i \in \mathcal{L}((\Omega_i)^L)$ *for* $i \in [K]$ *such that* $\mathsf{Var}[F_i] \leq 1$. *For every* $\epsilon > 0$, *there exists* $\gamma := \gamma(\epsilon, \rho) > 0$ *such that*

$$\left|\mathbb{E}[\prod_{1\leq i\leq K} F_i] - \mathbb{E}[\prod_{1\leq i\leq K} T_{1-\gamma} F_i]\right| \leq K\epsilon.$$

Since $\rho(\overline{\Omega}^{Qk}, \overline{\mu}') \leq \sqrt{\frac{Q-1}{Q}}$, we can apply the above theorem ($K \leftarrow Qk$, $\Omega_1 = \cdots = \Omega_K \leftarrow \overline{\Omega}$, $\nu \leftarrow \overline{\mu}'$, $\epsilon \leftarrow \frac{\zeta^Q}{4K}$, $F_{k(q-1)+i} \leftarrow \overline{f_q}$ for $q \in [Q]$ and $i \in [k]$) to have $\gamma := \gamma(Q, k, \zeta) \in (0, 1)$ such that

$$\left|\mathbb{E}_{x_{q,i}}[\prod_{1\leq q\leq Q, 1\leq i\leq k} f_q(x_{q,i})] - \mathbb{E}_{x_{q,i}}[\prod_{1\leq q\leq Q, 1\leq i\leq k} \overline{T}_{1-\gamma} f_q(x_{q,i})]\right| \leq \frac{\zeta^Q}{4}. \tag{10.4}$$

From block noised functions to individual noised functions, we state the following general theorem inspired by Wenner [Wen13].

Theorem 10.4.5. *Let* $(\Omega_1^{d_1} \times \cdots \times \Omega_K^{d_K}, \nu)$ *be joint probability spaces such that the marginal of each copy of* Ω_i *is* ν_i, *and the marginal of* $\Omega_i^{d_i}$ *is* $\nu_i^{\otimes d_i}$. *Fix* $F_i : (\Omega_i^{d_i})^L \to \mathbb{R}$ *for each* $i = 1, \ldots, K$ *with an associated projection* $\pi_i : [d_i L] \to [L]$ *such that* $|\pi_i^{-1}(j)| = d_i$ *for* $1 \leq j \leq L$. *For any* $0 \leq \rho \leq 1$, *the noise operator* $T_\rho F_i$ *and the block noise operator* $\overline{T}_\rho F_i$ *under* π_i *is defined as in Section 10.2. Fix a positive integer* J *and consider* F_i^{bad} *under* π_i *and* J. *Suppose* $\max_{1\leq i\leq K} \|F_i\|_2 \leq 1$ *and* $\xi := \max_{1\leq i\leq K} \|F_i^{\mathsf{bad}}\|_2$. *Then we have,*

$$\left|\mathbb{E}_{(x_1,\ldots,x_K)\sim\nu^{\otimes L}}[\prod_{1\leq i\leq K} \overline{T}_{1-\gamma} F_i(x_i)] - \mathbb{E}_{(x_1,\ldots,x_K)\sim\nu^{\otimes L}}[\prod_{1\leq i\leq K} T_{1-\gamma} F_i(x_i)]\right| \leq 2\cdot 3^K((1-\gamma)^J + \xi).$$

Proof. For each $1 \leq i \leq K$, we decompose F_i as the follows:

$$\begin{aligned}
F_i^{\mathsf{shattered}} &= \sum_{S\subseteq[d_iL]:S \text{ shattered under } \pi_i} (F_i)_S \\
F_i^{\mathsf{large}} &= \sum_{S\subseteq[d_iL]:S \text{ not shattered and } |\pi_i(S)|\geq J} (F_i)_S \\
F_i^{\mathsf{bad}} &= \sum_{S\subseteq[d_iL]:S \text{ not shattered and } |\pi_i(S)|< J} (F_i)_S \ .
\end{aligned}$$

Consider $C := \{\mathsf{shattered}, \mathsf{large}, \mathsf{bad}\}^K$. Expanding $F_i = (F_i^{\mathsf{shattered}} + F_i^{\mathsf{large}} + F_i^{\mathsf{bad}})$, we have

$$\prod_{1 \le i \le K} \overline{T}_{1-\gamma} F_i = \sum_{c \in C} \prod_{1 \le i \le K} \overline{T}_{1-\gamma} F_i^{c_i}$$

and

$$\prod_{1 \le i \le K} T_{1-\gamma} F_i = \sum_{c \in C} \prod_{1 \le i \le K} T_{1-\gamma} F_i^{c_i}$$

The quantity we want to bound can be also decomposable as

$$\Big|\sum_{c \in C} \mathbb{E}[\prod_{1 \le i \le K} \overline{T}_{1-\gamma} F_i^{c_i} - \prod_{1 \le i \le K} T_{1-\gamma} F_i^{c_i}]\Big|.$$

Since $\overline{T}_{1-\gamma} F_i^{\mathsf{shattered}} = T_{1-\gamma} F_i^{\mathsf{shattered}}$, the contribution of the case $c = \{\mathsf{shattered}\}^K$ is 0. We bound the other two cases of c.

- $c_{i'} = \mathsf{large}$ for some i':

$$\begin{aligned} |\mathbb{E}[\prod_{1 \le i \le K} \overline{T}_{1-\gamma} F_i^{c_i}]| &\le \|\overline{T}_{1-\gamma} F_{i'}^{\mathsf{large}}\|_2 \|\prod_{i \ne i'} \overline{T}_{1-\gamma} F_i^{c_i}\|_2 \\ &\le (1-\gamma)^J \|F_{i'}^{\mathsf{large}}\|_2 \le (1-\gamma)^J . \end{aligned}$$

  Similarly, $|\mathbb{E}[\prod_{1 \le i \le K} T_{1-\gamma} F_i^{c_i}]| \le (1-\gamma)^J$ and the contribution from such c is at most $2(1-\gamma)^J$.

- $c_{i'} = \mathsf{bad}$ for some i':

$$|\mathbb{E}[\prod_{1 \le i \le K} \overline{T}_{1-\gamma} F_i^{c_i}]| \le \|\overline{T}_{1-\gamma} F_{i'}^{\mathsf{bad}}\|_2 \|\prod_{i \ne i'} \overline{T}_{1-\gamma} F_i^{c_i}\|_2 \le \xi .$$

  Similarly, $|\mathbb{E}[\prod_{1 \le i \le K} T_{1-\gamma} F_i^{c_i}]| \le \xi$ and the contribution from such c is at most 2ξ.

Since there are at most 3^K choices for c, the total error is bounded by $2 \cdot 3^K((1-\gamma)^J + \xi)$. $\square$

By applying the above theorem with $K \leftarrow Qk$, $L \leftarrow L$, $\Omega_1, \dots, \Omega_K \leftarrow \Omega$, $d_1, \dots, d_K \leftarrow d$, $\nu \leftarrow \overline{\mu}'$, $F_{k(q-1)+1} = \cdots = F_{k(q-1)+k} \leftarrow f_q$, $\pi_{k(q-1)+1} = \cdots = \pi_{k(q-1)+k} \leftarrow \pi_q$, $\xi \leftarrow (\frac{J^2}{T})^{1/4}$, we have

$$\Big|\mathbb{E}_{x_{q,i}}[\prod_{1 \le q \le Q, 1 \le i \le k} \overline{T}_{1-\gamma} f_q(x_{q,i})] - \mathbb{E}_{x_{q,i}}[\prod_{1 \le q \le Q, 1 \le i \le k} T_{1-\gamma} f_q(x_{q,i})]\Big| \le 2 \cdot 3^{Qk}((1-\gamma)^J + (\frac{J^2}{T})^{1/4}).$$

Fixing J and T to satisfy $2 \cdot 3^{Qk}((1-\gamma)^J + (\frac{J^2}{T})^{1/4}) \leq \frac{\zeta^Q}{4}$ as well as the previous constraint, and combining with (10.4), we can conclude that

$$\left| \mathbb{E}_{x_{q,i}}[\prod_{1 \leq q \leq Q, 1 \leq i \leq k} f_q(x_{q,i})] - \mathbb{E}_{x_{q,i}}[\prod_{1 \leq q \leq Q, 1 \leq i \leq k} T_{1-\gamma} f_q(x_{q,i})] \right| \leq \frac{\zeta^Q}{2}. \tag{10.5}$$

In particular, if I induces less than $\frac{\zeta^Q}{4}$ fraction of hyperedges formed from e, from (10.5), we have

$$\mathbb{E}_{x_{q,i}}[\prod_{1 \leq q \leq Q, 1 \leq i \leq k} T_{1-\gamma} f_q(x_{q,i})] \leq \frac{3\zeta^Q}{4}. \tag{10.6}$$

STEP 4. **Invariance.** We now want to show

$$\mathbb{E}_{x_{q,i}}[\prod_{1 \leq q \leq Q, 1 \leq i \leq k} T_{1-\gamma} f_q(x_{q,i})] \approx \prod_{1 \leq q \leq Q} \mathbb{E}_{x_{q,i}}[\prod_{1 \leq i \leq k} T_{1-\gamma} f_q(x_{q,i})]],$$

unless f_q's share influential coordinates. Our invariance principle is similar to ones used in Wenner [Wen13] and Chan [Cha13]. With the goal of showing

$$\mathbb{E}_{x_1, \dots, x_K}[\prod_{1 \leq i \leq K} F_i(x_i)] \approx \mathbb{E}_{x_1}[F_1(x_1)] \mathbb{E}[\prod_{2 \leq i \leq K} F_i(x_i)],$$

one crucial property they used is that x_1 is independent of x_i for each $i = 2, \dots, K$ (even though any three x_i's are dependent).

Our $(x_{q,i})$ do not have such a property (any $x_{q,i}$ is dependent on $x_{q,i'}$ for $i \neq i'$), but it satisfies another property that any $x_{q,i}$ is independent of the joint distribution of $(x_{q',i'})_{q' \neq q, i' \in [k]}$ — everything not in the same hypercube. This property allows us to achieve the goal stated above.

The following lemma is the basic building block that enables the induction used in proof of the main invariance principle (Theorem 10.4.7) used in our framework. It is essentially implied by a theorem stated in a more general setup by Wenner [Wen13, Theorem 3.12]. For completeness, we present a proof below in simpler notation that fits for our purposes.

Lemma 10.4.6. *Let $(\Omega_1^k \times \Omega_2, \nu)$ be $(k+1)$ correlated spaces $(k \geq 2)$ such that each copy of Ω_1 has the same marginal, and any one copy of Ω_1 and Ω_2 are independent. Let $F \in \mathcal{L}_{[0,1]}(\Omega_1^L)$, and $G \in \mathcal{L}(\Omega_2^L)$. Suppose that $\sum_{1 \leq j \leq L} \mathsf{Inf}_j[F] \leq \Gamma$ and*

$$\sum_{1 \leq j \leq L} \mathsf{Inf}_j[F] \mathsf{Inf}_j[G] \leq \tau.$$

Then,

$$\left|\mathbb{E}_{x_1,\ldots,x_k,y}[\prod_{1\le i\le k} F(x_i)G(y)] - \mathbb{E}_{x_1,\ldots,x_k,y}[\prod_{1\le i\le k} F(x_i)]\mathbb{E}_y[G(y)]\right| \le 2^{k+1}\sqrt{\Gamma\tau}.$$

Proof. Let ν' be the distribution where the marginals of Ω_1^k and Ω_2 are the same as those of ν, but Ω_1^k and Ω_2 are independent. Fix $j \in [L]$. Let $(x_1,\ldots,x_k,y)$ be sampled such that $((x_1)_{j'},\ldots,(x_k)_{j'},y_{j'}) \sim \nu$ for $j' < j$ and $((x_1)_{j'},\ldots,(x_k)_{j'},y_{j'}) \sim \nu'$ for $j' \ge j$. Let $(x'_1,\ldots,x'_k,y')$ be the same except that $((x'_1)_j,\ldots,(x'_k)_j,y_j) \sim \nu$. We want to bound

$$\left|\mathbb{E}_{x_1,\ldots,x_k,y}[\prod_{1\le i\le k} F(x_i)G(y)] - \mathbb{E}_{x'_1,\ldots,x'_k,y'}[\prod_{1\le i\le k} F(x'_i)G(y')]\right|,$$

since the LHS with $j = 1$ and the RHS with $j = L$ are the two expectations we are interested in.

Decompose F into the following two parts.

$$F^{\mathsf{relevant}} = \sum_{S:j\in S} F_S$$
$$F^{\mathsf{not}} = \sum_{S:j\notin S} F_S$$

Note that $\|F^{\mathsf{relevant}}\|_2^2 = \mathsf{Inf}_j[F]$. Decompose $G = G^{\mathsf{relevant}} + G^{\mathsf{not}}$ in the same way. Let $C = \{\mathsf{relevant}, \mathsf{not}\}^{k+1}$. The term we wanted to bound now becomes

$$\left|\sum_{c\in C}\Big(\mathbb{E}_{x_1,\ldots,x_k,y}[\prod_{1\le i\le k} F^{c_i}(x_i)G^{c_{k+1}}(y)] - \mathbb{E}_{x'_1,\ldots,x'_k,y'}[\prod_{1\le i\le k} F^{c_i}(x'_i)G^{c_{k+1}}(y')]\Big)\right|. \quad (10.7)$$

If $c_{k+1} = \mathsf{not}$ or $c_1 = \cdots = c_k = \mathsf{not}$, the contribution from c is zero because the marginals of $((x_1)_j,\cdots,(x_k)_j)$ and y_j are the same with those of $((x'_1)_j,\ldots,(x'_k)_j)$ and y'_j respectively. Furthermore, the same conclusion holds when $c_{k+1} = \mathsf{relevant}$ and exactly one of $c_1,\ldots,c_k$ is $\mathsf{relevant}$, since one copy of Ω_1 and Ω_2 are independent and $((x_i)_j, y_j)$ and $((x'_i)_j, y'_j)$ have the same distribution. Thus a $c \in C$ with nonzero contribution to

(10.7) must satisfy $c_{i_1} = c_{i_2} = c_{k+1} = \textsf{relevant}$ for some $i_1 \neq i_2$. For such c,

$$
\begin{aligned}
&\left|\mathbb{E}_{x_1,\ldots,x_k,y}[\prod_{1\le i\le k} F^{c_i}(x_i)G^{c_{k+1}}(y)]\right| \\
&\le \|F^{\textsf{relevant}}(x_{i_1})G^{\textsf{relevant}}(y)\|_2\|F^{\textsf{relevant}}(x_{i_2})\|_2\|\prod_{i\neq i_1,i_2} F^{c_i}\|_\infty \quad \text{By Hölder inequality} \\
&= \|F^{\textsf{relevant}}\|_2\|G^{\textsf{relevant}}\|_2\|F^{\textsf{relevant}}\|_2\|\prod_{i\neq i_1,i_2} F^{c_i}\|_\infty \quad \text{By independence} \\
&\le \sqrt{\textsf{Inf}_j[F]^2\textsf{Inf}_j[G]},
\end{aligned}
$$

where the last inequality used the fact that $F^{\textsf{not}}(x) = \mathbb{E}_{x'}[F(x')|x'_{[L]\setminus j} = x_{[L]\setminus j}] \in [0,1]$ and $F^{\textsf{relevant}}(x) = F(x) - F^{\textsf{not}}(x) \in [-1,1]$. There are at most 2^k choices for such c and

$$
\left|\mathbb{E}_{x'_1,\ldots,x'_k,y}[\prod_{1\le i\le k} F^{c_i}(x'_i)G^{c_{k+1}}(y')]\right| \le \sqrt{\textsf{Inf}_j[F]^2\textsf{Inf}_j[G]}
$$

can be shown similarly, so

$$
\left|\mathbb{E}_{x_1,\ldots,x_k,y}[\prod_{1\le i\le k} F(x_i)G(y)] - \mathbb{E}_{x'_1,\ldots,x'_k,y'}[\prod_{1\le i\le k} F(x'_i)G(y')]\right| \le 2^{k+1}\sqrt{\textsf{Inf}_j[F]^2\textsf{Inf}_j[G]}.
$$

Summing over all $1 \le j \le J$, we conclude that

$$
\begin{aligned}
&\left|\mathbb{E}_{x_1,\ldots,x_k,y}[\prod_{1\le i\le k} F(x_i)G(y)] - \mathbb{E}_{x_1,\ldots,x_k}[\prod_{1\le i\le k} F(x_i)]\mathbb{E}_y[G(y)]\right| \\
&\le 2^{k+1}\sum_{1\le j\le L}\sqrt{\textsf{Inf}_j[F]^2\textsf{Inf}_j[G]} \\
&\le 2^{k+1}\sqrt{\sum_{1\le j\le L}\textsf{Inf}_j[F]\textsf{Inf}_j[G]}\sqrt{\sum_{1\le j\le L}\textsf{Inf}_j[F]} \qquad \text{(by Cauchy-Schwartz)} \\
&\le 2^{k+1}\sqrt{\Gamma\tau}\ .
\end{aligned}
$$

□

Given the lemma, we formalize our intuition and prove the following general theorem, which will also be used in our other results.

Theorem 10.4.7. *Let $(\Omega_1^{k_1}\times\cdots\times\Omega_Q^{k_Q}, \nu)$ be correlated spaces ($k_1,\ldots,k_{Q-1} \ge 2$, $k_Q \ge 1$) where each copy of Ω_q has the same marginal and independent of $\prod_{q'\neq q}\Omega_{q'}^{k_q}$. Let $k_{max} =$*

$\max_q k_q$ *and* $k_{sum} = \sum_q k_q$*. For* $1 \le q \le Q$*, let* $F_q \in \mathcal{L}_{[0,1]}(\Omega_q^L)$*. Suppose that for all* $1 \le q < Q$*,* $\sum_{1\le j\le L} \mathsf{Inf}_j[F_q] \le \Gamma$ *and*

$$\sum_{1\le j\le L} \mathsf{Inf}_j[F_q](\mathsf{Inf}_j[F_{q+1}] + \cdots + \mathsf{Inf}_j[F_Q]) \le \tau.$$

Then,

$$\left|\mathbb{E}_{x_{q,i}}[\prod_{1\le q\le Q, 1\le i\le k_q} F_q(x_{q,i})] - \prod_{1\le q\le Q} \mathbb{E}_{x_{q,i}}[\prod_{1\le i\le k_q} F_q(x_{q,i})]\right| \le Q\cdot 2^{k_{max}+1}\sqrt{\Gamma k_{sum}^2 \tau}.$$

Proof. . We use induction on Q. When $Q = 2$, the application of Lemma 10.4.6 (setting $F \leftarrow F_1$, $k \leftarrow k_1$, $\Omega_2 \leftarrow \Omega_2^{k_2}$, $G(x_{2,1}, \ldots, x_{2,k_2}) \leftarrow \prod_{1\le i\le k_2} F_2(x_{2,i})$) and applying Lemma 10.2.2 to have $\mathsf{Inf}_j[G] \le k_2^2 \mathsf{Inf}_j[F_2]$) implies the theorem.

Assuming the theorem holds for $Q-1$, the application of Lemma 10.4.6 with

- $F \leftarrow F_1$, $k \leftarrow k_1$, $\Omega_2 \leftarrow \Omega_2^{k_2} \times \cdots \times \Omega_Q^{k_Q}$, $G(x_{q,i}) \leftarrow \prod_{2\le q\le Q, 1\le i\le k_2} F_q(x_{q,i})$
- $\mathsf{Inf}_j[G] \le k_{sum}^2(\mathsf{Inf}_j[F_2] + \cdots + \mathsf{Inf}_j[F_Q])$ by Lemma 10.2.2

gives

$$\begin{aligned}
&\left|\mathbb{E}_{x_{q,i}}[\prod_{1\le q\le Q, 1\le i\le k_q} F_q(x_{q,i})] - \prod_{1\le q\le Q} \mathbb{E}_{x_{q,i}}[\prod_{1\le i\le k_q} F_q(x_{q,i})]\right| \\
\le&\left|\mathbb{E}_{x_{q,i}}[\prod_{1\le q\le Q, 1\le i\le k_q} F_q(x_{q,i})] - \mathbb{E}_{x_{1,i}}[\prod_{1\le i\le k_1} F_1(x_{1,i})]\mathbb{E}_{x_{q,i}}[\prod_{2\le q\le Q, 1\le i\le k_q} F_q(x_{q,i})]\right| \\
+&\left|\prod_{1\le q\le Q} \mathbb{E}_{x_{q,i}}[\prod_{1\le i\le k_q} F_q(x_{q,i})] - \mathbb{E}_{x_{1,i}}[\prod_{1\le i\le k_1} F_1(x_{1,i})]\mathbb{E}_{x_{q,i}}[\prod_{2\le q\le Q, 1\le i\le k_q} F_q(x_{q,i})]\right| \\
\le&2^{k_{max}+1}\sqrt{\Gamma k_{sum}^2\tau} + (Q-1)2^{k_{max}+1}\sqrt{\Gamma k_{sum}^2 \tau} \\
=&Q\cdot 2^{k_{max}+1}\sqrt{\Gamma k_{sum}^2\tau}\ .
\end{aligned}$$

□

By Lemma 1.13 of Wenner [Wen13], there exists $\Gamma = O(\frac{1}{\gamma})$ such that

$$\sum_{1\le j\le L} \overline{\mathsf{Inf}}_j[T_{1-\gamma} f_q] \le \sum_{1\le j\le R} \mathsf{Inf}_j[T_{1-\gamma} f_q] \le \Gamma.$$

Fix τ to satisfy $Q \cdot 2^{k+1}\sqrt{\Gamma(Qk)^2\tau} < \frac{\zeta^Q}{4}$. We have

$$\begin{aligned}
&\left|\mathbb{E}_{x_{q,i}}[\prod_{1\le q\le Q, 1\le i\le k} T_{1-\gamma}f_q(x_{q,i})] - \prod_{1\le q\le Q}\mathbb{E}_{x_{q,i}}[\prod_{1\le i\le k} T_{1-\gamma}f_q(x_{q,i})]\right| \\
\ge& \left|\prod_{1\le q\le Q}\mathbb{E}_{x_{q,i}}[\prod_{1\le i\le k} T_{1-\gamma}f_q(x_{q,i})]\right| - \left|\mathbb{E}_{x_{q,i}}[\prod_{1\le q\le Q, 1\le i\le k} T_{1-\gamma}f_q(x_{q,i})]\right| \\
\ge& \frac{\zeta^Q}{4} \quad \text{by (10.3) and (10.6) .}
\end{aligned}$$

Thus, applying Theorem 10.4.7 with $Q \leftarrow Q$, $k_1 = \cdots = k_Q \leftarrow k$,$\Omega_1 = \cdots = \Omega_Q = \overline{\Omega}$, $\nu \leftarrow \overline{\mu}'$, $L \leftarrow L$, $F_q \leftarrow \overline{T_{1-\gamma}f_q}$, $\mathsf{Inf}_j[F_q] \leftarrow \overline{\mathsf{Inf}}_j[T_{1-\gamma}f_q]$, there exists $q \in \{1, \ldots, Q-1\}$ such that

$$\sum_{1\le j\le L}\overline{\mathsf{Inf}}_j[T_{1-\gamma}f_q](\overline{\mathsf{Inf}}_j[T_{1-\gamma}f_{q+1}] + \cdots + \overline{\mathsf{Inf}}_j[T_{1-\gamma}f_Q]) > \tau. \tag{10.8}$$

STEP 5. **Decoding Strategy.** We use the standard strategy — each v_q samples a set $S \subseteq [R]$ according to $\|(f_q)_S\|_2^2$, and chooses a random element from S. For each $1 \le j \le L$, the probability that v chooses a label in $\pi^{-1}(j)$ is

$$\begin{aligned}
\sum_{S:S\cap\pi^{-1}(j)\neq\emptyset}\|(f_q)_S\|_2^2\frac{|S\cap\pi^{-1}(j)|}{|S|} &\ge \sum_{S:S\cap\pi^{-1}(j)\neq\emptyset}\|(f_q)_S\|_2^2\cdot\gamma(1-\gamma)^{\frac{|S|}{|S\cap\pi(j)|}} \\
&\ge \gamma\sum_{S:S\cap\pi^{-1}(j)\neq\emptyset}\|(f_q)_S\|_2^2\cdot(1-\gamma)^{|S|} \\
&= \gamma\overline{\mathsf{Inf}}_j[T_{1-\gamma}f_q]
\end{aligned}$$

where the first inequality follows from the fact that $\alpha \ge \gamma(1-\gamma)^{1/\alpha}$ for $\alpha > 0$ and $0 < \gamma < 1$. Fix q to be the one obtained in Step 4 that satisfies (10.8). The probability that $\pi_q(l(v_q)) = \pi_{q'}(l(v_{q'}))$ for some $q < q' \le Q$ is at least

$$\begin{aligned}
&\gamma^2\sum_{1\le j\le L}\overline{\mathsf{Inf}}_j[T_{1-\gamma}f_q]\max_{q<q'\le Q}\overline{\mathsf{Inf}}_j[T_{1-\gamma}f_{q'}] \\
\ge& \frac{\gamma^2}{Q}\sum_{1\le j\le L}\overline{\mathsf{Inf}}_j[T_{1-\gamma}f_q](\overline{\mathsf{Inf}}_j[T_{1-\gamma}f_{q+1}] + \cdots + \overline{\mathsf{Inf}}_j[T_{1-\gamma}f_Q]) \\
\ge& \frac{\gamma^2\tau}{Q}.
\end{aligned}$$

Suppose that the total fraction of hyperedges (of E') wholly contained within I is less than $\frac{\delta}{4}\cdot\frac{\zeta^Q}{4} = \epsilon^{O_{Q,k}(1)}$. Since $\frac{\delta}{2}$ fraction of hyperedges (of E) are good, for at least $\frac{\delta}{2} - \frac{\delta}{4} = \frac{\delta}{4}$

fraction of hyperedges the above analysis works, and these edges are weakly satisfied by the above randomized strategy with probability $\frac{\gamma^2\tau}{Q}$. Setting the soundness parameter in Theorem 10.2.3 $\eta := \frac{\delta}{4} \cdot \frac{\gamma^2\tau}{Q}$ completes the proof of the soundness Lemma 10.7.2, and therefore also Theorem 9.2.1.

Dependencies between Constants The above proof involves several constants that depend on each other. We summarize them in Table 10.1, in the order they are fixed in the proof.

Constants	How it is fixed	When it is fixed
Q, k, ϵ	Arbitrary $Q, k \geq 2$, $\epsilon > 0$	
δ	$\delta := \frac{(\frac{\epsilon}{2})^Q}{2^{Q+1}}$	STEP 1.
ζ	$\zeta := \zeta(Q, \epsilon, k)$ (by Theorem 10.3.8)	STEP 2.
γ	$\gamma := \gamma(Q, k, \zeta)$ (by Theorem 10.4.4)	STEP 3.
J, T	Large enough to satisfy $Q(\frac{J^2}{T})^{1/4} \leq \frac{\delta}{2}$ $2 \cdot 3^{Qk}((1-\gamma)^J + (\frac{J^2}{T})^{1/4}) \leq \frac{\zeta^Q}{4}$	STEP 3.
Γ	$\Gamma := O(\frac{1}{\gamma})$ (by [Wen13])	STEP 4.
τ	Small enough to satisfy $Q \cdot 2^{k+1}\sqrt{\Gamma(Qk)^2\tau} < \frac{\zeta^Q}{4}$	STEP 4.
η	$\eta := \frac{\delta}{4} \cdot \frac{\gamma^2\tau}{Q}$	STEP 5.

Table 10.1: List of the constants in the proof.

Requirements for Distributions. In the proof, we used the following three properties of the test distribution. We qualitatively describe them and how they are used in the proof. All the distributions in this part satisfy all the properties. Let $(\Omega_1 \times \cdots \times \Omega_K, \nu)$ be the test distribution for K points.

1. Let $\Omega_{i_1}, \ldots, \Omega_{i_k}$ correspond to k points queried in the same hypercube. We require the marginal distribution on $\Omega_{i_1} \times \cdots \times \Omega_{i_k}$ to have the *full support* — any $(x_{i_1}, \ldots, x_{i_k}) \in \Omega_{i_1} \times \cdots \times \Omega_{i_k}$ is sampled with nonzero probability. It is crucial in our application of the reverse hypercontractivity used in STEP 2.

2. When $(x_1, \ldots, x_K)$ is sampled from ν, we require that for any $i \in [K]$, x_i is not always determined by the other $K-1$ points. This is used when bounding correlations and smoothing functions in STEP 3.

3. When $(x_1, \ldots, x_K)$ is sampled from ν, we require that for any $i \in [K]$, x_i is completely independent from all the points not in the same hypercube. It is used in the application of the invariance principle in STEP 4.

10.5 k-HYPERGRAPH VERTEX COVER

In this section, we prove the following two theorems, both implying that it is NP-hard to approximate k-HYPERGRAPH VERTEX COVER with in a factor of $K - 1 - \epsilon$.

Theorem 10.5.1 (Restatement of Theorem 9.2.4)**.** *For any $\epsilon > 0$ and $K \geq 3$, given a K-uniform hypergraph $H = (V, E)$, it is NP-hard to distinguish the following cases.*

- *Completeness: There is a vertex cover of measure $\frac{1}{K-1}$.*
- *Soundness: Every $I \subseteq V$ of measure ϵ induces at least a fraction $\epsilon^{O_K(1)}$ of hyperedges.*

Theorem 10.5.2 (Restatement of Theorem 9.2.5)**.** *For any $\epsilon > 0$ and $K \geq 3$, given a K-uniform hypergraph $H = (V, E)$, it is NP-hard to distinguish the following cases.*

- *Completeness: There exist $V^* \subseteq V$ of measure ϵ and a coloring $c : [V \setminus V^*] \to [K-1]$ such that for every hyperedge of the induced hypergraph on $V \setminus V^*$, $K-2$ colors appear once and the other color twice. Therefore, H has a vertex cover of size at most $\frac{1}{K-1} + \epsilon$.*
- *Soundness: There is no independent set of measure ϵ.*

The above two theorems are not comparable to each other. In the completeness case, Theorem 9.2.4 ensures a smaller vertex cover, while Theorem 9.2.5 guarantees richer structure. In the soundness case, Theorem 9.2.4 gives a stronger density. Since they differ only in the test distribution, we prove Theorem 9.2.5 in details and introduce the distribution for Theorem 9.2.4 at the end of this section.

10.5.1 Distribution

We first define the distribution of K points, one in a single cell and the other $K - 1$ in a block of size d. Let $\Omega = \{*, 1, \ldots, K-1\}$ and $\overline{\Omega} = \Omega^d$. Let ω be the distribution on Ω such that $\omega(*) = \epsilon$ and $\omega(1) = \cdots = \omega(K-1) = \frac{1-\epsilon}{K-1}$. The K points $x \in \Omega$ and $y_1, \ldots, y_{K-1} \in \overline{\Omega}$ are sampled by the following procedure.

- Sample $x \sim \omega$.

- If $x = *$, sample $y_1, \ldots, y_{K-1} \sim \omega^{\otimes d}$ independently.

- If $x \neq *$, for each $1 \leq j \leq d$, sample $(y_1)_j, \ldots, (y_{K-1})_j \sim \mathbb{S}_{K-1}$ uniformly, and independently noise $(y_i)_j \leftarrow *$ with probability ϵ.

It is easy to see that the marginal distribution of each y_i is $\omega^{\otimes d}$. Let $(\Omega \times \overline{\Omega}^{K-1}, \overline{\mu}')$ denote the K correlated spaces corresponding to the above distribution, and let $\overline{\mu}$ denote the marginal distribution of $(y_1, \ldots, y_{K-1})$. Let $\overline{\Omega}_i$ $(1 \leq i \leq K-1)$ denote the copy of $\overline{\Omega}$ associated with y_i, and $\overline{\Omega}_i'$ be the product of the other $K-1$ spaces. With probability ϵ (when $x = *$), y_i is completely independent of the others. Even when $x \neq *$, y_i's marginal is $\omega^{\otimes d}$. By Lemma 10.2.1, we conclude that $\rho(\overline{\Omega}_i, \overline{\Omega}_i'; \overline{\mu}') \leq \sqrt{1-\epsilon}$.

However, bounding $\rho(\Omega, \overline{\Omega}^{K-1}; \overline{\mu}')$ (as the correlation between two spaces Ω and $\overline{\Omega}^{K-1}$) cannot be done in the same way. To get around this, we define the distribution $\overline{\mu}'_\beta$ be the same as $\overline{\mu}'$, but at the end each y_i is independently resampled with probability $1-\beta$. While we still use $\overline{\mu}'$ in the reduction, the fact that $\rho(\overline{\Omega}_i, \overline{\Omega}_i'; \overline{\mu}') \leq \sqrt{1-\epsilon}$ implies that our analysis, without much loss, can assume that each y_i is resampled as in $\overline{\mu}'_\beta$. In $\overline{\mu}'_\beta$, the same technique yields $\rho(\Omega, \overline{\Omega}^{K-1}; \overline{\mu}'_\beta) \leq \sqrt{1-(1-\beta)^{K-1}}$, which allows the usual analysis to proceed.

10.5.2 Reduction and Completeness

We now describe the reduction from MULTILAYERED LABEL COVER with A layers. Given a $G = (\cup_{1 \leq i \leq A} V_i, \cup_{i<j} E_{i,j})$ with a projection $\pi_e : [R_j] \to [R_i]$ for each hyperedge $e = (u, v)$ $(u \in V_i, v \in V_j)$, the resulting instance for k-HYPERGRAPH VERTEX COVER is (V', E'), where $V' = \cup_{1 \leq i \leq A} V_i \times \Omega^{R_i}$. The weight of (v, x) $(v \in V_i)$ is $\prod_{1 \leq j \leq R_i} \omega(x_j)$, so that the sum of the weights of the vertices in $\mathsf{Cloud}(v)$ is 1. For $v \in V_i$, let $\mathsf{Cloud}(v) := \{v\} \times \Omega^{R_i}$. The set of hyperedges E' is described by the following procedure.

- Sample $1 \leq a < b \leq A$ uniformly and $e = (u, v) \in E_{i,j}$ such that $u \in V_i, v \in V_j$.

- Sample $x \in \Omega^{R_a}, y_1, \ldots, y_{K-1} \in \Omega^{R_b}$ in the following way. For each $1 \leq j \leq R_a$, sample $x_j, ((y_i)_{\pi_e^{-1}(j)})_{i \in [K-1]}$ from $(\Omega \times \overline{\Omega}^{K-1}, \overline{\mu}')$.

- Add a hyperedge $((u,x),(v,y_1),\ldots,(v,y_{K-1}))$ to E'. We say that this hyperedge is *formed from* e, and the weight of this hyperedge is the probability that it is sampled given that e is sampled in the first step.

Given the reduction, completeness is easy to show.

Lemma 10.5.3. *If there is a labeling that satisfies every* $e \in E$*, there exist* $V^* \subseteq V'$ *of measure* ϵ *and* $c : V' \setminus V^* \to [K-1]$ *with the same measure for each color, such that in each hyperedge induced by* $V' \setminus V^*$*,* $K-1$ *colors appear once and the other color appears twice.*

Proof. Let $l : V \to [R_A]$ be a labeling that satisfies every edge in E. Let $V^* := \{(v,x) : (x)_{l(v)} = *\}$, and $c(v,x) = (x)_{l(v)}$. In each $\mathsf{Cloud}(v)$, V^* contains measure $\omega(*) = \epsilon$ and $c(i)$ contains $\omega(i) = \frac{1-\epsilon}{K-1}$. For each hyperedge $((u,x),(v,y_1),\ldots,(v,y_{K-1}))$ induced by $V' \setminus V^*$, $\{(v,y_1)_{l(v)},\ldots,(v,y_{K-1})_{l(v)}\} = [K-1]$. $\square$

10.5.3 Soundness

Unlike the previous reductions, the resulting instance is weighted — vertices and hyperedges can have different weights. The only reason is that (1) we used MULTILAYERED LABEL COVER and (2) and ω is not the uniform distribution. Once we fix an edge e of G, our hyperedge weights correspond to the above probability distribution and vertex weights correspond to its marginals. Therefore all the following probabilistic analysis works as in previous reductions.

Lemma 10.5.4. *For any* $\epsilon > 0$*, there exists* $\eta := \eta(\epsilon, K)$ *such that if* $I \subseteq V'$ *of measure* ϵ *induces less than* $\epsilon^{O_{Q,k}(1)}$ *fraction of hyperedges, the corresponding instance of* MULTILAYERED LABEL COVER *admits a labeling that satisfies* η *fraction of edges in* $E_{a,b}$ *for some* $1 \le a < b \le A$.

The proof is almost identical to the one presented in Section 10.7.2, with slightly more technical details dealing with noise.

STEP 1. **Fixing a Good Hyperedge.** Let $I \subseteq V'$ be of measure ϵ. Let f_v be the indicator function of $I \cap \mathsf{Cloud}(v)$. By averaging, $\frac{\epsilon}{2}$ fraction of vertices has $\mathbb{E}[f_v] \ge \frac{\epsilon}{2}$ — call these vertices *heavy*. Let $W_i \subseteq V_i$ be the set of heavy vertices in the ith layer.

By averaging, at least $\frac{\epsilon}{4}$ fraction of layers satisfy $|W_i| \geq \frac{\epsilon}{4}|V_i|$. Take $A = \lceil \frac{\epsilon}{16} \rceil$. By weak density, there exist $1 \leq a < b \leq A$ such that the fraction of edges in $E_{i,j}$ induced by W_a and W_b is at least $\frac{\epsilon^3}{1024}$. Let $L = R_a$ and $R = R_b$.

By the same argument as in Section 10.7.2, by adjusting the smoothness paramter T and an integer J, we can ensure that $\frac{\epsilon^3}{2048}$ fraction of edge $(u, v) \in E_{a,b}$ is good — both u and v are heavy and,

$$\|f_v^{\mathsf{bad}}\|_2 \leq (\frac{J^2}{T})^{1/4}$$

under π_e and J.

Throughout the rest of the section, fix such an edge $e = (u, v)$ and the associated projections $\pi := \pi_e$. For simplicity, let $f := f_u$ and $g := f_v$. We now measure the weight of hyperedges induced by I, which is

$$\mathbb{E}_{x,y_1,\ldots,y_{K-1}}[f(x) \prod_{1 \leq i \leq K-1} g(y_i)] \tag{10.9}$$

STEP 2. **Lower Bounding in Each Hypercube.** For each $1 \leq j \leq L$, with probability ϵ, $(y_i)_{\pi^{-1}(j)}$ are sampled completely independently from $\overline{\Omega}$. By Theorem 10.3.8 (setting $\Omega \leftarrow \overline{\Omega}$, $k \leftarrow K-1$, $\sigma \leftarrow \omega^{\otimes d}$, $\nu \leftarrow \overline{\mu}$, $\rho \leftarrow 1-\epsilon$, $F_1 = \cdots = F_{K-1} \leftarrow g$, $\epsilon \leftarrow \frac{\epsilon}{2}$), there exists $\zeta = \zeta(\epsilon, K) > 0$ such that for every $\gamma \in [0, 1]$,

$$\mathbb{E}_{y_1,\ldots,y_K \sim \overline{\mu}^{\otimes L}}[\prod_{1 \leq i \leq K-1} T_{1-\gamma} g(y_i)] \geq \zeta.$$

Note that $\overline{\mu}_\beta$ also satisfies the requirement of Theorem 10.3.8, so

$$\mathbb{E}_{y_1,\ldots,y_K \sim (\overline{\mu}_\beta)^{\otimes L}}[\prod_{1 \leq i \leq K-1} T_{1-\gamma} g(y_i)] \geq \zeta. \tag{10.10}$$

Let $\theta := \frac{\epsilon\zeta}{2}$ be the lower bound of $\mathbb{E}[f(x)]\mathbb{E}[\prod_i g(y_i)]$, which also holds for any noised versions of f, g and noised distributions.

STEP 3. **Smoothing Functions.** Due to the fact that $\rho(\Omega, \overline{\Omega}^{K-1}; \overline{\mu}')$ is not easily bounded, we insert the noise operator for $g(y_1), \ldots, g(y_{K-1})$ first using $\rho(\overline{\Omega}_i, \overline{\Omega}'_i; \overline{\mu}') \leq \sqrt{1-\epsilon}$ for $1 \leq i \leq K-1$. This follows from the following lemma from Mossel [Mos10], which is indeed the main lemma for Theorem 10.4.4.

Lemma 10.5.5 ([Mos10]). *Let $(\Omega_1 \times \Omega_2, \nu)$ be two correlated spaces with $\rho(\Omega_1, \Omega_2; \nu) \leq \rho < 1$, and the corresponding product spaces $((\Omega_1)^L \times (\Omega_2)^L, \nu^{\otimes L})$, and $F_i \in \mathcal{L}((\Omega_i)^L)$ for $i = 1, 2$ such that $\mathsf{Var}[F_i] \leq 1$. For any $\epsilon > 0$, there exists $\gamma := \gamma(\epsilon, \rho) > 0$ such that*

$$|\mathbb{E}[F_1 F_2] - \mathbb{E}[F_1 T_{1-\gamma} F_2] \leq \epsilon.$$

Applying the above lemma to $(\overline{\Omega}_i, \overline{\Omega}_i'; \overline{\mu}')$ iteratively for $i = 1, \ldots, K-1$, we have $\gamma_1 := \gamma_1(\epsilon, K, \theta)$ such that

$$\begin{aligned}
&\Big|\mathbb{E}_{x, y_i \sim \overline{\mu}'^{\otimes L}}[f(x) \prod_{1 \leq i \leq K-1} g(y_i)] - \mathbb{E}_{x, y_i \sim \overline{\mu}'^{\otimes L}}[f(x) \prod_{1 \leq i \leq K-1} \overline{T}_{1-\gamma_1} \overline{T}_{1-\gamma_1} g(y_i)]\Big| \\
=&\Big|\mathbb{E}_{x, y_i \sim \overline{\mu}'^{\otimes L}}[f(x) \prod_{1 \leq i \leq K-1} g(y_i)] - \mathbb{E}_{x, y_i \sim (\overline{\mu}'_{1-\gamma_1})^{\otimes L}}[f(x) \prod_{1 \leq i \leq K-1} \overline{T}_{1-\gamma_1} g(y_i)]\Big| \\
\leq&\frac{\theta}{8}.
\end{aligned}$$

Let $\beta := 1 - \gamma_1$, and use $\mathbb{E}$ to denote the expectation over $(x, y_1, \ldots, y_K) \sim (\overline{\mu}'_\beta)^{\otimes L}$ while $\mathbb{E}$ still denotes the expectation over $(x, y_1, \ldots, y_K) \sim \overline{\mu}'^{\otimes L}$. This implies

$$\mathbb{E}[f(x) \prod_{1 \leq i \leq K-1} \overline{T}_{1-\gamma_1} \overline{T}_{1-\gamma_1} g(y_i)] = \mathbb{E}[f(x) \prod_{1 \leq i \leq K-1} \overline{T}_{1-\gamma_1} g(y_i)].$$

Since $\rho(\Omega, \overline{\Omega}^{K-1}; \overline{\mu}'_\beta) \leq \sqrt{1 - (1-\beta)^{K-1}}$, another application of Lemma 10.5.5 will give γ_2 such that

$$\Big|\mathbb{E}[f(x) \prod_{1 \leq i \leq K-1} \overline{T}_{1-\gamma_1} g(y_i)] - \mathbb{E}[T_{1-\gamma_2} f(x) \prod_{1 \leq i \leq K-1} \overline{T}_{1-\gamma_1} g(y_i)]\Big| \leq \frac{\theta}{8}.$$

By applying Theorem 10.4.5 ($K \leftarrow K$, $L \leftarrow L$, $\Omega_1, \ldots, \Omega_K \leftarrow \Omega$, $\Omega_K = \Omega$, $d_1, \ldots, d_{K-1} \leftarrow d$, $d_K = 1$, $\nu \leftarrow \overline{\mu}'_\beta$, $F_1 = \cdots = F_{K-1} \leftarrow g$, $F_K \leftarrow f$, $\pi_1 = \cdots = \pi_{K-1} = \pi$, $\pi_K \leftarrow$ the identity, $\xi \leftarrow (\frac{J^2}{T})^{1/4}$), we have

$$\begin{aligned}
&\Big|\mathbb{E}[T_{1-\gamma_2} f(x) \prod_{1 \leq i \leq K-1} T_{1-\gamma_1} g(y_i)] - \mathbb{E}[T_{1-\gamma_2} f(x) \prod_{1 \leq i \leq K-1} \overline{T}_{1-\gamma_1} g(y_i)]\Big| \\
\leq&2 \cdot 3^K ((1-\gamma_1)^J + (\frac{J^2}{T})^{1/4}).
\end{aligned}$$

Fixing J and T to satisfy $2 \cdot 3^K((1-\gamma_1)^J + (\frac{J^2}{T})^{1/4}) \leq \frac{\theta}{8}$ as well as the previous constraint, we can conclude that

$$\left|\mathbb{E}[f(x) \prod_{1 \leq i \leq K-1} g(y_i)] - \mathbb{E}[T_{1-\gamma_2} f(x) \prod_{1 \leq i \leq K-1} T_{1-\gamma_1} g(y_i)]\right| \leq \frac{3\theta}{8}. \tag{10.11}$$

In particular, if I is independent, from (10.9) and (10.11)

$$\mathbb{E}[T_{1-\gamma_2} f(x) \prod_{1 \leq i \leq K-1} T_{1-\gamma_1} g(y_i)] \leq \frac{\theta}{2}. \tag{10.12}$$

STEP 4. **Invariance.** The marginal of y_i (resp. x) is $\omega^{\otimes R}$ (resp. $\omega^{\otimes L}$) on both $\overline{\mu}'^{\otimes L}$ and $\overline{\mu}^{\otimes L}$. Therefore, the Efron-Stein decomposition of f and g as well as the notion of (block) influence remain the same between $\overline{\mu}'$ and $\overline{\mu}'_\beta$. Since g is noised, there exists $\Gamma = O(\frac{1}{\gamma_1})$ such that

$$\sum_{1 \leq j \leq L} \overline{\mathsf{Inf}}_j[T_{1-\gamma_1} g] \leq \Gamma.$$

Fix τ to satisfy $Q \cdot 2^{K+1}\sqrt{\Gamma K^2 \tau} < \frac{\theta}{4}$. From (10.10) and (10.12),

$$\begin{aligned}
&\left|\mathbb{E}[T_{1-\gamma_2} f(x) \prod_{1 \leq i \leq K-1} T_{1-\gamma_1} g(y_i)] - \mathbb{E}[T_{1-\gamma_2} f(x)]\, \mathbb{E}[\prod_{1 \leq i \leq K-1} T_{1-\gamma_1} g(y_i)]\right| \\
\geq &\,\mathbb{E}[T_{1-\gamma_2} f(x)]\, \mathbb{E}[\prod_{1 \leq i \leq K-1} T_{1-\gamma_1} g(y_i)]| - \mathbb{E}[T_{1-\gamma_2} f(x) \prod_{1 \leq i \leq K-1} T_{1-\gamma_1} g(y_i)] \\
\geq &\frac{\theta}{2}.
\end{aligned}$$

Applying Theorem 10.4.7 ($Q \leftarrow 2$, $k_1 \leftarrow K-1$, $k_2 = 1$, $\Omega_1 = \overline{\Omega}$, $\Omega_2 \leftarrow \Omega$, $\nu \leftarrow \overline{\mu}'_\beta$, $L \leftarrow L$, $F_1 \leftarrow \overline{T_{1-\gamma_1} g}$, $F_2 \leftarrow T_{1-\gamma_2} f$, $\mathsf{Inf}_j[F_1] \leftarrow \overline{\mathsf{Inf}}_j[T_{1-\gamma_1} g]$),

$$\sum_{1 \leq j \leq L} \overline{\mathsf{Inf}}_j[T_{1-\gamma_1} g] \mathsf{Inf}_j[T_{1-\gamma_2} f] > \tau.$$

STEP 5. **Decoding Strategy.** We use the following standard strategy — v samples a set $S \subseteq [R]$ according to $\|g_S\|_2^2$, and chooses a random element from S. u also samples a set $S \subseteq [L]$ according to $\|f_S\|_2^2$, and chooses a random element from S. As shown in Section 10.7.2, for each $1 \leq j \leq L$, the probability that v chooses a label in $\pi^{-1}(j)$ is at least $\gamma_1 \overline{\mathsf{Inf}}_j[T_{1-\gamma_1} g]$, and the probability that u chooses j is at least $\gamma_2 \mathsf{Inf}[T_{1-\gamma_2} f]$.

The probability that $\pi_e(l(v)) = \pi(l(u))$ is at least

$$\gamma_1\gamma_2 \sum_{1 \le j \le L} \overline{\mathsf{Inf}}_j[T_{1-\gamma_1} g]\mathsf{Inf}_j[T_{1-\gamma_2} f] \ge \gamma_1\gamma_2\tau.$$

Suppose that I is indepenent. For at least $\frac{\epsilon^3}{2048}$ fraction of edges (of $E_{a,b}$) the above analysis works, and these edges are satisfied by the above randomized strategy with probability $\gamma_1\gamma_2\tau$. Setting $\eta := \frac{\epsilon^3}{2048} \cdot \gamma_1\gamma_2\tau$ completes the proof of soundness.

10.5.4 Distribution for Theorem 9.2.4

For Theorem 9.2.4, we again define the distribution of K points, one in a single cell and the other $K-1$ in a block of size d. Let $\Omega = \{0, 1\}$ and $\overline{\Omega} = \Omega^d$. Let ω be the $(1 - \frac{1}{K-1})$-biased distribution on Ω — $\omega(0) = \frac{1}{K-1}$ and $\omega(1) = 1 - \frac{1}{K-1}$. The K points $x \in \Omega$ and $y_1, \dots, y_{K-1} \in \overline{\Omega}$ are sampled by the following procedure.

- Sample $x \sim \omega$.
- If $x = 0$, sample $y_1, \dots, y_{K-1} \sim \omega^{\otimes d}$ independently.
- If $x = 1$, for each $1 \le j \le d$, sample $(y_1)_j, \dots, (y_{K-1})_j \sim \mu$, where μ is the uniform distribution on $K-1$ bit strings with exactly $(K-2)$ 1's.

$\Pr[(y_i)_j = 1] = \frac{1}{K-1} \cdot (1 - \frac{1}{K-1}) + (1 - \frac{1}{K-1})(\frac{K-2}{K-1}) = (1 - \frac{1}{K-1})$ for all $i \in [K-1]$ and $j \in [d]$, and $(y_i)_1, \dots, (y_i)_d$ are independent. Let $(\Omega \times \overline{\Omega}^{K-1}, \overline{\mu}')$ denote the K correlated spaces corresponding to the above distribution, and let $\overline{\mu}$ denote the marginal distribution of $(y_1, \dots, y_{K-1})$. Let $\overline{\Omega}_i$ $(1 \le i \le K-1)$ denote the copy of $\overline{\Omega}$ associated with y_i, and $\overline{\Omega}'_i$ be the product of the other $K-1$ spaces. With probability $\frac{1}{K-1}$ (when $x = 0$), y_i is completely independent of the others. Even when $x = 1$, y_i's marginal is $\omega^{\otimes d}$. By Lemma 10.2.1, we conclude that $\rho(\overline{\Omega}_i, \overline{\Omega}'_i; \overline{\mu}') \le \sqrt{\frac{K-2}{K-1}}$. Bounding $\rho(\Omega, \overline{\Omega}^{K-1}; \overline{\mu}')$ (as the correlation between two spaces Ω and $\overline{\Omega}^{K-1}$) can be done in the same way as the proof of Theorem 9.2.4 in this section: (1) define the distribution $\overline{\mu}'_\beta$ for the sake of analysis where each y_i is independently resampled with probability $1 - \beta$ after sampled according to $\overline{\mu}'$, (2) show that analyzing $\overline{\mu}'_\beta$ instead of $\overline{\mu}'$ incurs little extra error, and (3) use the standard technique to prove $\rho(\Omega, \overline{\Omega}^{K-1}; \overline{\mu}'_\beta) \le \sqrt{1 - (1-\beta)^{K-1}}$.

The fact that for each $1 \leq j \leq d$, at least one of $x, (y_1)_j, ..., (y_K)_j$ is 1 ensures completeness, and the bounded correlation ensures soundness. Furthermore, the fact that $y_1, ..., y_{K-1}$ become completely independent with probability $\frac{1}{K-1}$ (previously this was ϵ) implies $\zeta := \epsilon^{O_K(1)}$ and the same argument in Theorem 9.2.1 shows density in soundness.

10.6 Q-out-of-$(2Q+1)$-SAT

An instance of $(2Q+1)$-SAT is a tuple (V, Φ) consisting of the set of variables V and the set of clauses Φ. Each clause ϕ is described by $((v_1, z_1), \ldots, (v_{2Q+1}, z_{2Q+1}))$ where $v_q \in V$ and $z_q \in \{0, 1\}$. To be consistent with the notation we used for hypergraph coloring, we use the unconventional notation where 0 denotes True and 1 denotes False. Let $f : V \to \{0, 1\}$ be an assignment to variables. The number of literals of ϕ set to True by f is $|\{q : f(v_q) \oplus z_q = 0\}|$ where $\oplus$ denotes the sum over $\mathbb{Z}_2$.

10.6.1 Distribution

We first define the distribution of $2Q+1$ points, one in a single cell and the other $2Q$ in a block of size d. Let $\Omega = \{0, 1\}$ and $\overline{\Omega} = \Omega^d$. Let ω be the uniform distribution on Ω. $2Q+1$ points $x_0 \in \Omega$ and $x_{q,i} \in \overline{\Omega}$ for $1 \leq q \leq Q$ and $1 \leq i \leq k$ are sampled by the following procedure.

- Sample $q' \in \{0, \ldots, Q\}$ uniformly at random.
- If $q' = 0$,
  - Sample $x_0 \in \Omega$ uniformly independently.
  - For all $q \in [Q]$, sample $x_{q,1} \in \Omega^d$ independently and set $x_{q,2} = \mathbf{1}_d - x_{q,1}$, where $\mathbf{1}_d \in \Omega^d := (1, 1, \ldots, 1)$.
- If $q' > 0$,
  - For all $q \in [Q] \setminus \{q'\}$, sample $x_{q,1} \in \Omega^d$ independently and set $x_{q,2} = \mathbf{1}_d - x_{q,1}$.
  - Sample $x_0 \in \Omega$ independently. If $x_0 = 0$, sample $x_{q,1}, x_{q,2} \in \Omega^d$ independently. If $x_0 = 1$, sample $x_{q,1} \in \Omega^d$ independently and set $x_{q,2} = \mathbf{1}_d - x_{q,1}$.

Let $(\Omega \times \overline{\Omega}^{2Q}, \overline{\mu}')$ denote $2Q+1$ correlated spaces corresponding to the above distribution, and $\overline{\mu}$ denote the marginal distribution of $(x_{q,1}, x_{q,2})$, which is the same for all $q \in [Q]$. We bound $\rho(\Omega, \overline{\Omega}^{2Q}; \overline{\mu}')$.

Fix some $1 \leq q \leq Q$ and $1 \leq i \leq 2$. Let $\overline{\Omega}_{q,i}$ denote the copy of $\overline{\Omega}$ associated with $x_{q,i}$, and $\overline{\Omega}'_{q,i}$ be the product of the other $2Q$ copies. We have $\overline{\mu}' = \frac{1}{2(Q+1)}\alpha_q + (1 - \frac{1}{2(Q+1)})\beta_q$ where α_q denotes the distribution given $q' = q$ and $x_0 = 0$ (so that $x_{q,1}, x_{q,2}$ are sampled i.i.d.), and β_q denotes the distribution $q' \neq q$ or $x_0 = 1$. Since each entry of $x_{q,i}$ is sampled i.i.d. in α_q, $\rho(\overline{\Omega}_{q,i}, \overline{\Omega}'_{q,i}; \alpha_q) = 0$. In both α_q and β_q, the marginal of $x_{q,i}$ is $\omega^{\otimes d}$. By Lemma 10.2.1, we conclude that $\rho(\overline{\Omega}_{q,i}, \overline{\Omega}'_{q,i}; \overline{\mu}') \leq \sqrt{1 - \frac{1}{2(Q+1)}}$. Similarly, $\rho(\Omega, \overline{\Omega}^{2Q}; \overline{\mu}') \leq \sqrt{1 - \frac{1}{Q+1}}$. Therefore we have

$$\rho(\Omega, (\overline{\Omega}_{q,i})_{q,i}; \overline{\mu}') \leq \sqrt{1 - \frac{1}{2(Q+1)}}\,.$$

10.6.2 Reduction and Completeness

We now describe the reduction from $(Q+1)$-BIPARTITE HYPERGRAPH LABEL COVER. Given a $(Q+1)$-uniform hypergraph $H = (U \cup V, E)$ with Q projections from $[R]$ to $[L]$ for each hyperedge, the resulting instance for $(2Q+1)$-SAT is $(U' \cup V', \Phi)$ where $U' := (U \times \Omega^L)$ and $V' := (V \times \Omega^R)$. For $u \in U$ and $v \in V$, let $\mathsf{Cloud}(u) := \{u\} \times \Omega^L$ and $\mathsf{Cloud}(v) := \{v\} \times \Omega^R$. The clauses in Φ are described by the following procedure.

- Sample a random hyperedge $e = (u, v_1, \ldots, v_Q)$ with the associated projections $\pi_{e,v_1}, \ldots, \pi_{e,v_Q}$ from E.
- Sample $x_0 \in \Omega^L, (x_{q,i})_{1 \leq q \leq Q, 1 \leq i \leq 2} \in \Omega^R$ in the following way. For each $1 \leq j \leq L$, sample $(x_0)_j, ((x_{q,i})_{\pi^{-1}_{e,v_q}(j)})_{q,i}$ from $(\Omega \times \overline{\Omega}^{2Q}, \overline{\mu}')$.
- Sample $z_0, (z_{q,i})_{1 \leq q \leq Q, 1 \leq i \leq 2} \in \Omega$ i.i.d.
- Add a clause

  $$((u, x_0 \oplus z_0\, \mathbf{1}_L), z_0) \times ((v_q, x_{q,i} \oplus z_{q,i}\, \mathbf{1}_R), z_{q,i})_{1 \leq q \leq Q, 1 \leq i \leq 2}$$

  to Φ. We say this clause is *formed from* $e \in E$.

Given the reduction, completeness is easy to show.

Lemma 10.6.1. *If an instance of* $(Q+1)$-BIPARTITE HYPERGRAPH LABEL COVER *admits a labeling that strongly satisfies every hyperedge* $e \in E$*, there is an assignment* $f : U' \cup V' \to \Omega$ *that sets at least* Q *literals to* 0 *(which denotes True in our convention) in every clause of* Φ.

Proof. Let $l : U \cup V \to [R]$ be a labeling that strongly satisfies every hyperedge $e \in E$. For any $u \in U, x \in \Omega^L$, let $f(u,x) = x_{l(u)}$. For any $v \in V, x \in \Omega^R$, let $f(v,x) = x_{l(v)}$. For any clause

$$((u, x_0 \oplus z_0 \,\mathbf{1}_L), z_0) \times ((v_q, x_{q,i} \oplus z_{q,i} \,\mathbf{1}_R), z_{q,i})_{1 \le q \le Q, 1 \le i \le 2},$$

one of the following is true. Note that $f(u, x_0 \oplus z_0 \,\mathbf{1}_L) \oplus z_0 = (x_0)_{l(u)}$ and $f(v_q, x_{q,i} \oplus z_{q,i} \,\mathbf{1}_R) \oplus z_{q,i} = (x_{q,i})_{l(v_q)}$.

- Each $q \in [Q]$ satisfies $(x_{q,1})_{l(v_q)} \neq (x_{q,2})_{l(v_q)}$.
- For some $q \in [Q]$, all $q' \in [Q] \setminus \{q\}$ satisfy $(x_{q',1})_{l(v'_q)} \neq (x_{q',2})_{l(v'_q)}$, and if $(x_0)_{l(u)} = 1$, q also satisfies $(x_{q,1})_{l(v_q)} \neq (x_{q,2})_{l(v_q)}$.

In any case, $(2Q+1)$-tuple $((x_0)_{l(u)}) \times ((x_{q,i})_{l(v_q)})_{q,i}$ contains at least Q zeros, which means that any clause has at least Q literals set True. $\square$

10.6.3 Soundness

Lemma 10.6.2. *There exist* $\epsilon, \eta > 0$*, only depending on* Q*, such that if there is an assignment that satisfies more than* $(1-\epsilon)$ *fraction of hyperedges, the corresponding instance of* Q-HYPERGRAPH LABEL COVER *admits a labeling that weakly satisfies* η *fraction of hyperedges.*

The proof is almost identical to the one presented in Section 10.7.2. Let $g : U' \cup V' \to \Omega$ be any assignment. The fraction of clauses whose literals are all set to False is

$$\begin{aligned}
&\mathbb{E}_{u,v_1,\dots,v_Q}\mathbb{E}_{x_0,(x_{q,i})}\mathbb{E}_{z_0,(z_{q,i})}[(g(u, x_0 \oplus \mathbf{1}_L\, z_0) \oplus z_0) \prod_{1 \le q \le Q, 1 \le i \le 2} (g(v_q, x_{q,i} \oplus \mathbf{1}_R\, z_{q,i}) \oplus (z_0))] \\
=&\mathbb{E}_{u,v_1,\dots,v_Q}\mathbb{E}_{x_0,(x_{q,i})}[\mathbb{E}_{z_0}[(g(u, x_0 \oplus \mathbf{1}_L\, z_0) \oplus z_0)] \prod_{1 \le q \le Q, 1 \le i \le 2} \mathbb{E}_{z_{q,i}}[g(v_q, x_{q,i} \oplus \mathbf{1}_R\, z_{q,i}) \oplus z_{q,i}]] \\
=&\mathbb{E}_{u,v_1,\dots,v_Q}\mathbb{E}_{x_0,(x_{q,i})}[f(u, x_0) \prod_{1 \le q \le Q, 1 \le i \le 2} f(v, x_{q,i})]
\end{aligned}$$

where we define

$$f(u,x) := \mathbb{E}_{z\in\Omega}[f(u,x\oplus \mathbf{1}_L\, z)\oplus z)] \qquad u\in U$$
$$f(v,x) := \mathbb{E}_{z\in\Omega}[f(v,x\oplus \mathbf{1}_R\, z)\oplus z)] \qquad v\in V.$$

For $u \in U$, let $f_u \in \mathcal{L}_{[0,1]}(\Omega^L)$ be the restriction of f to $\{u\}\times\Omega^L$, and define $f_v \in \mathcal{L}_{[0,1]}(\Omega^R)$ similarly for $v\in V$. Note that $\mathbb{E}[f_u] = \mathbb{E}[f_v] = \frac{1}{2}$.

STEP 1. **Fixing a Good Hyperedge.** Since $\mathbb{E}[f_u] = \mathbb{E}[f_v] = \frac{1}{2}$ for all $u\in U$, and $v \in V$, we do not need to define heavy vertices. By the same argument as in Section 10.7.2, by adjusting the smoothness paramter T and the integer J, we can ensure that $\delta := \frac{1}{2}$ fraction of hyperedges are good for every vertex they contain, i.e., the hyperedge $e = (u, v_1, \ldots, v_Q)$ satisfies for each $q \in [Q]$,

$$\|f_{v_q}^{\mathsf{bad}}\|_2 \le (\frac{J^2}{T})^{1/4}$$

under π_{e,v_q} and J.

Throughout the rest of the section, fix such a hyperedge $e = (u, v_1, \ldots, v_Q)$ and the associated projections $\pi_{e,v_1}, \ldots, \pi_{e,v_Q}$. For simplicity, let $f_q := f_{v_q}$ and $\pi_q := \pi_{e,v_q}$ for $q\in[Q]$, and $f_{q+1} = f_u$. We now measure the fraction of clauses formed from e that are unsatisfied, which is

$$\mathbb{E}_{x_{q,i}}[f_u(x_0)\prod_{1\le q\le Q, 1\le i\le 2} f_q(x_{q,i})] \tag{10.13}$$

STEP 2. **Lower Bounding in Each Hypercube.** Fix any $q\in[Q]$. For each $1\le j\le L$, with probability $\frac{1}{2(Q+1)}$, $(x_{q,1})_{\pi_q^{-1}(j)}$ and $(x_{q,2})_{\pi_q^{-1}(j)}$ are sampled completely independently from $\overline{\Omega}$. By Theorem 10.3.8 (setting $\Omega \leftarrow \overline{\Omega}$, $k\leftarrow 2$, $\sigma \leftarrow \omega^{\otimes d}$, $\nu \leftarrow \overline{\mu}$, $\rho \leftarrow \sqrt{\frac{2Q+1}{2(Q+1)}}$, $F_1 = F_2 \leftarrow \overline{f_q}$, $\epsilon \leftarrow \frac{1}{2}$), there exists $\zeta = \zeta(Q) > 0$ such that for every $\gamma \in [0,1]$,

$$\mathbb{E}_{x_{q,1},x_{q,2}}\big[\, T_{1-\gamma} f_q(x_{q,1})\; T_{1-\gamma} f_q(x_{q,2})\,\big] \ge \zeta\,. \tag{10.14}$$

STEP 3. **Smoothing Functions.** Since $\rho(\Omega, (\overline{\Omega}_{q,i})_{q,i}; \overline{\mu}') \le \sqrt{1-\frac{1}{2(Q+1)}}$, we can apply Theorem 10.4.4 ($K \leftarrow 2Q+1$, $\Omega_1 = \cdots = \Omega_{K-1} \leftarrow \overline{\Omega}$, $\Omega_K \leftarrow \Omega$, $\nu \leftarrow \overline{\mu}'$, $\epsilon \leftarrow \frac{\zeta^Q}{8K}$,

$F_{2q-1} = F_{2q} \leftarrow \overline{f_q}$, $F_K \leftarrow f_u$) to have $\gamma := \gamma(Q, \zeta) \in (0, 1)$ such that

$$\left|\mathbb{E}_{x_{q,i}}[f_u(x_0) \prod_{1 \le q \le Q, 1 \le i \le 2} f_q(x_{q,i})] - \mathbb{E}_{x_{q,i}}[T_{1-\gamma} f_u(x_0) \prod_{1 \le q \le Q, 1 \le i \le 2} \overline{T}_{1-\gamma} f_q(x_{q,i})]\right| \le \frac{\zeta^Q}{8}. \tag{10.15}$$

By applying Theorem 10.4.5 ($K \leftarrow 2Q+1$, $L \leftarrow L$, $\Omega_1, \ldots, \Omega_K \leftarrow \Omega$, $d_1, \ldots, d_{K-1} \leftarrow d$, $d_K = 1$, $\nu \leftarrow \overline{\mu}'$, $F_{2q-1} = F_{2q} \leftarrow f_q$, $F_K \leftarrow f_u$, $\pi_{2q-1} = \pi_{2q} \leftarrow \pi_q$, $\pi_K \leftarrow$ the identity, $\xi \leftarrow (\frac{J^2}{T})^{1/4}$), we have

$$\left|\mathbb{E}_{x_{q,i}}[T_{1-\gamma} f_u(x_0) \prod_{1 \le q \le Q, 1 \le i \le 2} \overline{T}_{1-\gamma} f_q(x_{q,i})] - \mathbb{E}_{x_{q,i}}[T_{1-\gamma} f_u(x_0) \prod_{1 \le q \le Q, 1 \le i \le 2} T_{1-\gamma} f_q(x_{q,i})]\right|$$
$$\le 2 \cdot 3^{2Q+1}((1-\gamma)^J + (\frac{J^2}{T})^{1/4}) \,. \tag{10.16}$$

Fixing J and T to satisfy $2 \cdot 3^{2Q+1}((1-\gamma)^J + (\frac{J^2}{T})^{1/4}) \le \frac{\zeta^Q}{8}$ as well as the previous constraint, we can conclude from (10.15) and (10.16) that

$$\left|\mathbb{E}_{x_{q,i}}[f_u(x_0) \prod_{1 \le q \le Q, 1 \le i \le 2} f_q(x_{q,i})] - \mathbb{E}_{x_{q,i}}[T_{1-\gamma} f_u(x_0) \prod_{1 \le q \le Q, 1 \le i \le 2} T_{1-\gamma} f_q(x_{q,i})]\right| \le \frac{\zeta^Q}{4}.$$

In particular, if among the clauses formed from e, less than $\frac{\zeta^Q}{8}$ fraction of them are unsatisfied, from (10.13),

$$\mathbb{E}_{x_{q,i}}\left[T_{1-\gamma} f_u(x_0) \prod_{1 \le q \le Q, 1 \le i \le 2} T_{1-\gamma} f_q(x_{q,i})\right] \le \frac{3\zeta^Q}{8}. \tag{10.17}$$

STEP 4. **Invariance.** Since our functions are noised, there exists $\Gamma = O(\frac{1}{\gamma})$ such that

$$\sum_{1 \le j \le L} \overline{\mathsf{Inf}}_j[T_{1-\gamma} f_q] \le \Gamma.$$

Fix τ to satisfy $8Q \cdot \sqrt{\Gamma(2Q+1)^2\tau} < \frac{\zeta^Q}{8}$. We have

$$\left|\mathbb{E}_{x_{q,i}}[T_{1-\gamma} f_u(x_0) \prod_{1 \le q \le Q, 1 \le i \le 2} T_{1-\gamma} f_q(x_{q,i})] - \mathbb{E}[T_{1-\gamma} f_u] \prod_{1 \le q \le Q} \mathbb{E}_{x_{q,i}}[\prod_{1 \le i \le 2} T_{1-\gamma} f_q(x_{q,i})]\right|$$
$$\ge \mathbb{E}[T_{1-\gamma} f_u] \cdot \prod_{1 \le q \le Q} \mathbb{E}_{x_{q,i}}[\prod_{1 \le i \le 2} T_{1-\gamma} f_q(x_{q,i})] - \mathbb{E}_{x_{q,i}}[T_{1-\gamma} f_u(x_0) \prod_{1 \le q \le Q, 1 \le i \le 2} T_{1-\gamma} f_q(x_{q,i})]$$
$$\ge \frac{1}{2}\zeta^Q - \frac{3\zeta^Q}{8} = \frac{\zeta^Q}{8} \quad \text{(using (10.14) and (10.17))} \,.$$

Now, applying Theorem 10.4.7 ($Q \leftarrow Q+1$, $k_1 = \cdots = k_Q \leftarrow k$, $k_{Q+1} \leftarrow 1$, $\Omega_1 = \cdots = \Omega_Q = \overline{\Omega}$, $\Omega_{Q+1} \leftarrow \Omega$, $\nu \leftarrow \overline{\mu}'$, $L \leftarrow L$, $F_q \leftarrow \overline{T_{1-\gamma} f_q}$ for $q \in [Q]$, $F_{Q+1} \leftarrow T_{1-\gamma} f_u$, $\mathsf{Inf}_j[F_q] \leftarrow \overline{\mathsf{Inf}}_j[T_{1-\gamma} f_q]$ for $q \in [Q]$), there exists $q \in \{1, \ldots, Q\}$ such that

$$\sum_{1 \le j \le L} \overline{\mathsf{Inf}}_j[T_{1-\gamma} f_q](\overline{\mathsf{Inf}}_j[T_{1-\gamma} f_{q+1}] + \cdots + \overline{\mathsf{Inf}}_j[T_{1-\gamma} f_Q] + \mathsf{Inf}_j[f_u]) > \tau.$$

STEP 5. **Decoding Strategy.** We use the standard strategy — each v_q samples a set $S \subseteq [R]$ according to $\|(f_q)_S\|_2^2$, and chooses a random element from S. u also samples a set $S \subseteq [L]$ according to $\|(f_u)_S\|_2^2$, and chooses a random element from S. As shown in Section 10.7.2, for each $1 \le j \le L$, the probability that v chooses a label in $\pi^{-1}(j)$ is at least $\gamma \overline{\mathsf{Inf}}_j[T_{1-\gamma} f_q]$, and the probability that u chooses j is at least $\gamma \mathsf{Inf}_j[T_{1-\gamma} f_u]$.

Fix q to be the one obtained from Theorem 10.4.7. The probability that $\pi_q(l(v_q)) = \pi_{q'}(l(v_{q'}))$ for some $q < q' \le Q$ or $\pi_q(l(v_q)) = l(u)$ is at least

$$\begin{aligned}
&\gamma^2 \sum_{1 \le j \le L} \overline{\mathsf{Inf}}_j[T_{1-\gamma} f_q] \max[\max_{q < q' \le Q} \overline{\mathsf{Inf}}_j[T_{1-\gamma} f_{q'}], \mathsf{Inf}_j[T_{1-\gamma} f_u]] \\
\ge &\frac{\gamma^2}{Q+1} \sum_{1 \le j \le L} \overline{\mathsf{Inf}}_j[T_{1-\gamma} f_q](\overline{\mathsf{Inf}}_j[T_{1-\gamma} f_{q+1}] + \cdots + \overline{\mathsf{Inf}}_j[T_{1-\gamma} f_Q] + \mathsf{Inf}_j[T_{1-\gamma} f_u]) \\
\ge &\frac{\gamma^2 \tau}{Q+1}.
\end{aligned}$$

If the total fraction of unsatisfied clauses is at most $\epsilon := \frac{1}{4} \cdot \frac{\zeta^Q}{8}$, since at least $\frac{1}{2}$ fraction of hyperedges are good, at least $\frac{1}{4}$ fraction of hyperedges are weakly satisfied by the above randomized strategy with probability $\frac{\gamma^2 \tau}{Q+1}$. Setting $\eta := \frac{1}{4} \cdot \frac{\gamma^2 \tau}{Q+1}$ completes the proof of soundness.

10.7 Hardness of MAX 2-COLORING under Low Discrepancy

In this section we consider the hardness of MAX 2-COLORING when promised discrepancy as low as one.

10.7.1 Reduction from MAX CUT

Let $K = 2t + 1$. Let $G = (V, E)$ be an instance of MAX CUT, where each edge has weight 1. Let $n = |V|$ and $m = |E|$. We produce a hypergraph $H = (V', E')$ where $V' = V \times [K]$. For each $u \in V$, let $\mathsf{cloud}(u) := \{u\} \times [K]$. For each edge $(u, v) \in E$, we add $N := 2\binom{K}{t}\binom{K}{t+1}$ hyperedges

$$\{U \cup V : U \subseteq \mathsf{cloud}(u), V \subseteq \mathsf{cloud}(v), |U| + |V| = K, ||U| - |V|| = 1\},$$

each with weight $\frac{1}{N}$. Call these hyperedges *created by* (u, v). The sum of weights is m for both G and H.

Completeness. Given a coloring $C : V \mapsto \{B, W\}$ that cuts at least $(1 - \alpha)m$ edges of G, we color H so that for every $v \in V$, each vertex in $\mathsf{cloud}(v)$ is given the same color as v. If $(u, v) \in E$ is cut, all hyperedges created by (u, v) will have discrepancy 1. Therefore, the total weight of hyperedges with discrepancy 1 is at least $(1 - \alpha)m$.

Soundness. Given a coloring $C' : V' \mapsto \{B, W\}$ such that the total weight of non-monochromatic hyperedges is $(1 - \beta)m$, $v \in V$ is given the color that appears the most in its cloud (K is odd, so it is well-defined). Consider $(u, v) \in E$. If no hyperedge created by (u, v) is monochromatic, it means that u and v should be given different colors by the above majority algorithm (if they are given the same color, say white, then there are at least $t + 1$ white vertices in both clouds, so we have at least one monochromatic hyperedge).

This means that for each $(u, v) \in E$ that is uncut by the above algorithm (lost weight 1 for MAX CUT objective), at least one hyperedge created by (u, v) is monochromatic, and we lost weight at least $\frac{1}{N}$ there for our problem. This means that the total weight of cut edges for MAX CUT is at least $(1 - \beta N)m$.

The Result. The following theorem shows hardness of MAX CUT.

Theorem 10.7.1 ([KKMO07]). *Let $G = (V, E)$ be a graph with $m = |E|$. For sufficiently small $\epsilon > 0$, it is UG-hard to distinguish the following cases.*

- *There is a 2-coloring that cuts at least $(1 - \epsilon)|E|$ edges.*
- *Every 2-coloring cuts at most $(1 - (2/\pi)\sqrt{\epsilon})|E|$ edges.*

Our reduction proves the following.

Theorem 10.7.2. *Given a hypergraph $H = (V, E)$, it is UG-hard to distinguish the following cases.*

- *There is a 2-coloring where at least $(1-\epsilon)$ fraction of hyperedges have discrepancy 1.*
- *Every 2-coloring cuts (in a standard sense) at most $(1-(2/\pi)\frac{\sqrt{\epsilon}}{N})$ fraction of hyperedges.*

Note that $N = 2\binom{K}{t}\binom{K}{t+1} \leq (2/\pi)2^K \cdot 2^K \leq (2/\pi)2^{2K}$. If we take $\epsilon = 2^{-6K}$ for large enough K, we cannot distinguish the following two cases:

- There is a 2-coloring where at least $(1-2^{-6K})$ fraction of hyperedges have discrepancy 1.
- Every 2-coloring cuts (in a standard sense) at most $(1-2^{-5K})$ fraction of hyperedges.

This proves Theorem 9.2.3.

10.7.2 NP-Hardness

In this subsection, we show that given a hypergraph which admits a 2-coloring with discrepancy at most 2, it is NP-hard to find a 2-coloring that has less than $K^{-O(K)}$ fraction of monochromatic hyperedges. Note that while the inapproximability factor is worse than the previous subsection, we get NP-hardness and it holds when the input hypergraph is promised to have *all* hyperedges have discrepancy at most 2.

Distributions. We first define the distribution $\overline{\mu}'$ for each *block*. $2Q$ points $x_{q,i} \in \{1,2\}^d$ for $1 \leq q \leq Q$ and $1 \leq i \leq 2$ are sampled by the following procedure.

- Sample $q' \in [Q]$ uniformly at random.
- Sample $x_{q',1}, x_{q',2} \in \{1,2\}^d$ i.i.d.
- For $q \neq q'$, $1 \leq j \leq d$, sample a permutation $((x_{q,1})_j, (x_{q,2})_j) \in \{(1,2),(2,1)\}$ uniformly at random.

Reduction and Completeness. We now describe the reduction from Q-HYPERGRAPH LABEL COVER. Given a Q-uniform hypergraph $H = (V, E)$ with Q projections from $[R]$ to $[L]$ for each hyperedge (let $d = R/L$), the resulting instance of $2Q$-hypergraph coloring is $H' = (V', E')$ where $V' = V \times \{1, 2\}^R$. Let $\mathsf{cloud}(v) := \{v\} \times \{1, 2\}^R$. The set E' consists of hyperedges generated by the following procedure.

- Sample a random hyperedge $e = (v_1, \ldots, v_Q) \in E$ with associated projections $\pi_{e,v_1}, \ldots, \pi_{e,v_Q}$ from E.
- Sample $(x_{q,i})_{1 \le q \le Q, 1 \le i \le 2} \in \{1, 2\}^R$ in the following way. For each $1 \le j \le L$, independently sample $((x_{q,i})_{\pi^{-1}_{e,v_q}(j)})_{q,i}$ from $((\{1, 2\}^d)2Q, \overline{\mu}')$.
- Add a hyperedge between $2Q$ vertices $\{(v_q, x_{q,i})\}_{q,i}$ to E'. We say this hyperedge is *formed from* $e \in E$.

Given the reduction, completeness is easy to show.

Lemma 10.7.1. *If an instance of* Q-HYPERGRAPH LABEL COVER *admits a labeling that strongly satisfies every hyperedge* $e \in E$*, there is a coloring* $c : V' \to \{1, 2\}$ *of the vertices of* H' *such that every hyperedge* $e' \in E'$ *has at least* $(Q - 1)$ *vertices of each color.*

Proof. Let $l : V \to [R]$ be a labeling that strongly satisfies every hyperedge $e \in E$. For any $v \in V, x \in \{1, 2\}^R$, let $c(v, x) = x_{l(v)}$. For any hyperedge $e' = \{(v_q, x_{q,i})\}_{q,i} \in E'$, $c(v_q, x_{q,i}) = (x_{q,i})_{l(v_q)}$, and all but one q satisfies $\{(x_{q,1})_{l(v_q)}, (x_{q,2})_{l(v_q)}\} = \{1, 2\}$. Therefore, the above strategy ensures that every hyperedge of E' contains at least $(Q - 1)$ vertices of each color. □

Soundness. The following lemma establishes the soundness of our reduction.

Lemma 10.7.2. *There exists* $\eta := \eta(Q)$ *such that if* $I \subseteq V'$ *of measure* $\frac{1}{2}$ *induces less than* $Q^{-O(Q)}$ *fraction of hyperedges in* H'*, the corresponding instance of* Q-HYPERGRAPH LABEL COVER *admits a labeling that weakly satisfies a fraction* η *of hyperedges.*

Proof. Consider a vertex v and hyperedge $e \in E$ that contains v with a permutation $\pi = \pi_{e,v}$. Let $f : \{1, 2\}^R \mapsto [0, 1]$ be a *noised* indicator function of $I \cap \mathsf{cloud}(v)$ with $\mathbb{E}_{x \in \{1,2\}^R}[f(x)] \ge \frac{1}{2} - \epsilon$ for small $\epsilon > 0$ that will be determined later. We define the inner product

$$\langle f, g \rangle = \mathbb{E}_{x \in \{1,2\}^R}[f(x)g(x)].$$

f admits the Fourier expansion

$$\sum_{S \subseteq [R]} \hat{f}(S)\chi_S$$

where

$$\chi_S(x_1, \ldots, x_K) = \prod_{i \in S} (-1)^{x_i}, \qquad \hat{f}(S) = \langle f, \chi_S \rangle.$$

In particular, $\hat{f}(\emptyset) = \mathbb{E}[f(x)]$, and

$$\sum_S \hat{f}(S)^2 = \mathbb{E}[f(x)^2] \le \mathbb{E}[f(x)] \tag{10.18}$$

A subset $S \subseteq [R]$ is said to be *shattered* by π if $|S| = |\pi(S)|$. For a positive integer J, we decompose f as the following:

$$f^{\mathsf{good}} = \sum_{S\text{: shattered}} \hat{f}(S)\chi_S$$
$$f^{\mathsf{bad}} = f - f^{\mathsf{good}}.$$

By adding a suitable noise and using smoothness of LABEL COVER, for any $\delta > 0$, we can assume that $||f^{\mathsf{bad}}||^2 \le \delta$. See [GL15d] for the details.

Each time a $2Q$-hyperedge is sampled is formed from e, two points are sampled from each cloud. Let x, y be the points in $\mathsf{cloud}(v)$. Recall that they are sampled such that for each $1 \le j \le L$,

- With probability $\frac{1}{Q}$, for each $i \in \pi^{-1}(j)$, x_i and y_i are independently sampled from $\{1, 2\}$.
- With probability $\frac{Q-1}{Q}$, for each $i \in \pi^{-1}(j)$, (x_i, y_i) are sampled from $\{(1, 2), (2, 1)\}$.

We can deduce the following simple properties.

1. $\mathbb{E}_{x,y}[\chi_{\{i\}}(x)\chi_{\{i\}}(y)] = -\frac{Q-1}{Q}$. Let $\rho := -\frac{Q-1}{Q}$.
2. $\mathbb{E}_{x,y}[\chi_{\{i\}}(x)\chi_{\{j\}}(y)] = 0$ if $i \ne j$.
3. $\mathbb{E}_{x,y}[\chi_S(x)\chi_T(y)] = 0$ unless $\pi(S) = \pi(T) = \pi(S \cap T)$.

We are interested in lower bounding

$$\mathbb{E}_{x,y}[f(x)f(y)] \geq \mathbb{E}[f^{\mathsf{good}}(x)f^{\mathsf{good}}(y)] - 3\|f^{\mathsf{bad}}(x)\|^2\|f\|^2 \geq \mathbb{E}[f^{\mathsf{good}}(x)f^{\mathsf{good}}(y)] - 3\delta.$$

By the property 3.,

$$\begin{aligned}
\mathbb{E}[f^{\mathsf{good}}(x)f^{\mathsf{good}}(y)] &= \sum_{S:\text{ shattered}} \hat{f}(S)^2\rho^{|S|} \\
&= \mathbb{E}[f]^2 + \sum_{S:\text{ shattered}} \hat{f}(S)^2\rho^{|S|} \\
&\geq \mathbb{E}[f]^2 + \rho(\sum_{|S|>1} \hat{f}(S)^2) \qquad \text{since } \rho \text{ is negative} \\
&\geq \mathbb{E}[f]^2 + \rho(\mathbb{E}[f] - \mathbb{E}[f]^2) \quad \text{ by (10.18)} \\
&\geq \mathbb{E}[f]^2(1+\rho) - \epsilon \quad \text{ since } \mathbb{E}[f] \geq \frac{1}{2} - \epsilon \Rightarrow \mathbb{E}[f] - \mathbb{E}[f]^2 \leq \mathbb{E}[f]^2 + \epsilon \\
&\geq \frac{\mathbb{E}[f]^2}{Q} - \epsilon.
\end{aligned}$$

By taking ϵ and δ small enough, we can ensure that

$$\mathbb{E}[f(x)f(y)] \geq \zeta := \frac{1}{5Q}. \tag{10.19}$$

The soundness analysis of Guruswami and Lee [GL15d] ensures ((10.19) replaces their Step 2) that there exists $\eta := \eta(Q)$ such that if the fraction of hyperedges induced by I is less than $Q^{-O(Q)}$, the LABEL COVER instance admits a solution that satisfies η fraction of constraints. We omit the details. □

Corollary to MAX 2-COLORING under discrepancy $O(\log K)$. The above NP-hardness, combined with the reduction techinque from MAX CUT in Section 10.7.1, shows that given a K-uniform hypergraph, it is NP-hard to distinguish whether it has discrepancy at most $O(\log K)$ or any 2-coloring leaves at least $2^{-O(K)}$ fraction of hyperedges monochromatic. Even though the direction reduction from MAX CUT results in a similar inapproximability factor with discrepancy even 1, this result does not rely on the UGC and hold even *all* edges (compared to *almost* in Section 10.7.1) have discrepancy $O(\log K)$.

Let $r = \Theta(\frac{K}{\log K})$ so that $s = \frac{K}{r} = \Theta(\log K)$ is an integer. Given a r-uniform hypergraph, it is NP-hard to distinguish whether it has discrepancy at most 2 or any 2-coloring leaves at least $r^{-O(r)}$ fraction of hyperedges monochromatic. Given a r-uniform hypergraph, the reduction replaces each vertex v with $\mathsf{cloud}(v)$ that contains $(2s-1)$ new vertices. Each hyperedge $(v_1, \ldots, v_r)$ is replaced by $d := (\binom{2s-1}{s})^r \leq (2^s)^r = 2^K$ hyperedges

$$\{\cup_{i=1}^r V_i : V_i \subset \mathsf{cloud}(v_i), |V_i| = s\}.$$

If the given r-uniform hypergraph has discrepancy at most 2, the resulting K-uniform hypergraph has discrepancy at most $2s = O(\log K)$.

If the resulting K-uniform hypergraph admits a coloring that leaves α fraction of hyperedges monochromatic, giving v the color that appears more in $\mathsf{cloud}(v)$ is guaranteed to leaves at most $d\alpha$ fraction of hyperedges monochromatic. Therefore, if any 2-coloring of the input r-uniform hypergraph leaves at least $r^{-O(r)}$ fraction of hyperedges monochromatic, any 2-coloring of the resulting K-uniform hypergraph leaves at least $\frac{r^{-O(r)}}{d} = 2^{-O(K)}$ fraction of hyperedges.

10.7.3 Hardness under Almost Colorability

Let K be such that $\ell := \sqrt{K}$ be an integer and let $\chi := K - \ell$. We prove the following hardness result for any $\epsilon > 0$ assuming the Unique Games Conjecture: given a K-uniform hypergraph such that there is a χ-coloring that have at least $(1-\epsilon)$ fraction of hyperedges rainbow, it is NP-hard to find a 2-coloring that leaves at most $(\frac{1}{2})^{K-1}$ fraction of hyperedges monochromatic.

The main technique for this result is to show the existence of a *balanced pairwise independence distribution* with the desired support. Let μ be a distribution on $[\chi]^K$. μ is called balanced pairwise independent if for any $i \neq j \in [K]$ and $a, b \in [\chi]$,

$$\Pr_{(x_1,\ldots,x_K)\sim\mu}[x_i = a, x_j = b] = \frac{1}{\chi^2}.$$

For example, the uniform distribution on $[\chi]^K$ is a balanced pairwise distribution. We now consider the following distribution μ to sample $(x_1, \ldots, x_K) \in [\chi]^K$.

- Sample $S \subseteq [K]$ with $|S| = \chi$ uniformly at random. Let $S = \{s_1 < \cdots < s_\chi\}$.
- Sample a permutation $\pi : [\chi] \mapsto [\chi]$.
- Sample $y \in [\chi]$.
- For each $i \in [K]$, if $i = s_j$ for some $j \in [\chi]$, output $x_i = \pi(\chi)$. Otherwise, output $x_i = y$.

Note that for any supported by $(x_1, \ldots, x_K)$, we have $\{x_1, \ldots, x_K\} = [\chi]$. Therefore, μ is supported on *rainbow strings*. We now verify pairwise independence. Fix $i \neq j \in [K]$ and $a, b \in [\chi]$.

- If $a = b$, by conditioning on wheter i, j are in S or not,

$$\begin{aligned}
\Pr_{\mu}[x_i = a, x_j = b] =& \Pr[x_i = a, x_j = b | i, j \in S] \Pr[i, j \in S] + \\
& \Pr[x_i = a, x_j = b | i \in S, j \notin S] \Pr[i \in S, j \notin S] + \\
& \Pr[x_i = a, x_j = b | i \notin, j \in S] \Pr[i \notin, j \in S] + \\
& \Pr[x_i = a, x_j = b | i, j \notin S] \Pr[i, j \notin S] \\
=& 0 \cdot (\frac{\chi(\chi-1)}{K(K-1)}) + 2 \cdot (\frac{1}{\chi^2}) \cdot (\frac{l\chi}{K(K-1)}) + (\frac{1}{\chi}) \cdot (\frac{\ell(\ell-1)}{K(K-1)}) \\
=& \frac{2\ell\chi + \chi(\ell^2 - \ell)}{\chi^2 K(K-1)} = \frac{\chi K + \chi\sqrt{K}}{\chi^2 K(K-1)} = \frac{\sqrt{K}(\sqrt{K}+1)}{\chi K(\sqrt{K}+1)(\sqrt{K}-1)} \\
=& \frac{1}{\chi(K - \sqrt{K})} = \frac{1}{\chi^2}.
\end{aligned}$$

- If $a \neq b$, by the same conditioning,

$$\begin{aligned}
& \Pr_{\mu}[x_i = a, x_j = b] \\
=& (\frac{1}{\chi(\chi-1)}) \cdot (\frac{\chi(\chi-1)}{K(K-1)}) + 2 \cdot (\frac{1}{\chi^2}) \cdot (\frac{\ell\chi}{K(K-1)}) + 0 \cdot (\frac{\ell(\ell-1)}{K(K-1)}) \\
=& \frac{\chi^2 + 2l\chi}{\chi^2 K(K-1)} = \frac{\chi + 2\ell}{\chi K(K-1)} = \frac{K + \sqrt{K}}{\chi K(K-1)} = \frac{1}{\chi^2}.
\end{aligned}$$

Given such a balanced pairwise independent distribution supported on rainbow strings, a standard procedure following the work of Austrin and Mossel [AM09] shows that it is UG-hard to outperform the random 2-coloring. We omit the details.

Chapter 11

Algorithms for Coloring

11.1 Algorithmic Techniques

Our algorithms for MAX 2-COLORING are straightforward applications of semidefinite programming, namely, we use natural vector relaxations of the promised properties, and round using a random hyperplane. The analysis however, is highly non-trivial and boils down to approximating a multivariate Gaussian integral. In particular, we show a (to our knowledge, new) upper bound on the Gaussian measure of simplicial cones in terms of simple properties of these cones. We should note that this upper bound is sensible only for simplicial cones that are well behaved with respect to the these properties. (The cones we are interested in are those given by the intersection of hyperplanes whose normal vectors constitute a solution to our vector relaxations). We believe our analysis to be of independent interest as similar approaches may work for other K-CSPs.

Gaussian Measure of Simplicial Cones. As can be seen via an observation of Kneser [Kne36], the Gaussian measure of a simplicial cone is equal to the fraction of spherical volume taken up by a spherical simplex (a spherical simplex is the intersection of a simplicial cone with a ball centered at the apex of the cone). This however, is a very old problem in spherical geometry, and while some things are known, like a nice differential formula due to Schlafli (see [Sch58]), closed forms upto four dimensions (see [MY05]), and a complicated power series expansion due to Aomoto [Aom77], it is likely hopeless to achieve a closed form solution or even an asymptotic formula for the volume of general spherical simplices.

Zwick [Zwi98a] considered the performance of hyperplane rounding in various 3-CSP formulations, and this involved analyzing the volume of a 4-dimensional spherical simplex. Due to the complexity of this volume function, the analysis was tedious, and non-analytic for many of the formulations. His techniques were based on the Schlafli differential formula, which relates the volume differential of a spherical simplex to the volume functions of its codimension-2 faces and dihedral angles. However, to our knowledge not much is known about the general volume function in even 6 dimensions. This suggests that Zwick's techniques are unlikely to be scalable to higher dimensions.

On the positive side, an asymptotic expression is known in the case of symmetric spherical simplices, due to H. E. Daniels [Rog64] who gave the analysis for regular cones of angle $\cos^{-1}(1/2)$. His techniques were extended by Rogers [Rog61] and Boeroeczky and Henk [BJH99] to the whole class of regular cones.

We combine the complex analysis techniques employed by Daniels with a lower bound on quadratic forms in the positive orthant, to give an upper bound on the Gaussian measure of a much larger class of simplicial cones.

Column Subset Selection. Informally, the cones for which our upper bound is relevant are those that are high dimensional in a strong sense, i.e. the normal vectors whose corresponding hyperplanes form the cone, must be such that no vector is too close to the linear span of any subset of the remaining vectors.

When the normal vectors are solutions to our rainbow colorability SDP relaxation, this need not be true. However, this can be remedied. We consider the column matrix of these normal vectors, and using spectral techniques, we show that there is a reasonably large subset of columns (vectors) that are well behaved with respect to condition number. We are then able to apply our Gaussian Measure bound to the cone given by this subset, admittedly in a slightly lower dimensional space.

11.2 Approximate MAX 2-COLORING

In this section we show how the properties of $(K+\ell)$-strong colorability and $(K-\ell)$-rainbow colorability in K-uniform hypergraphs allow one to 2-color the hypergraph, such that the respective fractions of monochromatic edges are small. For $\ell = o(\sqrt{K})$, these guarantees handsomely beat the naive random algorithm (color every vertex blue or red uniformly and independently at random), wherein the expected fraction of monochromatic edges is $1/2^{K-1}$.

Our algorithms are straightforward applications of semidefinite programming, namely, we use natural vector relaxations of the above properties, and round using a random hyperplane. The analysis however, is quite involved.

11.2.1 Semidefinite Relaxations

Our SDP relaxations for low-discrepancy, rainbow-colorability, and strong-colorability are the following. Given that $\langle v_i, v_j\rangle = \frac{-1}{\chi-1}$ when unit vectors $v_1, \ldots, v_\chi$ form a χ-regular simplex centered at the origin, it is easy to show that they are valid relaxations.

Discrepancy ℓ.

$$\begin{aligned} &\left|\left|\sum_{i\in e} u_i\right|\right|_2 \le \ell && \forall\, e \in E \\ &||u_i||_2 = 1 && \forall\, i \in [n] \\ &u_i \in \mathbb{R}^n && \forall\, i \in [n] \end{aligned} \tag{11.1}$$

Feasibility. For K, ℓ such that $(K-\ell) \mod 2 \equiv 0$, consider any K-uniform hypergraph $H = (V = [n], E)$, and any 2-coloring of H of discrepancy ℓ. Pick any unit vector $w \in \mathbb{R}^n$. For each vertex of the first color in the coloring, assign the vector w, and for each vertex of the second color assign the vector $-w$. This is a feasible assignment, and hence Relaxation 11.1 is a feasible relaxation for any hypergraph of discrepancy ℓ.

$(K-\ell)$-Rainbow Colorability.

$$\begin{aligned} &\left|\left|\sum_{i\in e} u_i\right|\right|_2 \le \ell && \forall\, e \in E \\ &\langle u_i, u_j\rangle \ge \frac{-1}{K-\ell-1} && \forall\, e \in E,\ \forall\, i < j \in e \\ &||u_i||_2 = 1 && \forall\, i \in [n] \\ &u_i \in \mathbb{R}^n && \forall\, i \in [n] \end{aligned} \tag{11.2}$$

Feasibility. Consider any K-uniform hypergraph $H = (V = [n], E \subseteq \binom{V}{K})$, and any $(K-\ell)$-rainbow coloring of H. As testified by the vertices of the $(K-\ell)$-simplex, we can always choose unit vectors $w_1 \dots w_{K-\ell} \in \mathbb{R}^n$ satisfying,

$$\forall\, i < j \in [K-\ell],\ \langle w_i, w_j \rangle = \frac{-1}{K-\ell-1},$$

It is not hard to verify that consequently,

$$\forall\, a_1, \dots, a_{K-\ell} \in [l],\ \sum_{i \in [K-\ell]} a_i = K, \quad \text{we have,} \quad \left|\left| \sum_{i \in e} a_i w_i \right|\right|_2 \leq \ell$$

For each vertex of the color i, assign the vector w_i. This is a feasible assignment, and hence Relaxation 11.2 is a feasible relaxation for any hypergraph of rainbow colorability $K-\ell$.

$(K+\ell)$**-Strong Colorability.**

$$\begin{aligned} &\langle u_i, u_j \rangle = \frac{-1}{K+\ell-1} && \forall\, e \in E,\ \forall\, i < j \in e \\ &||u_i||_2 = 1 && \forall\, i \in [n] \\ &u_i \in \mathbb{R}^n && \forall\, i \in [n] \end{aligned} \tag{11.3}$$

Feasibility. Consider any K-uniform hypergraph $H = (V = [n], E \subseteq \binom{V}{K})$, and any $(K+\ell)$-strong coloring of H. As testified by the vertices of the $(K+\ell)$-simplex, we can always choose unit vectors $w_1 \dots w_{K+\ell} \in \mathbb{R}^n$ satisfying,

$$\forall\, i < j \in [K-\ell],\ \langle w_i, w_j \rangle = -\frac{1}{K+\ell-1},$$

It is not hard to verify that consequently,

$$\forall\, J \subset [K+\ell],\ |J| = K,\ \left|\left| \sum_{i \in J} w_i \right|\right|_2 = \ell$$

For each vertex of the color i, assign the vector w_i. This is a feasible assignment, and hence the Relaxation 11.3 is a feasible relaxation for any hypergraph of strong colorability $K+\ell$.

Our rounding scheme is the same for all the above relaxations.

Rounding Scheme. Pick a standard n-dimensional Gaussian random vector r. For any $i \in [n]$, if $\langle v_i, r \rangle \geq 0$, then vertex i is colored blue, and otherwise it is colored red.

11.2.2 Setup of Analysis

We now setup the framework for analyzing all the above relaxations.

Consider a standard n-dimensional Gaussian random vector r, i.e. each coordinate is independently picked from the standard normal distribution $\mathcal{N}(0,1)$. The following are well known facts (the latter being due to Renyi),

Lemma 11.2.1. *$r/\left\|r\right\|_2$ is uniformly distributed over the unit sphere in $\mathbb{R}^n$.*

Note. Lemma 11.2.1 establishes that our rounding scheme is equivalent to random hyperplane rounding.

Lemma 11.2.2. *Consider any $j < n$. The projections of r onto the pairwise orthogonal unit vectors $e_1, \ldots, e_j$ are independent and have distribution $\mathcal{N}(0,1)$.*

Next, consider any K-uniform hypergraph $H = (V = [n], E \subseteq \binom{V}{K})$ that is feasible for any of the aforementioned formulations. Our goal now, is to analyze the expected number of monochromatic edges. To obtain this expected fraction with high probability, we need only repeat the rounding scheme polynomially many times, and the high probability of a successful round follows by Markov's inequality. Thus we are only left with bounding the probability that a particular edge is monochromatic.

To this end, consider any edge $e \in E$ and let the vectors corresponding to the vertices in e be $u'_1, \ldots, u'_K$. Consider a K-flat $\mathcal{F}$ (subspace of $\mathbb{R}^n$ congruent to $\mathbb{R}^K$), containing $u'_1, \ldots, u'_K$. Applying Lemma 11.2.2 to the standard basis of $\mathcal{F}$, implies that the projection of r into $\mathcal{F}$ has the standard K-dimensional Gaussian distribution. Now since projecting r onto $\mathrm{Span}(u'_1, \ldots u'_K)$ preserves the inner products $\{\langle r, u'_i \rangle\}_i$, we may assume without loss of generality that $u'_1, \ldots, u'_K$ are vectors in $\mathbb{R}^K$, and the rounding scheme corresponds to picking a random K-dimensional Gaussian vector r, and proceeding as before.

Let U be the $K \times K$ matrix whose columns are the vectors $u'_1, \ldots, u'_K$ and μ represent the Gaussian measure in $\mathbb{R}^K$. Then the probability of e being monochromatic in the rounding is given by,

$$\mu\big(\{x \in \mathbb{R}^K \mid U^T x \geq 0\}\big) + \mu\big(\{x \in \mathbb{R}^K \mid U^T x < 0\}\big) = 2\mu\big(\{x \in \mathbb{R}^K \mid U^T x \geq 0\}\big) \tag{11.4}$$

In other words, this boils down to analyzing the Gaussian measure of the cone given by $U^T x \geq 0$. We thus take a necessary detour.

11.2.3 Gaussian Measure of Simplicial Cones

In this section we show how to bound the Gaussian measure of a special class of simplicial cones. This is one of the primary tools in our analysis of the previously introduced SDP relaxations. We first state some preliminaries.

Simplicial Cones and Equivalent Representations. A simplicial cone in $\mathbb{R}^K$, is given by the intersection of a set of K linearly independent halfspaces. For any simplicial cone with apex at position vector p, there is a unique set (upto changes in lengths), of K linearly independent vectors, such that the direct sum of $\{p\}$ with their positive span produces the cone. Conversely, a simplicial cone given by the direct sum of $\{p\}$ and the positive span of K linearly independent vectors, can be expressed as the intersection of a unique set of K halfspaces with apex at p. We shall refer to the normal vectors of the halfspaces above, as simply normal vectors of the cone, and we shall refer to the spanning vectors above, as simplicial vectors. We represent a simplicial cone C with apex at p, as (p, U, V) where U is a column matrix of unit vectors $u_1, \ldots, u_K$ (normal vectors), V is a column matrix of unit vectors $v_1, \ldots, v_K$ (simplicial vectors) and

$$C = \left\{ x \in \mathbb{R}^K \;\middle|\; u_1^T x \geq p_1, \ldots, u_K^T x \geq p_K \right\} = \left\{ p + x_1 v_1 + \cdots + x_K v_K \;\middle|\; x \geq 0, x \in \mathbb{R}^K \right\}$$

Switching Between Representations. Let $C \equiv (0, U, V)$ be a simplicial cone with apex at the origin. It is not hard to see that any v_i is in the intersection of exactly $K-1$ of the K halfspaces determined by U, and it is thus orthogonal to exactly $K-1$ vectors of the form u_j. We may assume without loss of generality that for any v_i, the only column vector of U not orthogonal to it, is u_i. Thus clearly $V^T U = D$ where D is some non-singular diagonal matrix. Let $A_U = U^T U$ and $A_V = V^T V$, be the gram matrices of the vectors. A_U and A_V are positive definite symmetric matrices with diagonal entries equal to one (they comprise of the pairwise inner products of the normal and simplicial vectors respectively). Also, clearly,

$$V = U^{-T} D, \qquad A_V = D A_U^{-1} D \tag{11.5}$$

One then immediately obtains: $(A_V)_{ij} = \frac{a_{ij}}{\sqrt{a_{ii} a_{jj}}}$, and $(A_U)_{ij} = \frac{-a'_{ij}}{\sqrt{a'_{ii} a'_{jj}}}$. where a_{ij} and a'_{ij} are the cofactors of the $(i,j)^{th}$ entries of A_U and A_V respectively.

Formulating the Integral. Let $C \equiv (0, U, V)$ be a simplicial cone with apex at the origin, and for $x \in \mathbb{R}^K$, let dx denote the differential of the standard K-dimensional Lebesgue measure. Then the Gaussian measure of C is given by,

$$
\begin{aligned}
&\frac{1}{\pi^{K/2}} \int\limits_{U^T x \geq 0} e^{-||x||_2^2}\,\mathrm{dx} \\
&= \frac{\det(V)}{\pi^{K/2}} \int\limits_{\mathbb{R}_+^K} e^{-||Vx||_2^2}\,\mathrm{dx} && \text{Subst. } x \leftarrow Vx \\
&= \frac{\det(V)}{\pi^{K/2}} \int\limits_{\mathbb{R}_+^K} e^{-\left|\left|U^{-T}Dx\right|\right|_2^2}\,\mathrm{dx} && \text{By Eq. (11.5)} \\
&= \frac{\det(V)}{\pi^{K/2}\det(D)} \int\limits_{\mathbb{R}_+^K} e^{-\left|\left|U^{-T}x\right|\right|_2^2}\,\mathrm{dx} && \text{Subst. } x \leftarrow Dx \\
&= \frac{1}{\pi^{K/2}\det(U)} \int\limits_{\mathbb{R}_+^K} e^{-\left|\left|U^{-T}x\right|\right|_2^2}\,\mathrm{dx} && (11.6) \\
&= \frac{1}{\pi^{K/2}\sqrt{\det(A_U)}} \int\limits_{\mathbb{R}_+^K} e^{-x^T A_U^{-1} x}\,\mathrm{dx} && (11.7)
\end{aligned}
$$

For future ease of use, we give a name to some properties.

Definition 11.2.3. *The* para-volume *of a set of vectors (resp. a matrix U), is the volume of the parallelotope determined by the set of vectors (resp. the column vectors of U).*

Definition 11.2.4. *The* sum-norm *of a set of vectors (resp. a matrix U), is the length of the sum of the vectors (resp. the sum of the column vectors of U).*

Walkthrough of Symmetric Case Analysis. We next state some simple identities that can be found in say, [Rog64], some of which were originally used by Daniels to show that the Gaussian measure of a symmetric cone in $\mathbb{R}^K$ of angle $\cos^{-1}(1/2)$ (between any two simplicial vectors) is $\frac{(1+o(1))\ e^{K/2-1}}{\sqrt{2}^{K+1}\sqrt{K}^{K-1}\sqrt{\pi}^K}$. We state these identities, while loosely describing the analysis of the symmetric case, to give the reader an idea of their purpose.

First note that the gram matrices S_U and S_V, of the symmetric cone of angle $\cos^{-1}(1/2)$ are given by:

$$S_U = (1 + 1/K)\mathrm{I} - \mathbf{1}\,\mathbf{1}^T/K \qquad S_V = (\mathrm{I} + \mathbf{1}\,\mathbf{1}^T)/2$$

Thus $x^T S_U^{-1} x$ is of the form,

$$\alpha\,||x||_1^2 + \beta\,||x||_2^2 \tag{11.8}$$

The key step is in linearizing the $||x||_1^2$ term in the exponent, which allows us to separate the terms in the multivariate integral into a product of univariate integrals, and this is easier to analyze.

Lemma 11.2.5 (Linearization)**.** $\sqrt{\pi}e^{-s^2} = \int_{-\infty}^{\infty} e^{-t^2+2its}\ \mathrm{dt}$

Observation 11.2.6. *Let $f : (-\infty, \infty) \mapsto \mathbb{C}$ be a continuous complex function. Then,* $\left|\int\limits_{-\infty}^{\infty} f(t)\ \mathrm{dt}\right| \leq \int\limits_{-\infty}^{\infty} |f(t)|\ \mathrm{dt}.$

On applying Lemma 11.2.5 to Eq. (11.6) in the symmetric case, one obtains a product of identical univariate complex integrals. Specifically, by Eq. (11.6), Eq. (11.8), and Lemma 11.2.5, we have the expression,

$$\begin{aligned}
\int\limits_{\mathbb{R}_+^K} e^{-\beta||x||_2^2-\alpha||x||_1^2}\ \mathrm{dx} &= \int\limits_{-\infty}^{\infty} e^{-t^2} \int\limits_{\mathbb{R}_+^K} e^{-\beta(x_1^2+\dots x_K^2)+2it\sqrt{\alpha}(x_1+\dots x_K)}\ \mathrm{dx}\ \mathrm{dt} \\
&= \int\limits_{-\infty}^{\infty} e^{-t^2} \left(\int\limits_{0}^{\infty} e^{-\beta s^2+2it\sqrt{\alpha}s}\ \mathrm{ds} \right)^K
\end{aligned}$$

The inner univariate complex integral is not readily evaluable. To circumvent this, one can change the line of integration so as to shift mass form the inner integral to the outer integral. Then we can apply the crude upper bound of Observation 11.2.6 to the inner integral, and by design, the error in our estimate is small.

Lemma 11.2.7 (Changing line of integration)**.** *Let $g(t)$ be a real valued function for real t. If, when interpreted as a complex function in the variable $t = a + ib$, $g(a + ib)$ is an entire function, and furthermore, $\lim\limits_{a\to\infty} g(a + ib) = 0$ for some fixed b, then we have,* $\int_{-\infty}^{\infty} g(t)\ \mathrm{dt} = \int_{-\infty}^{\infty} g(a + ib)\ \mathrm{da}$.

Squared L_1 Inequality. Motivated by the above linearization technique, we prove the following lower bound on quadratic forms in the positive orthant:

Lemma 11.2.8. *Consider any $K \times K$ matrix A, and $x \in \mathbb{R}_+^K$, such that x is in the column space of A. Let $A^\dagger$ denote the Moore-Penrose pseudo-inverse of A. Then, $x^T A^\dagger x \geq \frac{||x||_1^2}{\mathrm{sum}(A)}$.*

Proof. Consider any x in the positive orthant and column space of A. Let $v_1, \ldots, v_q$ be the eigenvectors of A corresponding to it's non-zero eigenvalues. We may express x in the form $x = \sum_i \beta_i v_i$, so that

$$||x||_1 = \langle \mathbf{1}, x \rangle = \sum_{i \in [q]} \beta_i \langle \mathbf{1}, v_i \rangle \Rightarrow ||x||_1^2 = (\sum_{i \in [q]} \beta_i \langle \mathbf{1}, v_i \rangle)^2.$$

We also have

$$x^T A^\dagger x = x^T (\sum_{i \in [q]} \lambda_i^{-1} v_i v_i^T) x = \sum_{i \in [q]} \lambda_i^{-1} \beta_i^2.$$

Now by Cauchy-Schwartz,

$$\left(\sum_i \lambda_i \langle \mathbf{1}, v_i \rangle^2 \right) \left(\sum_{i \in [q]} \lambda_i^{-1} \beta_i^2 \right) \geq ||x||_1^2 .$$

Therefore, we have

$$x^T A^\dagger x \geq \frac{||x||_1^2}{\sum_{i \in [q]} \lambda_i \langle \mathbf{1}, v_i \rangle^2} = \frac{||x||_1^2}{\mathbf{1}^T A \, \mathbf{1}} = \frac{||x||_1^2}{\mathrm{sum}(A)}.$$

$\square$

Equipped with all necessary tools, we may now prove our result.

Our Gaussian Measure Bound. Let $C \equiv (0, U, V)$ be a simplicial cone with apex at the origin. We now show an upper bound on the Gaussian measure of C that depends surprisingly on only the para-volume and sum-norm of U. Since Gaussian measure is at most 1, it is evident when viewing our bound that it can only be useful for simplicial cones wherein the sum-norm of their normal vectors is $O(\sqrt{K})$, and the para-volume of their normal vectors is not too small.

Theorem 11.2.1. *Let $C \equiv (0, U, V)$ be a simplicial cone with apex at the origin. Let $\ell = ||\sum_i u_i||_2$ (i.e. sum-norm of the normal vectors), then the Gaussian measure of C is at most* $\left(\frac{e}{2\pi K}\right)^{K/2} \frac{\ell^K}{\sqrt{\det(A_U)}}$

Proof. By the sum-norm property, the sum of entries of A_U is ℓ^2. Also by the definition of a simplicial cone, U, and cosequently A_U, must have full rank. Thus we may apply

Lemma 11.2.8 over the entire positive orthant. We proceed to analyze the multivariate integral in Eq. (11.6), by first applying Lemma 11.2.8 and then linearizing the exponent using Lemma 11.2.5. Post-linearization, our approach is similar to the presentation of Boeroeczky and Henk [BJH99]. We have,

$$
\begin{aligned}
I \leftarrow \int\limits_{\mathbb{R}_+^K} e^{-x^T A_U^{-1} x}\,\mathrm{dx} \quad &\leq \quad \int\limits_{\mathbb{R}_+^K} e^{-||x||_1^2/\ell^2}\,\mathrm{dx} && \text{by Lemma 11.2.8} \\
= \quad \ell^K \int\limits_{\mathbb{R}_+^K} e^{-||y||_1^2}\,\mathrm{dy} & && \text{Subst. } y \leftarrow x/\ell \\
= \frac{\ell^K}{\sqrt{\pi}} \int\limits_{\mathbb{R}_+^K} \int\limits_{-\infty}^{\infty} e^{-\,t^2\,+\,2it\sum\limits_{i\in[K]} y_i}\,\mathrm{dt}\,\mathrm{dy} & && \text{by Lemma 11.2.5} \\
= \frac{\ell^K}{\sqrt{\pi}} \int\limits_{-\infty}^{\infty} e^{-t^2} \prod_{i\in[K]} \left(\int\limits_0^{\infty} e^{2ity_i}\,\mathrm{dy_i} \right)\,\mathrm{dt} & && \\
= \frac{\ell^K}{\sqrt{\pi}} \int\limits_{-\infty}^{\infty} e^{-t^2} \left(\int_0^{\infty} e^{2its}\,\mathrm{ds} \right)^K \mathrm{dt} & && \\
= \frac{\ell^K}{\sqrt{\pi}} \int\limits_{-\infty}^{\infty} e^{-a^2+b^2-2abi} \left(\int_0^{\infty} e^{-2bs+2asi}\,\mathrm{ds} \right)^K \mathrm{da} & && \text{by Lemma 11.2.7} \\
= \frac{e^{K/2}\,\ell^K}{\sqrt{\pi}(2K)^{K/2}} \int\limits_{-\infty}^{\infty} e^{-a^2} \left(2be^{-ia/b} \int_0^{\infty} e^{-2bs+2asi}\,\mathrm{ds} \right)^K \mathrm{da} & && \text{Fixing } b = \sqrt{K/2} \\
= \frac{e^{K/2}\,\ell^K}{\sqrt{\pi}(2K)^{K/2}} \left| \int\limits_{-\infty}^{\infty} e^{-a^2} \left(2be^{-ia/b} \int_0^{\infty} e^{-2bs+2asi}\,\mathrm{ds} \right)^K \mathrm{da} \right| & && \text{Since expr. is positive} \\
\leq \frac{e^{K/2}\,\ell^K}{\sqrt{\pi}(2K)^{K/2}} \int\limits_{-\infty}^{\infty} e^{-a^2} \left(2b \int_0^{\infty} e^{-2bs}\,\mathrm{ds} \right)^K \mathrm{da} & && \text{By Observation 11.2.6} \\
= \frac{e^{K/2}\,\ell^K}{\sqrt{\pi}(2K)^{K/2}} \int\limits_{-\infty}^{\infty} e^{-a^2}\,\mathrm{da} \;=\; \frac{e^{K/2}\,\ell^K}{(2K)^{K/2}} & &&
\end{aligned}
$$

Lastly, the claim follows by substituting the above in Eq. (11.6). □

11.2.4 Analysis of Hyperplane Rounding given Strong Colorability

In this section we analyze the performance of random hyperplane rounding on K-uniform hypergraphs that are $(K+\ell)$-strongly colorable.

Theorem 11.2.2. *Consider any $(K+\ell)$-strongly colorable K-uniform hypergraph $H = (V, E)$. The expected fraction of monochromatic edges obtained by performing random hyperplane rounding on the solution of Relaxation 11.3, is $O\Big(\ \ell^{K-1/2}\big(\frac{e}{2\pi}\big)^{K/2}\ \frac{1}{K^{(K-1)/2}}\ \Big)$.*

Proof. Let U be any $K\times K$ matrix whose columns are unit vectors $u_1, \ldots, u_K \in Re^K$ that satisfy the edge constraints in Relaxation 11.3. Recall from Section 11.2.2, that to bound the probability of a monochromatic edge we need only bound the expression in Eq. (11.4) for U of the above form. By Relaxation 11.3, the gram matrix $A_U = U^T U$, is exactly, $A_U = (1+\alpha)\mathrm{I} - \alpha\,\mathbf{1}\,\mathbf{1}^T$ where $\alpha = \frac{1}{K+\ell-1}$. By matrix determinant lemma (determinant formula for rank one updates), we know

$$\det(A_U) = (1+\alpha)^K\left(1 - \frac{K\alpha}{1+\alpha}\right) \geq \left(\frac{\ell}{K+\ell}\right) = \Omega\left(\frac{\ell}{K}\right)$$

Further, Relaxation 11.3 implies the length of $\sum_i u_i$, is at most ℓ. The claim then follows by combining Eq. (11.4) with Theorem 11.2.1. □

Note. Being that any edge in the solution to the strong colorability relaxation corresponds to a symmetric cone, Theorem 11.2.2 is directly implied by prior work on the volume of symmetric spherical simplices. It is in the next section, where the true power of Theorem 11.2.1 is realized.

Remark. As can be seen from the asymptotic volume formula of symmetric spherical simplices, $\sqrt{\pi K/(2e)}$ is a sharp threshold for ℓ, i.e. when $\ell > (1+o(1))\sqrt{\pi K/(2e)}$, hyperplane rounding does worse than the naive random algorithm, and when $\ell < (1 - o(1))\sqrt{\pi K/(2e)}$, hyperplane rounding beats the naive random algorithm.

11.2.5 Analysis of Hyperplane Rounding given Rainbow Colorability

In this section we analyze the performance of random hyperplane rounding on K-uniform hypergraphs that are $(K-\ell)$-rainbow colorable.

Let U be the $K \times K$ matrix whose columns are unit vectors $u_1, \ldots, u_K \in \mathbb{R}^K$ satisfying the edge constraints in Relaxation 11.2. We need to bound the expression in Eq. (11.4) for U of the above form. While we'd like to proceed just as in Section 11.2.4, we are limited by the possibility of U being singular or the parallelotope determined by U having arbitrarily low volume (as u_1 can be chosen arbitrarily close to the span of $u_2, \ldots, u_K$ while still satisfying $||\sum_i u_i||_2 \leq \ell$).

While U can be bad with respect to our properties of interest, we will show that some subset of the vectors in U are reasonably well behaved with respect to para-volume and sum-norm.

Finding a Well Behaved Subset. We'd like to find a subset of U with high para-volume, or equivalently, a principal sub-matrix of A_U with reasonably large determinant. To this end, we express the gram matrix $A_U = U^T U$ as the sum of a symmetric skeleton matrix B_U and a residue matrix E_U. Formally, $E_U = A_U - B_U$ and $B_U = (1+\beta)I - \beta\, \mathbf{1}\,\mathbf{1}^T$ where $\beta = \frac{1}{K-\ell-1}$. We have (assuming $\ell = o(K)$), $\mathrm{sum}(A_U) \leq \ell^2$ and $\mathrm{sum}(B_U) = K - K(K-1)\beta = \frac{-\ell}{1-o(1)}$. Let $s \leftarrow \mathrm{sum}(E_U) \leq \ell^2 - \mathrm{sum}(B_U) = \ell^2 + \frac{\ell}{1-o(1)}$.

We further observe that E_U is symmetric, with all diagonal entries zero. Also since $u_1, \ldots, u_K$ satisfy Relaxation 11.2, all entries of E_U are non-negative.

By an averaging argument, at most cK^δ columns of E_U have column sums greater than $s/(cK^\delta)$ for some parameters δ, c to be determined later. Let $S \subsetneq [K]$ be the set of indices of the columns having the lowest $K - cK^\delta$ column sums. Let $\tilde{K} \leftarrow |S| = K - cK^\delta$, and let A_S, B_S, E_S be the corresponding matrices restricted to S (in both columns and rows).

Spectrum of B_S and E_S.

Observation 11.2.9. *For a square matrix X, let $\lambda_{\min}(X)$ denote its minimum eigenvalue. The eigenvalues of B_S are exactly $(1+\beta)$ with multiplicity $(\tilde{K}-1)$, and $(1+\beta-\tilde{K}\beta)$ with multiplicity 1. Thus $\lambda_{\min}(B_S) = 1+\beta-\tilde{K}\beta$. This is true since B_S merely shifts all eigenvalues of $-\beta\,\mathbf{1}\,\mathbf{1}^T$ by $1+\beta$.*

While we don't know as much about the spectrum of E_S, we can still say some useful things.

Observation 11.2.10. *Since E_S is non-negative, by Perron-Frobenius theorem, its spectral radius is equal to its max column sum, which is at most $s/(cK^\delta)$. Thus $\lambda_{\min}(E_S) \geq -s/(cK^\delta)$.*

Now that we know some information about the spectra of B_S and E_S, the next natural step is to consider the behaviour of spectra under matrix sums.

Spectral properties of Matrix sums. The following identity is well known.

Observation 11.2.11. *If X and Y are symmetric matrices with eigenvalues $x_1 > x_2 > \cdots > x_m$ and $y_1 > y_2 > \cdots > y_m$ and the eigenvalues of $A + B$ are $z_1 > z_2 > \cdots > z_m$, then*

$$\forall\, 0 \le i + j \le m, \;\; z_{m-i-j} \ge x_{m-i} + y_{m-j}.$$

In particular, this implies $\lambda_{\min}(X + Y) \ge \lambda_{\min}(X) + \lambda_{\min}(Y)$.

We may finally analyze the spectrum of A_S.

Properties of A_S.

Observation 11.2.12 (Para-Volume)**.** *Let the eigenvalues of A_S be $a_1 > a_2 > \cdots > a_{\tilde{K}}$ By Observation 11.2.9, Observation 11.2.10, and Observation 11.2.11 we have (Assuming $\ell < cK^{\delta/2}$),*

$$\begin{aligned} \lambda_{\min}(A_S) = a_{\tilde{K}} \quad &\ge \quad 1 + \beta - \tilde{K}\beta - \frac{s}{cK^\delta} = \quad \frac{c}{K^{1-\delta}} - \frac{\ell^2}{cK^\delta} - o(1) \\ a_2, a_3, \ldots, a_{\tilde{K}-1} \quad &\ge \quad 1 + \beta - \frac{s}{cK^\delta} \qquad\quad = \quad 1 - \frac{\ell^2}{cK^\delta} - o(1) \end{aligned}$$

Consequently,

$$\det(A_S) \ge \left(\frac{c}{K^{1-\delta}} - \frac{\ell^2}{cK^\delta} - o(1)\right)\left(1 - \frac{\ell^2}{cK^\delta} - o(1)\right)^{\tilde{K}} \ge \left(\frac{c}{K^{1-\delta}} - \frac{\ell^2}{cK^\delta} - o(1)\right) e^{-K}$$

In particular, note that A_S is non-singular and has non-negligible para-volume when

$$\frac{\ell^2}{cK^\delta} = \frac{c}{2K^{1-\delta}}, \quad \textit{i.e. } \ell \approx cK^{\delta-1/2} \quad \textit{or, } \; \delta \approx \frac{1}{2}\frac{\log(\ell/c)}{\log K}$$

Observation 11.2.13 (Sum-Norm)**.** *Since E_U is non-negative,* $\mathrm{sum}(E_S) \le \mathrm{sum}(E_U) = s$. *Also we know that the sum of entries of A_S is*

$$\mathrm{sum}(B_S) + \mathrm{sum}(E_S) = \tilde{K}(1+\beta) - \tilde{K}(\tilde{K}-1)\beta + s \le cK^\delta + s \tag{11.9}$$

The Result. We are now equipped to prove our result.

Theorem 11.2.3. *For $\ell < \sqrt{K}/100$, consider any $(K-\ell)$-rainbow colorable K-uniform hypergraph $H = (V, E)$. Let $\theta = 1/2 + \log(\ell)/\log(K)$ and $\eta = 19(1-\theta)/40$. The expected fraction of monochromatic edges obtained by performing random hyperplane rounding on the solution of Relaxation 11.2, is at most*

$$\frac{1}{2.1^K\, K^{\eta K}}$$

Proof. Let U be any $K \times K$ matrix whose columns are unit vectors $u_1, \ldots, u_K \in \mathbb{R}^K$ that satisfy the edge constraints in Relaxation 11.3. Recall from Section 11.2.2, that to bound the probability of a monochromatic edge we need only bound the expression in Eq. (11.4) for U of the above form.

By Section 11.2.5, we can always choose a matrix U_S whose columns $\tilde{u}_1, \ldots, \tilde{u}_{\tilde{K}}$ are from the set $\{u_1, \ldots, u_K\}$, such that the gram matrix $A_S = U_S^T U_S$ satisfies Eq. (11.9) and Observation 11.2.12. Clearly the probability of all vectors in U being monochromatic is at most the probability of all vectors in U_S being monochromatic.

Thus just as in Section 11.2.2, to find the probability of U_S being monochromatic, we may assume without loss of generality that we are performing random hyperplane rounding in $\mathbb{R}^{\tilde{K}}$ on any $\tilde{K}$-dimensional vectors $\tilde{u}_1, \ldots, \tilde{u}_{\tilde{K}}$ whose gram (pairwise inner-product) matrix is the aforementioned A_S.

Specifically, by combining Eq. (11.9) and Observation 11.2.12 with Theorem 11.2.1, our expression is at most:

$$\left(\frac{e}{2\pi}\right)^{\tilde{K}/2}\left(\frac{cK^{\delta}+s}{K}\right)^{\tilde{K}/2}\frac{1}{\sqrt{det(A_U)}} \leq 3.2^{\tilde{K}/2}\left(\frac{(1-o(1))c}{K^{1-\delta}}\right)^{\tilde{K}/2} \leq \frac{1}{2.1^K\, K^{(1-c)(1-\delta)K}}$$

assuming $c = 1/20$, $\delta \geq 1/2$ and $\ell < \sqrt{K}/100$ (constraint on ℓ ensures that non-singularity conditions of Observation 11.2.12 are satisfied). The claim follows. $\square$

Remark. Yet again we see a threshold for ℓ, namely, when $\ell < \sqrt{K}/100$, hyperplane rounding beats the naive random algorithm, and for $\ell = \Omega(\sqrt{K})$, it fails to do better. In fact, as we'll see in the next section, assuming the UGC, we show a hardness result when $\ell = \Omega(\sqrt{K})$.

11.3 Approximate MIN COLORING

In this section, we provide approximation algorithms for the MIN COLORING problem under strong colorability, rainbow colorability, and low discrepancy assumptions. Our approach is standard, namely, we first apply degree reduction algorithms followed by the usual paradigm pioneered by Karger, Motwani and Sudan [KMS98b], for coloring bounded degree (hyper)graphs. Consequently, our exposition will be brief and non-linear.

In the interest of clarity, all results henceforth assume the special cases of Discrepancy 1, or $(K-1)$-rainbow colorability, or $(K+1)$-strong colorability. All arguments generalize easily to the cases parameterized by l.

11.3.1 Approximate MIN COLORING in Bounded Degree Hypergraphs

The Algorithm. INPUT: K-uniform hypergraph $H = ([n], E)$ with max-degree t and m edges, having Discrepancy 1, or being $(K-1)$-rainbow colorable, or being $(K+1)$-strong colorable.

1. Let $u_1, \ldots, u_n$ be a solution to the SDP relaxation from Section 11.2.1 corresponding to the assumption on the hypergraph.
2. Let H_1 be a copy of H, and let γ, τ be parameters to be determined shortly.
3. Until no vertex remains in the hypergraph, Repeat:
   Find an independent set $\mathcal{I}$ in the residual hypergraph, of size at least γn by repeating the below process until $|\mathcal{I}| \geq \gamma n$:
   (A) Pick a random vector r from the standard multivariate normal distribution.
   (B) For all i, if $\langle u_i, r \rangle \geq \tau$, add vertex i to $\mathcal{I}$.
   (C) For every edge e completely contained in $\mathcal{I}$, delete any single vertex in e, from $\mathcal{I}$.

   Color $\mathcal{I}$ with a new color and remove $\mathcal{I}$ and all edges involving vertices in $\mathcal{I}$, from H_1.

Analysis. First note that by Lemma 11.2.2, for any fixed vector a, $\langle a, r\rangle$ has the distribution $\mathcal{N}(0,1)$. Note that all SDP formulations in Section 11.2.1 satisfy,

$$\left\|\sum_{j\in[K]} u_{i_j}\right\|_2 \leq 1 \tag{11.10}$$

Now consider any edge $e = (i_1, \ldots, i_K)$. In any fixed iteration of the inner loop, the probability of e being contained in $\mathcal{I}$ at Step (B), is at most the probability of

$$\langle r, \sum_{j\in[K]} u_{i_j}\rangle \geq K\tau$$

However, by Lemma 11.2.2 and Eq. (11.10), the inner product above is dominated by the distribution $\mathcal{N}(0,1)$. Thus in any fixed iteration of the inner loop, let H_1 have n_1 vertices and m_1 edges, we have

$$\begin{aligned}
\mathbb{E}[\mathcal{I}] &\geq n_1\Phi(\tau) - m_1\Phi(K\tau) \\
&\geq n_1 e^{-\tau^2/2} - \frac{n_1 t}{K} e^{-K^2\tau^2/2} \\
&= \Omega(\gamma n_1) \qquad\qquad \text{setting, } \tau^2 = \frac{2\log t}{K^2-1}, \ \text{ and } \ \gamma = t^{-1/(K^2-1)}
\end{aligned}$$

Now by applying Markov's inequality to the vertices not in $\mathcal{I}$, we have, $\Pr[|\mathcal{I}| < \gamma n_1] \leq 1 - \Omega(\gamma)$. Thus for a fixed iteration of the outer loop, with high probability, the inner loop doesn't repeat more than $O(\log n_1/\gamma)$ times.

Lastly, the outermost loop repeats $O(\log n/\gamma)$ times, using one color at each iteration. Thus with high probability, in polynomial time, the algorithm colors H with

$$t^{\frac{1}{K^2-1}} \log n \ \text{ colors.}$$

Important Note. We can be more careful in the above analysis for the rainbow and strong colorability cases. Specifically, the crux boils down to finding the gaussian measure of the cone given by $\left\{x \,\middle|\, U^T x \geq \tau\right\}$ instead of zero. Indeed, on closely following the proof of Theorem 11.2.1 we obtain for strong and rainbow coloring respectively (assuming max-degree n^K),

$$n^{\frac{1}{K}\left(1-\frac{3\beta}{2}\right)} \log n \qquad \text{and} \qquad n^{\frac{1}{K}\left(1-\frac{5\beta}{4}\right)} \log n, \qquad \text{where } \ \beta = \frac{\log K}{\log n}$$

While these improvements are negligible for small K, they are significant when K is reasonably large with respect to n.

11.3.2 Main MIN COLORING Result

Combining results from Section 11.3.1 with our degree reduction approximation schemes from the forthcoming sections, we obtain the following.

Theorem 11.3.1. *Consider any K-uniform hypergraph $H = (V, E)$ with n vertices. In $n^{c+O(1)}$ time, one can color H with*

$$\min\left\{\left(\frac{n}{c\log n}\right)^{\alpha},\ n^{\frac{1}{K}\left(1-\frac{3\beta}{2}\right)}, \left(\frac{m}{n}\right)^{\frac{1}{K^2}}\right\}\log n \textit{ colors,} \quad \textit{if } H \textit{ is } (K+1)\textit{-strongly colorable.}$$

$$\min\left\{\left(\frac{n}{c}\right)^{\alpha},\ n^{\frac{1}{K}\left(1-\frac{5\beta}{4}\right)}, \left(\frac{m}{n}\right)^{\frac{1}{K^2}}\right\}\log n \textit{ colors,} \qquad \textit{if } H \textit{ is } (K-1)\textit{-rainbow colorable.}$$

$$\min\left\{\left(\frac{n}{c}\right)^{\alpha},\ \left(\frac{m}{n}\right)^{\frac{1}{K^2}}\right\}\log n \textit{ colors,} \qquad \textit{if } H \textit{ has discrepancy } 1.$$

where, $\alpha = \dfrac{1}{K+2-o(1)}, \quad \beta = \dfrac{\log K}{\log n}$

Remark. In all three promise cases the general polytime min-coloring guarantee parameterized by ℓ, is roughly $n^{\ell^2/K}$. Thus, the threshold value of ℓ, for which standard min-coloring techniques improve with K, is $o(\sqrt{K})$.

Degree Reduction Schemes under Promise. Wigderson [Wig83] and Alon et al. [AKMH96] studied degree reduction in the cases of 3-colorable graphs and 2-colorable hypergraphs, respectively. Assuming our proposed structures, we are able to combine some simple combinatorial ideas with counterparts of the observations made by Wigderson and Alon et al., to obtain degree reduction approximation schemes. Such approximation schemes are likely not possible assuming only 2-colorability.

11.3.3 Degree Redution under Strong Colorability

Let $H = \left(V, E \subseteq \binom{V}{K}\right)$ be a K-uniform $(K+1)$-strongly colorable hypergraph with n vertices and m edges. In this section, we give an algorithm that in $n^{c+O(1)}$ time, partially colors H with $3n(K+1)\log K/(t^{1/(K-1)}c\log n)$ colors, such that no edge in the colored subgraph is monochromatic, and furthermore, the subgraph induced by the the uncolored vertices has max-degree t.

The following observations motivate the structure of our algorithm.

Observation 11.3.1. *For any $(K+1)$-strong coloring $f : V \mapsto [K+1]$, of a K-uniform hypergraph H, and any subset of vertices $\overline{V}$ satisfying, $\forall u, v \in \overline{V}, f(u) = f(v) = j$ (all of the same color), the subgraph F of H, induced by $N(\overline{V})$, is K-uniform and K-strongly colorable. This is because f is a strong coloring of F, and moreover, $\forall\, v \in N(\overline{V}),\ f(v) \neq j$, since v has a neighbor in $\overline{V}$ with color j. Thus we can 2-color such a subgraph F in polynomial time.*

Observation 11.3.2. *By Observation 11.3.1, in order to $3(K+1)$-color the subgraph induced by $\overline{V} \cup N(\overline{V})$ for an arbitrary subset $\overline{V}$ of vertices, we need only search through all possible $(K+1)$-colorings of $\overline{V}$, and then attempt to 2-color the neighborhood of each color class with two new colors. This process will always terminate with some proper coloring of $\overline{V} \cup N(\overline{V})$.*

We are now prepared to state the algorithm.

The Algorithm *SCDegreeReduce*.

1. Let H_1 be a copy of H.
2. **While** H_1 contains a vertex of degree greater than t:
   - (A) Let H_2 be a copy of H_1.
   - (B) Sequentially pick arbitrary vertices $\overline{V} = \{v_1, v_2 \dots v_s\}$ of degree at least t from H_2, wherein we remove from H_2 the vertices $\{v_i\} \cup N(v_i)$ and all involved edges, after picking v_i and before picking v_{i+1}. We only stop when we have either picked $c \log n / \log K$ vertices, or H_2 has max-degree t.
   - (C) For every possible assignment of $K+1$ new colors $\{c_1, \dots c_{K+1}\}$ to the vertices in $\overline{V}$:
     - (C1) Let $C_i = \left\{u \;\middle|\; v \in \overline{V},\ color(v) = c_i,\ u \in N_{H_1}(v)\right\}$. Then for each $i \in [K+1]$, 2-color the subgraph of H_1 induced by $N_{H_1}(C_i)$ using two new colors and the proper 2-coloring algorithm for r-uniform, r-strongly colorable graphs.

(C2) **If** no edge is monochromatic:
Stick with this $3(K+1)$-coloring of $\overline{V} \cup N_{H_1}(\overline{V})$, remove $\overline{V} \cup N_{H_1}(\overline{V})$ and all edges containing any of these vertices, from H_1, and stop iterating through assignments of $\overline{V}$.

(C3) **If** some edge is monochromatic:
Discard the coloring and continue iterating through assignments of $\overline{V}$.

End While

3. **Output** the partial coloring of H and the residual graph H_1 of max-degree t.

The Result.

Theorem 11.3.2. *Let $H = \left(V, E \subseteq \binom{V}{K}\right)$ be a K-uniform $(K+1)$-strongly colorable hypergraph with n vertices. Algorithm 11.3.4 partially colors H in $n^{c+O(1)}$ time, with at most $\frac{3n(K+1)\log K}{t^{1/(K-1)}c\log n}$ colors, such that:*

1. *The subgraph of H induced by the colored vertices has no monochromatic vertices.*
2. *The subgraph of H induced by the uncolored vertices has maximum degree t.*

Proof. Observation 11.3.1 combined with the fact that step (C1) uses two new colors for each C_i, establishes that step (C) of Algorithm 11.3.4 will always terminate with some proper coloring of $\overline{V} \cup N_{H_1}(\overline{V})$. Furthermore, any edge intersecting $\overline{V_1} \cup N_{H_1}(\overline{V_1})$ and $\overline{V_2} \cup N_{H_1}(\overline{V_2})$ for V_1 and V_2 taken from different iterations of Algorithm 11.3.4, cannot be monochromatic since we use new colors in each iteration. Thus the partial coloring is proper.

For the claim on number of colors, observe that a vertex of degree at least t, must have at least $(K-1)t^{1/(K-1)}$ distinct neighbors. Thus step (C) can be run at most $n/t^{1/(K-1)}$ times, using $3(K+1)$ new colors each time.

Lastly for the runtime, note that for each run of step (C), there are at most $(K+1)^{c\log n/\log K} = n^{c+O(1)}$ assignments to try, and the rest of the work takes $n^{O(1)}$ time. $\square$

Remark. We contrast Theorem 11.3.2 with the results of Alon et al. [AKMH96], who give a polynomial time algorithm for degree reduction in 2-colorable K-uniform hypergraphs using $O(n/t^{1/(K-1)})$ colors. The strong coloring property, gives us additional power, namely, we obtain an approximation scheme, and furthermore, for constant c, Theorem 11.3.2 uses fewer colors than the result of Alon et al., by a factor of about $K \log n / \log K$.

The arguments in this section and the next are readily generalizable - One can modify the degree reduction algorithm, such that the bound on colors used, would be a function of the strong colorability parameter of the hypergraph.

11.3.4 Degree Reduction under Low Discrepancy

For odd K, let $H = (V, E)$ be a K-uniform hypergraph with n vertices, that admits a discrepancy 1 coloring. In this section, we give an algorithm that in $n^{c+O(1)}$ time, partially colors H with $3n(K+1)/(t^{\frac{1}{K-1}} c \log n)$ colors, such that no edge in the induced colored subgraph is monochromatic, and furthermore, the subgraph induced by the the uncolored vertices has max-degree t.

First, we present a warmup algorithm that exposes the key ideas. The following observations motivate the structure of our algorithm.

Observation 11.3.3. *For any discrepancy* 1 *coloring $f : V \mapsto \{-1, 1\}$, of a K-uniform hypergraph H, and any size $K-1$ subset of vertices S, we have:*

(A) If $N(S)$ is an independent set, we can properly 2-color the subgraph induced by $S \cup N(S)$.

(B) If $N(S)$ contains an edge, then the set S has discrepancy 0 in the coloring f. This is because, an edge cannot be monochromatic in the coloring f, and by assumption, S must be have a neighbor with color -1 and a neighbor with color $+1$.

Though Observation 11.3.3 and Observation 11.3.1 are functionally similar, the two-pronged nature of Observation 11.3.3 almost wholly accounts for the gap in power between the respective degree reduction algorithms. Intuitively, the primary weakness comes from the fact that $N(S)$ being an independent set tells us nothing about the discrepancy of S.

Nevertheless, we may still exploit some aspects of this observation.

Observation 11.3.4. *Consider any discrepancy* 1 *coloring* $f : V \mapsto \{-1, 1\}$*, of a* K*-uniform hypergraph* H*, and any set of subsets* $S_1, \ldots S_m$ *each of size* $(K-1)$ *and discrepancy* 0 *in the coloring* f*. The* $(K-1)$*-uniform hypergraph* F *with vertex set* $\bigcup_i S_i$ *and edge set* $\{S_1, \ldots S_m\}$*, has a discrepancy* 0 *coloring (*f*). Thus we can properly* 2*-color* F *in polynomial time.*

We are now ready to state the warmup algorithm, whose correctness is evident from Observation 11.3.3 and Observation 11.3.4

Warmup Algorithm.

1. Let H_1 be a copy of H, and set $\mathrm{MARKED} \leftarrow \phi$
2. **While** H_1 contains a size $(K-1)$ subset S such that $N_{H_1}(S) > t$:
   (A) **If** $N_{H_1}(S)$ contains an edge:
   Delete from H_1 all edges that completely contain S. Also, add S to MARKED.
   (B) **If** $N_{H_1}(S)$ is an independent set:
   Use 2 new colors, color S one color and $N_{H_1}(S)$ the other, remove $S \cup N_{H_1}(S)$ and all edges containing any of these vertices from H_1.
   End While
3. Let F be the $(K-1)$-uniform hypergraph whose vertex set is the union of the sets in MARKED, and whose edge set is MARKED. Using 2 new colors, properly 2-color the vertices of F using the 2-coloring algorithm for discrepancy 0 hypergraphs. Remove these vertices and all involved edges, from H_1.
4. **Output** the partial coloring of H and the residual graph H_1 of max-degree t.

The Algorithm *LDDegreeReduce*.

1. Let H_1 and H_2 be copies of H, $\mathrm{MARKED} \leftarrow \phi$ and $T \leftarrow \phi$.
2. **While** H_2 contains a size $(K-1)$ subset S of vertices, such that $|N_{H_2}(S)| > t$:
   (A) Delete $N_{H_2}(S)$ and all edges involving these vertices, from H_2.
   (B) **If** $N_{H_1}(S)$ contains an edge:
   Delete from H_1 all edges that completely contain S. Also, add S to MARKED.

(C) **If** $N_{H_1}(S)$ is an independent set:
Add S to T.

(D) For every size c subset $\overline{V} = \{S'_1, \ldots, S'_c\}$ of T:

Fix two new colors c_1, c_2.
For every possible assignment of c_1, c_2 to $\overline{V}$, such that each S'_i has discrepancy 2, (We define $bias(S'_i) = c_1$ (resp. c_2) for coloring bias towards c_1 (resp. c_2)):

(D1) For $i = 1, 2$, let $C_i = \{u \mid S' \in \overline{V},\ bias(S') = c_i,\ u \in N_{H_1}(S')\}$. Then color $N_{H_1}(C_1)$ with just c_2 and $N_{H_1}(C_2)$ with just c_1.

(D2) **If** no edge is monochromatic:
Stick with this proper 2-coloring of the vertices in $\overline{V}, N_{H_1}(\overline{V})$.
Remove $\overline{V}$ from T, i.e. $T \leftarrow T \setminus \overline{V}$
Remove $\bigcup_i (S'_i \cup N_{H_1}(S'_i))$ and all edges containing any of these vertices, from H_1 and H_2, and stop iterating through assignments of $\overline{V}$.

(D3) **If** some edge is monochromatic:
Discard the coloring and continue iterating through assignments of $\overline{V}$.

End While

3. For every subset B of T of size less than c:
   (1) Let $A \leftarrow T \setminus B$.

   (2) Using two new colors, run the proper 2-coloring algorithm for discrepany zero hypergraphs on the $(K-1)$-uniform hypergraph whose edge set is A.

   (3) Using two new colors, iterate through all assignments of B, and attempt to 2-color $N_{H_1}(B)$ just as in Step (D1).

   (4) **If** both colorings succeed:
   Stick with this proper 2-coloring of the vertices in $T, N_{H_1}(B)$.

   Remove from H_1 the vertices $\bigcup_{S' \in A} S'$ and $\bigcup_{S' \in B}(S' \cup N_{H_1}(S'))$ and all edges involving any of these vertices, and stop iterating through subsets of T.

   (5) **If** either coloring fails:
   Discard the coloring and continue iterating through subsets of T.

4. **Output** the proper partial coloring of H and the residual graph H_1 of max-degree $\binom{n-1}{K-2}t$.

The Result.

Theorem 11.3.3. *For odd K, let $H = \left(V, E \subseteq \binom{V}{K}\right)$ be a K-uniform discrepancy 1 hypergraph with n vertices. Algorithm 11.3.4 partially colors H in $n^{c+O(1)}$ time, with at most $2n/ct$ colors, such that:*

1. *The subgraph of H induced by the colored vertices has no monochromatic vertices.*
2. *The subgraph of H induced by the uncolored vertices has maximum degree $\binom{n-1}{K-2}t$.*

Proof. The proof goes very similarly to that of Theorem 11.3.2, thus we just state the key observations required to complete the proof.

(A) In any discrepancy 1 coloring of H, any size $K-1$ set S' either has discrepancy 2, or discrepancy 0.

(B) Consider any discrepancy 1 coloring of H. If a size $K-1$ set S' has discrepancy 2, then $N(S')$ is monochromatic.

(C) At the end of any iteration of Step 2., there is no size c subset of T such that every set in the subset has discrepancy 2 in any discrepancy 1 coloring of H.

(D) When we reach Step 3., at least $|T| - c$ sets in T, all have discrepancy 0 in EVERY discrepancy 1 coloring of H.

□

11.3.5 Degree Reduction under Rainbow Colorability

Now, the equivalent algorithm in the case of rainbow colorability is virtually identical to that of Section 11.3.4. Thus we merely state the result.

Theorem 11.3.4. *Let $H = \left(V, E \subseteq \binom{V}{K}\right)$ be a K-uniform $(K-1)$-rainbow colorable hypergraph with n vertices. Algorithm 11.3.4 partially colors H in $n^{c+O(1)}$ time, with at most $(K-1)n/ct$ colors, such that:*

1. *The subgraph of H induced by the colored vertices has no monochromatic vertices.*
2. *The subgraph of H induced by the uncolored vertices has maximum degree $\binom{n-1}{K-2}t$.*

Part IV

Subgraph Transversal and Graph Partitioning

Chapter 12

Subgraph Transversal Overview

12.1 Introduction

Given a collection of subsets $S_1, ..., S_m$ of the underlying set U, the SET TRANSVERSAL problem asks to find the smallest subset of U that intersects every S_i, and the SET PACKING problem asks to find the largest subcollection $S_{i_1}, ..., S_{i_{m'}}$ which are pairwise disjoint.[1] It is clear that optimum of the former is always at least that of the latter (i.e. weak duality holds). Studying the (approximate) reverse direction of the inequality (i.e. strong duality) as well as the complexity of both problems for many interesting classes of set systems is arguably the most studied paradigm in combinatorial optimization.

This work focuses on set systems where the size of each set is bounded by a constant k. With this restriction, MINIMUM SET TRANSVERSAL and MAXIMUM SET PACKING are known as HYPERGRAPH VERTEX COVER and k-SET PACKING, respectively. This assumption significantly simplifies the problem since there are at most n^k sets. While there is a simple factor k-approximation algorithm for both problems, it is NP-hard to approximate k-HYPERGRAPH VERTEX COVER and k-SET PACKING within a factor less than $k - 1$ [DGKR05] and $O(\frac{k}{\log k})$ [HSS06] respectively.

Given a large graph $G = (V_G, E_G)$ and a fixed graph $H = (V_H, E_H)$ with k vertices, one of the natural attempts to further restrict set systems is to set $U = V_G$, and take the collection of subsets to be all *copies* of H in G (formally defined in the next subsection). This

[1]These problems are called many different names in the literature. SET TRANSVERSAL is also called HYPERGRAPH VERTEX COVER, SET COVER (of the dual set system), and HITTING SET. SET PACKING is also called HYPERGRAPH MATCHING. We try to use TRANSVERSAL / PACKING unless another name is established in the literature (e.g. k-HYPERGRAPH VERTEX COVER).

natural representation in graphs often results in a deeper understanding of the underlying structure and better algorithms, with MAXIMUM MATCHING ($H = K_2$) being the most well-known example. Kirkpatrick and Hell [KH83] proved that MAXIMUM MATCHING is essentially the only case where H-PACKING can be solved exactly in polynomial time — unless H is the union of isolated vertices and edges, it is NP-hard to decide whether V_G can be partitioned into k-subsets each inducing a subgraph containing H. A similar characterization for the edge version (i.e. $U = E_G$) was obtained much later by Dor and Tarsi [DT97].

We extend these results by studying the approximability of H-TRANSVERSAL and H-PACKING. We use the term *strong inapproximability* to denote NP-hardness of approximation within a factor $\Omega(k/\text{polylog}(k))$. We give a simple sufficient condition that implies strong inapproximability — if H is 2-vertex connected, H-TRANSVERSAL and H-PACKING are almost as hard to approximate as k-HYPERGRAPH VERTEX COVER and k-SET PACKING. We also show that there is a 1-connected H where H-TRANSVERSAL admits an $O(\log k)$-approximation algorithm, so 1-connectivity is not sufficient for strong inapproximability for H-TRANSVERSAL. It is an interesting open problem whether 1-connectivity is enough to imply strong inapproximability of H-PACKING, or there is a class of connected graphs where H-PACKING admits a significantly nontrivial approximation algorithm (e.g. factor k^ϵ for some $\epsilon < 1$).

One of our algorithms introduces another natural problem called k-VERTEX SEPARATOR, where given a graph $G = (V, E)$, we want remove the fewest number of vertices such that every connected component has strictly less than k verticese. k-VERTEX SEPARATOR can be regarded as a special case of a more general class of problems where we are given a graph G and a set of pattern graphs $\mathcal{H}$ with k vertices and asked to remove the minimum number of vertices to ensure G does not have any graph in $\mathcal{H}$ as a subgraph (in this case $\mathcal{H}$ is the set of all connected graphs with k vertices). Note that it is still a special case of k-HYPERGRAPH VERTEX COVER and admits a simple k-approximation algorithm. k-VERTEX SEPARATOR also can be considered as a special case of *graph partitioning* problems, and our algorithm exploits this connection between graph partitioning and H-TRANSVERSAL.

Our results give a unified answer to questions left open in many independent works studying a special cases. In the subsequent sections, we state our main results, review related work, and state potential future directions.

12.2 Problems and Our Results

12.2.1 H-TRANSVERSAL and H-PACKING

Hardness. Given an undirected graphs $G = (V_G, E_G)$ and $H = (V_H, E_H)$ with $|V_H| = k$, we define the following problems.

- H-TRANSVERSAL asks to find the smallest $F \subseteq V_G$ such that the subgraph of G induced by $V_G \setminus F$ does not have H as a subgraph.

- H-PACKING asks to find the maximum number of pairwise disjoint k-subsets of $S_1, ..., S_m$ of V_G such that the subgraph induced by each S_i has H as a subgraph.

Our main result states that 2-connectivity of H is sufficient to make H-TRANSVERSAL and H-PACKING hard to approximate.

Theorem 12.2.1. *If H is a 2-vertex connected with k vertices, unless* $\mathsf{NP} \subseteq \mathsf{BPP}$, *no polynomial time algorithm approximates H-*TRANSVERSAL *within a factor better than $k - 1$, and H-*PACKING *within a factor better than* $\Omega(\frac{k}{\log^7 k})$.

Our hardness results for transversal problems rely on hardness of k-HYPERGRAPH VERTEX COVER which is NP-hard to approximate within a factor better than $k - 1$ [DGKR05]. Our hardness results for packing problems rely on hardness of MAXIMUM INDEPENDENT SET on graphs with maximum degree k and girth strictly *greater than* g (MIS-k-g). Almost tight inapproximability of MAXIMUM INDEPENDENT SET on graphs with maximum degree k (MIS-k) is recently proved in Chan [Cha13], which rules out an approximation algorithm with ratio better than $\Omega(\frac{k}{\log^4 k})$. We are able to extend his result to MIS-k-g with losing only a polylogarithmic factor. All applications in this part require $g = \Theta(k)$.

Theorem 12.2.2. *For any constants k and g, unless* $\mathsf{NP} \subseteq \mathsf{BPP}$, *no polynomial time algorithm approximates* MIS-k-g *within a factor of* $\Omega(\frac{k}{\log^7 k})$.

We remark that assuming the Unique Games Conjecture (UGC) slightly improves our hardness ratios through better hardness of k-HYPERGRAPH VERTEX COVER [KR08] and MIS-k [AKS09], and even simplifies the proof for some problems (e.g. k-Clique Transversal) through structured hardness of k-HYPERGRAPH VERTEX COVER [BK10].

Algorithms. Let k-Star denote $K_{1,k-1}$, the complete bipartite graph with 1 and $k-1$ vertices on each side. The following theorem shows that k-STAR TRANSVERSAL admits a good approximation algorithm, so the assumption of 2-connectedness in Theorem 12.2.1 is required for strong inapproximability of H-TRANSVERSAL.

Theorem 12.2.3. k-STAR TRANSVERSAL *can be approximated within a factor of* $O(\log k)$ *in time* $\mathrm{poly}(n, k)$.

This algorithmic result matches $\Omega(\log k)$-hardness of k-STAR TRANSVERSAL via a simple reduction from MINIMUM DOMINATING SET on degree-k graphs [CC08]. This problem has the following equivalent but more natural interpretation: given a graph $G = (V_G, E_G)$, find the smallest $F \subseteq V_G$ such that the subgraph induced by $V_G \setminus F$ has maximum degree at most $k-2$. Our algorithm, which uses iterative roundings of 2-rounds of Sherali-Adams hierarchy of linear programming (LP) followed by a simple greedy algorithm for *Constrained Set Cover*, is also interesting in its own right.

We also obtain a positive result for k-PATH TRANSVERSAL. Let $l(G)$ be the length of the longest path of G including both endpoints (e.g., length of a single edge is 2). Given a graph $G = (V, E)$ and $k \in \mathbb{N}$, k-PATH TRANSVERSAL asks to find the smallest subset $S \subseteq V$ such that $l(G|_{V \setminus S}) < k$. Finding a path of length k has played a central role in devopment of FPT algorithms — it is NP-hard to do for general k, but there are various algorithms that run in time $2^{O(k)} n^{O(1)}$ using color coding or algebraic algorithms.

k-PATH TRANSVERSAL is motivated by applications in transportation / wireless sensor networks, and has also been actively studied as k-PATH VERTEX COVER or P_k-HITTING SET [TZ11, BKKS11, BJK+13, Cam15, Kat16] in terms of their approximability and fixed parameter tractability. Tu and Zhou [TZ11] gave a 2-approximation algorithm for 3-PATH TRANSVERSAL. Camby [Cam15] recently gave a 3-approximation algorithm for 4-PATH TRANSVERSAL. In the same doctoral thesis, Camby [Cam15] asked whether we can get $(1-\delta)k$-approximation for k-PATH TRANSVERSAL for a general k and a universal constant $\delta > 0$. We show that it admits $O(\log k)$-approximation in FPT time. Note that the superpolynomial dependence on k is necessary for any approximation from NP-hardness of finding a k-path.

Theorem 12.2.1. *There is an* $O(\log k)$*-approximation algorithm for* k-PATH TRANSVERSAL *that runs in time* $2^{O(k^3 \log k)} n^{O(1)}$.

Table 12.1 summarizes our results.

Promises	H-TRANSVERSAL	H-PACKING
2-connected	Hard to approximate within $k-1$	Hard to approximate within $\Omega(\frac{k}{\text{polylog}(k)})$
k-Star	Admits $O(\log k)$-approximation	Hard to approximate within $\Omega(\frac{k}{\text{polylog}(k)})$
k-Path	Admits $O(\log k)$-approximation	?

Table 12.1: Summary of our algorithmic and hardness results for H-TRANSVERSAL and H-PACKING for different H.

12.2.2 Graph Partitioning

Our main result for graph partitioning is the following algorithm for k-VERTEX SEPARATOR. For fixed constants $b, c > 1$, an algorithm for k-VERTEX SEPARATOR is called an (b, c)-bicriteria approximation algorithm if given an instance $G = (V, E)$ and $k \in \mathbb{N}$, it outputs $S \subseteq V$ such that (1) each connected component of $G|_{S \setminus V}$ has at most bk vertices and (2) $|S|$ is at most c times the optimum of k-VERTEX SEPARATOR.

Theorem 12.2.2. *For any $\epsilon \in (0, 1/2)$, there is a polynomial time $(\frac{1}{1-2\epsilon}, O(\frac{\log k}{\epsilon}))$-bicriteria approximation algorithm for k-*VERTEX SEPARATOR.

Setting $\epsilon = \frac{1}{4}$ and running the algorithm yields $S \subseteq V$ with $|S| \leq O(\log k) \cdot \mathsf{Opt}$ such that each component in $G|_{V \setminus S}$ has at most $2k$ vertices. Performing an exhaustive search in each connected component yields the following true approximation algorithm whose running time depends exponentially only on k.

Corollary 12.2.3. *There is an $O(\log k)$-approximation algorithm for k-*VERTEX SEPARATOR *that runs in time $n^{O(1)} + 2^{O(k)}n$.*

This gives an FPT approximation algorithm when parameterized by only k, and its approximation ratio $O(\log k)$ improves the simple $(k+1)$-approximation from k-HYPERGRAPH VERTEX COVER. When $\mathsf{Opt} \gg k$, it runs even faster than the time lower bound $k^{\Omega(\mathsf{Opt})}n^{\Omega(1)}$ for the exact algorithm assuming the Exponential Time Hypothesis [DDvH14].

The natural question is whether superpolynomial dependence on k is necessary to achieve *true* $O(\log k)$-approximation. The following theorem proves hardness of k-VERTEX SEPARATOR based on Densest k-Subgraph. In particular, a polynomial time $O(\log k)$-approximation algorithm for k-VERTEX SEPARATOR will imply $O(\log^2 n)$-approximation algorithm for Densest k-Subgraph. Given that the best approximation algorithm achieves $\approx O(n^{1/4})$-approximation [BCC+10] and $n^{\Omega(1)}$-rounds of the Sum-of-Squares hierarchy have a gap at least $n^{\Omega(1)}$ [BCV+12], such a result seems unlikely or will be considered as a breakthrough.

Theorem 12.2.4. *If there is a polynomial time f-approximation algorithm for k-*VERTEX SEPARATOR*, then there is a polynomial time $2f^2$-approximation algorithm for Densest k-Subgraph.*

k-EDGE SEPARATOR is the edge deletion version of k-VERTEX SEPARATOR where given a graph $G = (V, E)$, we want to remove the fewest number of edges to make each connected component have strictly less than k vertices. For k-EDGE SEPARATOR, we prove that the true $O(\log k)$-approximation can be achieved in polynomial time. This shows a stark difference between the vertex version and the edge version.

Theorem 12.2.5. *There is an $O(\log k)$-approximation algorithm for k-*EDGE SEPARATOR *that runs in time $n^{O(1)}$.*

When $k = n^{o(1)}$ so that $\rho = n^{-(1-o(1))}$, our algorithm outperforms the previous best approximation algorithm for ρ-separator [KNS09, ENRS99].[2]

While most of graph partitioning algorithms deal with the edge version, we focus on the vertex version because (1) it exhibits richer connections to k-HYPERGRAPH VERTEX COVERand FPT as mentioned, (2) usually the vertex version is considered to be harder in the graph partitioning literature. We present the algorithm for the edge version in Appendix 15.5.

12.3 Related Work and Special Cases

After the aforementioned work characterizing those pattern graphs H admitting the existence of a polynomial-time exact algorithm for H-PACKING [KH83, DT97], Lund and Yannakakis [LY93] studied the maximization version of H-TRANSVERSAL (i.e. find the largest $V' \subseteq V_G$ such that the subgraph induced by V' does not have H as a subgraph), and showed it is hard to approximate within factor $2^{\log^{1/2-\epsilon} n}$ for any $\epsilon > 0$. They also mentioned the minimization version of two extensions of H-TRANSVERSAL. The most general node-deletion problem is APX-hard for every nontrivial hereditary (i.e. closed under node deletion) property, and the special case where the property is characterized by a finite number of forbidden subgraphs (i.e. $\{H_1, ..., H_l\}$-TRANSVERSAL in our terminology) can be approximated with a constant ratio. They did not provide explicit constants (one trivial approximation ratio for $\{H_1, ..., H_l\}$-TRANSVERSAL is $\max(|V_{H_1}|, ..., |V_{H_l}|)$),

[2]Both papers only present a bicriteria approximation algorithm, but they can be combined with our final *cleanup* step to achieve true approximation by *adding* $O(\log k)$ to the approximation ratio. See Appendix 15.5.

and our result can be viewed as a quantitative extension of their inapproximability results for the special case of H-TRANSVERSAL.

H-TRANSVERSAL / H-PACKING has been also studied outside the approximation algorithms community. The duality between our H-TRANSVERSAL and H-PACKING is closely related to the famous Erdős-Pósa property actively studied in combinatorics. The recent work of Jansen and Marx [JM15] considered problems similar to our H-PACKING with respect to fixed-parameter tractability (FPT).

Many other works on H-TRANSVERSAL / H-PACKING focus on a special case where H is a cycle or clique. We define k-Cycle (resp. k-Clique) to be the cycle (resp. clique) on k vertices.

Cycles. The initial motivation for our work was to prove a super-constant factor inapproximability for the FEEDBACK VERTEX SET (FVS) problem without relying on the Unique Games Conjecture. Given a (directed) graph G, the FVS problem asks to find a subset F of vertices with the minimum cardinality that intersects every cycle in the graph (equivalently, the induced subgraph $G \setminus F$ is acyclic). One of Karp's 21 NP-complete problems, FVS has been a subject of active research for many years in terms of approximation algorithms and fixed-parameter tractability (FPT). For FPT results, see [Bod94, CLL$^+$08, CPPW11, CCHM12] and references therein.

FVS on undirected graphs has a 2-approximation algorithm [BBF95, BG96, CGHW98], but the same problem is not well-understood in directed graphs. The best approximation algorithm [Sey95, ENSS98, ENRS00] achieves an approximation factor of $O(\log n \log \log n)$. The best hardness result follows from a simple approximation preserving reduction from VERTEX COVER, which implies that it is NP-hard to approximate FEEDBACK VERTEX SET within a factor of 1.36 [DS05]. Assuming UGC [Kho02b], it is NP-hard to approximate FVS in directed graphs within any constant factor [GMR08, Sve13] (we give a simpler proof in [GL16c]). The main challenge is to *bypass* the UGC and to show a super-constant inapproximability result for FVS assuming only $\mathsf{P} \neq \mathsf{NP}$ or $\mathsf{NP} \not\subseteq \mathsf{BPP}$.

By Theorem 12.2.1, we prove that k-CYCLE TRANSVERSAL is hard to approximate within factor $\Omega(k)$. The following theorem improves the result of Theorem 12.2.1 in the sense that in the completeness case, a small number of vertices not only intersect cycles of length exactly k, but intersect every cycle of length $3, 4, ..., O(\frac{\log n}{\log \log n})$.

Theorem 12.3.1. *Fix an integer $k \geq 3$ and $\epsilon \in (0, 1)$. Given a graph $G = (V_G, E_G)$ (directed or undirected), unless $\mathsf{NP} \subseteq \mathsf{BPP}$, there is no polynomial time algorithm to tell apart the following two cases.*

- *Completeness: There exists $F \subseteq V_G$ with $\frac{1}{k-1} + \epsilon$ fraction of vertices that intersects every cycle of at most length $O(\frac{\log n}{\log \log n})$ (hidden constant in O depends on k and ϵ).*
- *Soundness: Every subset F with less than $1-\epsilon$ fraction of vertices does not intersect at least one cycle of length k. Equivalently, any subset with more than ϵ fraction of vertices has a cycle of length exactly k in the induced subgraph.*

This can be viewed as some (modest) progress towards showing inapproximability of FVS in the following sense. Consider the following standard linear programming (LP) relaxation for FVS.

$$\min \sum_{v \in V_G} x_v \quad \text{subject to} \quad \sum_{v \in C} x_v \geq 1 \quad \forall \text{ cycle } C \text{ ,} \quad \text{and} \quad 0 \leq x_v \leq 1 \;\; \forall v \in V_G$$

The integrality gap of the above LP is upper bounded by $O(\log n)$ for undirected graphs [BYGNR98] and $O(\log n \log \log n)$ for directed graphs [ENSS98]. Suppose in the completeness case, there exist c fraction of vertices that intersect every cycle of length at most $\log^{1.1} n$ (or any number bigger than the known integrality gaps). If we remove these vertices and consider the above LP on the remaining subgraphs, since every cycle is of length at least $\log^{1.1} n$, setting $x_v = 1/\log^{1.1} n$ is a feasible solution, implying that the optimal solution to the LP is at most $n/\log^{1.1} n$. Since the integrality gap is at most $O(\log n \log \log n)$, we can conclude that the remaining cycles can be hit by at most $O(n \log \log n / \log^{0.1} n) = o(n)$ vertices, extending the completeness result to every cycle. Thus, improving our result to hit cycles of length $\omega(\log n \log \log n)$ in the completeness case will prove a factor-$\omega(1)$ inapproximability of FVS.

Another interesting aspect about Theorem 12.3.1 is that it also holds for undirected graphs. This should be contrasted with the fact that undirected graphs admit a 2-approximation algorithm for FVS, suggesting that to overcome $\log n$-cycle barrier mentioned above, some properties of directed graphs must be exploited. Section B.2 and B.3 of the arXiv version of this work [GL15b] present a directed graph specific approach using a different reduction technique called *labeling gadget* to prove a similar result only on directed graphs. It has an additional advantage of being derandomized and assumes only $\mathsf{P} \neq \mathsf{NP}$.

For cycles of bounded length, Kortsarz et al. [KLN10] studied k-CYCLE EDGE TRANSVERSAL, and suggested a $(k-1)$-approximation algorithm as well as proved that improving the ratio 2 for K_3 will have the same impact on VERTEX COVER, refuting the Unique Games Conjecture [KR08].

For the dual problem of packing cycles of any length, called VERTEX-DISJOINT CYCLE PACKING (VDCP), the results of [KNS+07, FS07] imply that the best approximation factor by any polynomial time algorithm lies between $\Omega(\sqrt{\log n})$ and $O(\log n)$. In a closely related problem EDGE-DISJOINT CYCLE PACKING (EDCP), the same papers showed that $\Theta(\log n)$ is the best possible. In directed graphs the vertex and edge version have the same approximability, the best known algorithms achieves $O(\sqrt{n})$-approximation while the best hardness result remains $\Omega(\log n)$.

Variants of k-CYCLE PACKING have also been considered in the literature. Rautenbach and Regen [RR09] studied k-CYCLE EDGE PACKING on graphs with girth k and small degree. Chalermsook et al. [CLN14] studied a variant of k-CYCLE PACKING on directed graphs for $k \geq n^{1/2}$ where we want to pack as many disjoint cycles of length at most k as possible, and proved that it is NP-hard to approximate within a factor of $n^{1/2-\epsilon}$. This matches the algorithm implied by [KNS$^+$07].

Cliques. MINIMUM MAXIMAL (RESP. MAXIMUM) CLIQUE TRANSVERSAL asks to find the smallest subset of vertices that intersects every maximal (resp. maximum) clique in the graph. In mathematics, Tuza [Tuz91] and Erdős et al. [EGT92] started to estimate the size of the smallest such set depending on structure of graphs. See the recent work of Shan et al. [SLK14] and references therein. In computer science, exactly computing the smallest set on special classes of graphs appears in many works [GPR00, LCS02, CKL01, DLMS08, Lee12].

Both the edge and vertex version of k-CLIQUE PACKING also have been studied actively both in mathematics and computer science. In mathematics, the main focus of research is lower bounding the maximum number of edge or vertex-disjoint copies of K_k in very dense graphs (note that even K_3 does not exist in $K_{n,n}$ which has $2n$ vertices and n^2 edges). See the recent paper [Yus14] or the survey [Yus07] of Yuster. The latter survey also mentions approximation algorithms, including APX-hardness and the general approximation algorithm for k-SET PACKING which now achieves $\frac{k+1+\epsilon}{3}$ for the vertex version and $\frac{\binom{k}{2}+1+\epsilon}{3}$ for the edge version [Cyg13]. Feder and Subi [FS12] considered H-EDGE PACKING and showed APX-hardness when H is k-cycle or k-clique. Chataigner et al. [CMWY09] considered an interesting variant where we want to pack vertex-disjoint cliques of any size to maximize the total number of edges of the packed cliques, and proved APX-hardness and a 2-approximation algorithm. Exact algorithms for special classes of graphs have been considered in [BCD97, GPRC$^+$01, HKNP05, Klo12].

Graph Partitioning. Graph partitioning is a general task of removing a small number of edges or vertices to make the resulting graph consist of smaller connected components. In this context, the edge versions have received more attention.

One of the most well-studied formulations is called l-*Balanced Partitioning*. Given a graph $G = (V, E)$ and $l \in \mathbb{N}$, the goal is to remove the smallest number of edges so that the resulting graph has l ($l \geq 2$) connected components with (roughly) the same number $\frac{n}{l}$ of vertices. [3] The case $l = 2$ has been studied extensively and produced elegant approximation algorithms. The best results are $O(\log n)$-true approximation (i.e., each component must have $\frac{n}{2}$ vertices) [Rac08] and $O(\sqrt{\log n})$-bicriteria approximation (i.e., each component must have at most $\frac{2n}{3}$ vertices) [ARV09]. The extension to $l \geq 3$ has been studied more recently. While it is NP-hard to achieve any nontrivial true approximation for general l [AR06], Krauthgamer et al. [KNS09] presented an $O(\sqrt{\log n \log l})$-bicriteria approximation where the resulting graph is guaranteed to have each connected component with at most $\frac{2n}{l}$ vertices.

The true approximation for l-Balanced Partitioning is ruled out by encoding the Integer 3-Partition problem in graphs, and hard instances contain disjoint cliques of size at most $\frac{n}{l}$. Even et al. [ENRS99] defined a similar problem called *ρ-Separator*, which is exactly our k-EDGE SEPARATOR with $\rho = \frac{k}{n}$. They "believe that the definition of ρ-Separator captures type of partitioning that is actually required in applications", since "instead of limiting the number of resulting parts, which is not always important for divide-and-conquer applications or for parallelism, it limits only the sizes or weights of each part." They provided a bicriteria approximation algorithm that removes at most $O(\frac{1+\epsilon}{\epsilon} \log n) \cdot \mathsf{Opt}$ edges to make sure that each component has size $(1 + \epsilon)\rho n$ for any $\epsilon > 0$, which is improved to $O(\frac{1+\epsilon}{\epsilon}\sqrt{\log(1/\epsilon\rho) \log n}) \cdot \mathsf{Opt}$ by Krauthgamer et al. [KNS09]. The previous algorithms' primary focus is when ρ is a constant (so that $k = \Omega(n)$), and their performance deteriorates when k is small. In particular, when $k = O(n^{1-\epsilon})$ and $\rho = O(n^{-\epsilon})$ for some $\epsilon > 0$, the best guarantee from the above line of work gives an $O(\log n)$-bicriteria approximation algorithm.

Some of the ideas can be used for the analogous vertex versions, but they have not received the same amount of attention. Often additional algorithmic ideas were required to achieve the same guarantee [FHL08], or matching the same guarantee is proved to be NP-hard under some complexity assumptions [LRV13].

Fixed Parameter Tractability. Given a graph G and an integer k, the optimum of k-VERTEX SEPARATOR has been known as *k-Component Order Connectivity* in mathematics. We refer to the survey by Gross et al. [GHI$^+$13] for more background.

Let Opt be the optimal value. For small values of k and Opt, the complexity of exact algorithms has been studied in terms of their fixed parameter tractability (FPT). While the

[3] In the literature it is called k-Balanced Partitioning. We use l in order to avoid confusion between l-Balanced Partitioning and k-Edge Separator ($l = \frac{n}{k}$).

trivial algorithm takes $n^{O(\mathsf{Opt})}$ time to find the exact solution for k-VERTEX SEPARATOR, Drange et al. [DDvH14] presented an exact algorithm that runs in time $k^{O(\mathsf{Opt})}n$, so the problem is in FPT when parameterized by both k and Opt. They complemented their result by showing that the problem is $\mathbf{W}[1]$-hard when parameterized by Opt or k. They also showed that any exact algorithm that runs in time $k^{o(\mathsf{Opt})}n^{O(1)}$ will refute the Exponential Time Hypothesis.

Given an instance of a problem with a parameter κ, an approximation algorithm is said to be an FPT c-approximation algorithm if it runs in time $f(\kappa) \cdot n^{O(1)}$ for some function f and achieves c-approximation. See the survey of Marx [Mar08] and the recent work of Chitnis et al. [CHK13]. For k-VERTEX SEPARATOR, the simple $(k+1)$-approximation runs in polynomial time regardless of Opt and k, but any exact algorithm requires both Opt and k to be parameterized. It is an interesting question whether significantly improved approximation is possible when only one of them is parameterized.

12.3.1 Organization

Chapter 13 presents our hardness results. Section 13.1 recalls and extends previous hardness results for the problems we reduce from. Section 13.2 and Section 13.3 prove hardness of H-TRANSVERSAL and H-PACKING respectively. Section 13.4 proves Theorem 12.3.1 to illustrate the connection to FVS.

Chapter 14 gives an $O(\log k)$-approximation algorithm for k-STAR TRANSVERSAL, proving Theorem 12.2.3. Chapter 15 gives an $O(\log k)$-approximation algorithm for k-PATH TRANSVERSAL, proving Theorem 12.2.1.

Chapter 13

Hardness of H-Transversal / Packing

13.1 Preliminaries

Notation. A k-uniform hypergraph is denoted by $P = (V_P, E_P)$ such that each $e \in E_P$ is a k-subset of V_P. We denote e as an *ordered* k-tuple $e = (v^1, \ldots, v^k)$. The ordering can be chosen arbitrarily given P, but should be fixed throughout. If v indicates a vertex of some graph, we use a superscript v^i to denote another vertex of the same graph, and e^i to denote the ith (hyper)edge. For an integer m, let $[m] = \{1, 2, \ldots, m\}$. Unless otherwise stated, the *measure* of $F \subseteq V$ is obtained under the uniform measure on V, which is simply $\frac{|F|}{|V|}$.

k-HYPERGRAPH VERTEX COVER. An instance of k-HYPERGRAPH VERTEX COVER consists of a k-uniform hypergraph P, where the goal is to find a set $C \subseteq V_P$ with the minimum cardinality such that it intersects every hyperedge. The result of Dinur, Guruswami, Khot and Regev [DGKR05] states that

Theorem 13.1.1 ([DGKR05]). *Given a k-uniform hypergraph ($k \geq 3$) and $\epsilon > 0$, it is NP-hard to tell apart the following cases:*

- *Completeness: There exists a vertex cover of measure $\frac{1+\epsilon}{k-1}$.*
- *Soundness: Every vertex cover has measure at least $1 - \epsilon$.*

*Therefore, it is NP-hard to approximate k-*HYPERGRAPH VERTEX COVER *within a factor $k - 1 + 2\epsilon$.*

MIS-k. Given a graph $G = (V_G, E_G)$, a subset $S \subseteq V_G$ is *independent* if the subgraph induced by S does not contain any edge. The MAXIMUM INDEPENDENT SET (MIS) problem asks to find the largest independent set, and MIS-k indicates the same problem where G is promised to have maximum degree at most k. The recent result of Chan [Cha13] implies

Theorem 13.1.2 ([Cha13])**.** *Given a graph G with maximum degree at most k, it is NP-hard to tell apart the following cases:*

- *Completeness: There exists an independent set of measure $\Omega(1/(\log k))$.*
- *Soundness: Every subset of vertices of measure $O(\frac{\log^3 k}{k})$ contains an edge.*

Therefore, it is NP-hard to approximate MIS-k *within a factor* $\Omega(\frac{k}{\log^4 k})$.

13.2 H-TRANSVERSAL

In this section, given a 2-connected graph $H = (V_H, E_H)$ with k vertices, we give a reduction from k-HYPERGRAPH VERTEX COVER to H-TRANSVERSAL. The simplest try will be, given a hypergraph $P = (V_P, E_P)$ (let $n = |V_P|, m = |E_P|$), to produce a graph $G = (V_G, E_G)$ where $V_G = V_P$, and for each hyperedge $e = (v^1, \dots, v^k)$ add $|E_H|$ edges that form a *canonical copy* of H to E_G. While the soundness follows directly (if $F \subseteq V_P$ contains a hyperedge, the subgraph induced by F contains H), the completeness property does not hold since edges that belong to different canonical copies may form an unintended non-canonical copy. To prevent this, a natural strategy is to replace each vertex by a set of many vertices (call it a *cloud*), and for each hyperedge $(v^1, \dots, v^k)$, add many canonical copies on the k clouds (each copy consists of one vertex from each cloud). If we have too many canonical copies, soundness works easily but completeness is hard to show due to the risk posed by non-canonical copies, and in the other extreme, having too few canonical copies could result in the violation of the soundness property. Therefore, it is important to control the structure (number) of canonical copies that ensure both completeness and soundness at the same time.

Our technique, which we call *random matching*, proceeds by creating a carefully chosen number of *random* copies of H for each hyperedge to ensure both completeness and soundness. We remark that properties of random matchings are also used to bound the number of short non-canonical paths in inapproximability results for edge-disjoint paths on undirected graphs [AZ06, ACG$^+$10]. The details in our case are different as we create *many* copies of H based on a *hypergraph*.

Fix $\epsilon > 0$, apply Theorem 13.1.1, let $c := \frac{1+\epsilon}{k-1}$, $s := 1 - \epsilon$ be the measure of the minimum vertex cover in the completeness and soundness case respectively, and $d := d(k, \epsilon)$ be the maximum degree of hard instances. Let a and B be integer constants greater than 1, which will be determined later. Lemma 13.2.1 and Lemma 13.2.4 with these parameters imply the first half of Theorem 12.2.1.

Reduction. Without loss of generality, assume that $V_H = [k]$. Given a hypergraph $P = (V_P, E_P)$, construct an undirected graph $G = (V_G, E_G)$ such that

- $V_G = V_P \times [B]$. Let $n = |V_P|$ and $N = |V_G| = nB$. For $v \in V_P$, let $\mathsf{cloud}(v) := \{v\} \times [B]$ be the copy of $[B]$ associated with v.
- For each hyperedge $e = (v^1, \dots, v^k)$, for aB times, take $l^1, \dots, l^k$ independently and uniformly from $[B]$. For each edge $(i, j) \in H$ $(1 \leq i < j \leq k)$, add $((v^i, l^i), (v^j, l^j))$ to E_G. Each time we add $|E_H|$ edges isomorphic to H, and we have aB of such copies of H per each hyperedge. Call such copies *canonical*.

Completeness. The next lemma shows that if P has a small vertex cover, G also has a small H-TRANSVERSAL.

Lemma 13.2.1. *Suppose P has a vertex cover C of measure c. For any $\epsilon > 0$, with probability at least $3/4$, there exists a subset $F \subseteq V_G$ of measure at most $c + \epsilon$ such that the subgraph induced by $V_G \setminus F$ has no copy of H.*

Proof. Let $F = C \times [B]$. We consider the expected number of copies of H that avoid F and argue that a small fraction of additional vertices intersect all of these copies. Choose k vertices $(v^1, l^1), \dots, (v^k, l^k)$ which satisfy

- $v^1 \in V_P$ can be any vertex.
- $l^1, \dots, l^k \in B$ can be arbitrary labels.
- For each $(i, j) \in E_H$, there must be a hyperedge of P containing both i and j.

There are n possible choices for v^1, B choices for each l^i, and at most kd choices for each v^i $(i > 1)$. The number of possibilities to choose such $(v^1, l^1), \dots, (v^k, l^k)$ is bounded by $n(dk)^k B^k$. Note that no other k-tuple of vertices induce a connected graph and contain a copy of H. Further discard the tuple when two vertices are the same.

We calculate the probability that the subgraph induced by $((v^1, l^1), \dots, (v^k, l^k))$ contains a copy *in this order* — formally, for all $(i, j) \in E_H$, $((v^i, l^i), (v^j, l^j)) \in E_G$. For

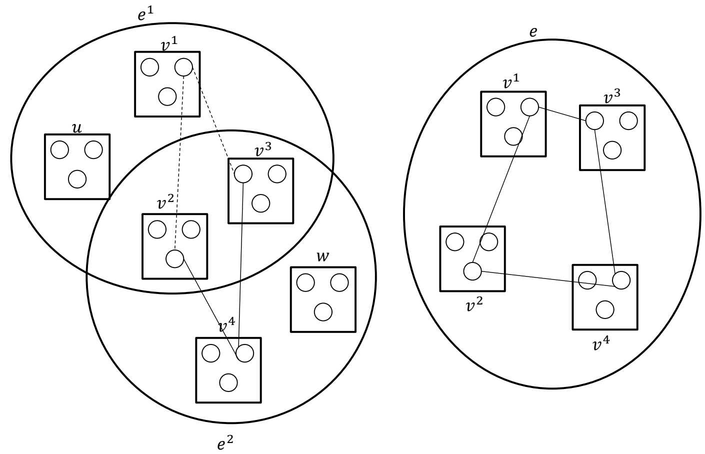

Figure 13.1: Two examples where $k = 4$ and H is a 4-cycle. On the left, purported edges are divided into two groups (dashed and solid edges). Each copy of canonical cycle should match the labels of three vertices to ensure it covers 2 designated edges (6 labels total). On the right, one canonical copy can cover all the edges, and it only needs to match the labels of four vertices (4 labels total).

each $(i, j) \in E_H$, we call a pair $((v^i, l^i), (v^j, l^j)) \in \binom{V_G}{2}$ a *purported edge*. For a set of purported edges, we say that this set can be *covered by a single canonical copy* if one copy of canonical copy of H can contain all purported edges with nonzero probability. Suppose that all $|E_H|$ purported edges can be covered by a single canonical copy of H. It is only possible when there is a hyperedge whose k vertices are exactly $\{v^1, \ldots, v^k\}$. In this case, $((v^1, l^1), \ldots, (v^k, l^k))$ intersects F. (right case of Figure 13.1). When $|E_H|$ purported edges have to be covered by more than one canonical copy, some vertices must be covered by more than one canonical copy, and each canonical copy covering the same vertex should give the same label to that vertex. This redundancy makes it unlikely to have all k edges exist at the same time. (left case of Figure 13.1). The below claim formalizes this intuition.

Claim 13.2.2. *Suppose that $((v^1, l^1), \ldots, (v^k, l^k))$ cannot be covered by a single canonical copy. Then the probability that it forms a copy of H is at most $\frac{(adk)^{k^2}}{B^k}$.*

Proof. Fix $2 \leq p \leq |E_H|$. Partition $|E_H|$ purported edges into p nonempty *groups* $I_1, \ldots, I_p$ such that each group can be covered by a single canonical copy of H. There are at most $p^{|E_H|}$ possibilities to partition. For each $v \in V_P$, there are at most d hyperedges containing v and at most aBd canonical copies intersecting $\mathsf{cloud}(v)$. Therefore, all edges in one group can be covered simultaneously by at most aBd canonical copies. There are at most $(aBd)^p$ possibilities to assign a canonical copy to each group. Assume that one canonical copy is responsible for exactly one group. This is without loss of generality since if one canonical copy is responsible for many groups, we can merge them and this case can be dealt with smaller p.

Focus on one group I of purported edges, and one canonical copy $L = (V_L, E_L)$ which is supposed to cover them. Let $I' \subseteq V_G$ be the set of vertices which are incident on the edges in I. Suppose $V_L = \{(u^1, l'^1), \ldots, (u^k, l'^k)\}$, which is created by a hyperedge $f = (u^1, \ldots, u^k) \in E_P$. We calculate the probability that L contains all edges in I over the choice of labels $l'^1, \ldots, l'^k$ for L. One necessary condition is that $\{v | (v, l) \in I' \text{ for some } l \in [B]\}$ (i.e. the set I' projected to V_P) is contained in f. Otherwise, some vertices of I' cannot be covered by L. Another necessary condition is $v^i \neq v^j$ for any $(v^i, l^i) \neq (v^j, l^j) \in I'$. Otherwise (i.e. $(v, l^i), (v, l^j) \in I'$ for $l^i \neq l^j$), since L gives only one label to each vertex in $f \subseteq V_P$, (v, l^i) and (v, l^j) cannot be contained in L simultaneously. Therefore, we have a nice characterization of I': It consists of at most one vertex from the cloud of each vertex in f.

The probability that L contains I is at most the probability that for each $(v^i, l^i) \in I'$, l^i is equal to the label L assigns to v^i, which is $B^{-|I'|}$. Now we need the following lemma saying that the sum of $|I'|$ is large, which relies on 2-connectivity of H.

Lemma 13.2.3. *Fix $p \geq 2$. For any partition $I_1, ..., I_p$ of purported edges into p non-empty groups, $\sum_{i=1}^{p} |I'_i| \geq k + p$.*

Proof. Let t be the number of vertices contained in at least two I'_is. Call them *boundary vertices*. Note that exactly $k - t$ vertices belongs to exactly one I'_i. For $i = 1, ..., p$, let b_i be the number of boundary vertices in $|I'_i|$. Since (I'_i, I_i) is a proper subgraph of H and H is 2-vertex connected, $b_i \geq 2$ for each i, which implies $\sum_i b_i \geq 2p$. Furthermore, each boundary vertex contributes to at least two b_i's, so $\sum_i b_i \geq 2t$. Therefore, $\sum_i b_i \geq \max(2p, 2t)$ and

$$\sum_{i=1}^{p} |I'_i| = (k - t) + \sum_{i=1}^{p} b_i \geq (k - t) + \max(2p, 2t) \geq k + p.$$

□

We conclude that for each partition, the probability of having all the edges is at most

$$(aBd)^p \prod_{q=1}^{p} B^{-|I'_q|} = \frac{(aBd)^p}{B^{k+p}} = \frac{(ad)^p}{B^k} .$$

The probability that $((v^1, l^1), \ldots, (v^k, l^k))$ forms a copy is therefore bounded by

$$\sum_{p=2}^{|E_H|} p^{|E_H|} \frac{(ad)^p}{B^k} \leq \frac{(adk)^{k^2}}{B^k} .$$

$\square$

Therefore, the expected number of copies that avoid F is bounded by $n(kd)^k B^k \cdot \frac{(adk)^{k^2}}{B^k}$. With probability at least $3/4$, the number of such copies is at most $4n(adk)^{2k^2}$. Let $B \geq \frac{4(adk)^{2k^2}}{\epsilon}$. Then these copies of H can be covered by at most $\epsilon n B = \epsilon N$ vertices. $\square$

Soundness. The soundness claim above is easier to establish. By an averaging argument, a subset I of V_G of measure 2ϵ must contain ϵB vertices from the clouds corresponding to a subset S of measure ϵ in V_P. There must be a hyperedge e contained within S, and the chosen parameters ensure that one of the canonical copies corresponding to e is likely to lie within I.

Lemma 13.2.4. *For $a = a(k, \epsilon)$ and $B = \Omega(\log |E_P|)$, if every subset of V_P of measure at least ϵ contains a hyperedge in the induced subgraph, with probability at least $3/4$, every subset of V_G with measure 2ϵ contains a canonical copy of H.*

Proof. We want to show that the following property holds for every hyperedge $e = (v^1, \ldots, v^k)$: if a subset of vertices $I \subseteq V_G$ has at least ϵ fraction of vertices from each $\mathsf{cloud}(v^i)$, then I will contain a canonical copy. Fix $A^1 \subseteq \mathsf{cloud}(v^1), \ldots, A^k \subseteq \mathsf{cloud}(v^k)$ be such that for each i, $|A^i| \geq \epsilon B$. There are at most 2^{kB} ways to choose such A's. The probability that one canonical copy associated with e is not contained in $(v^1, A^1) \times \cdots \times (v^k, A^k)$ is at most $1 - \epsilon^k$. The probability that none of canonical copy associated with e is contained in $(v^1, A^1) \times \cdots \times (v^k, A^k)$ is $(1 - \epsilon^k)^{aB} \leq \exp(-aB\epsilon^k)$.

By union bound over all $A^1, \ldots, A^k$, the probability that there exists $A^1, \ldots, A^k$ containing no canonical copy is at most $\exp(kB - aB\epsilon^k) = \exp(-B) \leq \frac{1}{4|E_P|}$ by taking a large enough constant depending on k and ϵ, and $B = \Omega(\log |E_P|)$. Therefore, with probability at least $3/4$, the desired property holds for all hyperedges.

Let I be a subset of V_G of measure at least 2ϵ. By an averaging argument, at least ϵ fraction of *good* vertices $v \in V_P$ satisfy that $|\mathsf{cloud}(v^i) \cap I| \geq \epsilon B$. By the soundness property of P, there is a hyperedge e contained in the subgraph induced by the good vertices, and the above property for e ensures that I contains a canonical copy. $\square$

13.3 H-PACKING and MIS-k-g

Given a 2-connected graph H, the reduction from MIS-k-k to H-PACKING is relatively straightforward. Here we assume that hard instances of MIS-k-k are indeed k-regular for simplicity. Given an instance $M = (V_M, E_M)$ of MIS-k-k, we take $G = (V_G, E_G)$ to be its *line graph* — $V_G = E_M$, and $e, f \in V_G$ are adjacent if and only if they share an endpoint as edges of M.

For each vertex $v \in V_M$, let $\mathsf{star}(v) := \{e \in V_G : v \in e\}$. $\mathsf{star}(v)$ induces a k-clique, and for $v, u \in V_M$, $\mathsf{star}(v)$ and $\mathsf{star}(u)$ share one vertex if u and v are adjacent, and share no vertex otherwise. Given an independent set S of M, we can find $|S|$ pairwise disjoint stars in G, which gives $|S|$ vertex-disjoint copies of H. On the other hand, 2-connectivity of H and large girth of M implies that any copy of H must be entirely contained in one star, proving that many disjoint copies of H in G also give a large independent set of M with the same cardinality, completing the reduction from MIS-k-k to H-PACKING. The following theorem formalizes the above intuition.

Lemma 13.3.1. *For a 2-connected graph H with k vertices, there is an approximation-preserving reduction from* MIS-k-k *to* H-PACKING.

Proof. Let $M = (V_M, E_M)$ be an instance of MIS-k-k M with maximum degree k and girth greater than k. First, let $G = (V_G = E_M, E_G)$ be the line graph of M. For each vertex $v \in V_M$ with degree strictly less than k, we add $k - \mathsf{deg}(v)$ new vertices to V_G. Let $\mathsf{star}(v) \subseteq V_G$ be the union of the edges of M incident on v and the newly added vertices for v. Note that $|\mathsf{star}(v)| = k$ for all $v \in V_M$. Add edges to G to ensure that every $\mathsf{star}(v)$ induces a k-clique. For two vertices u and v of M, $\mathsf{star}(u)$ and $\mathsf{star}(v)$ share exactly one vertex if u and v are adjacent in M, and share no vertex otherwise.

Let S be an independent set of M. The $|S|$ stars $\{\mathsf{star}(v)\}_{v \in S}$ are pairwise disjoint and each induces a k-clique, so G contains at least $|S|$ disjoint copies of H.

We claim that any k-subset of V_G that induces a 2-connected subgraph must be $\mathsf{star}(v)$ for some v. Assume towards contradiction, let T be a k-subset inducing a 2-connected subgraph of G that cannot be contained in a single star. We first show T must contain two disjoint edges of M. Take any $(u, v) \in T$. Since $T \not\subseteq \mathsf{star}(u)$, T contains an edge of M not incident on u. If it is not incident on v either, we are done. Otherwise, let (w, v) be this edge. The same argument from $T \not\subseteq \mathsf{star}(v)$ gives another edge (w', u) in T. If $w \neq w'$, (w, v) and (w', u) are disjoint. Otherwise, w, u, v form a triangle in M, contradicting a large girth. Let (u, v), (w, x) be two disjoint edges of M in contained in T.

Since the subgraph of G vertex-induced by T is 2-connected, there are two internally vertex-disjoint paths P_1, P_2 in G from (u, v) to (w, x). The sum of the two lengths is at most k, where the length of a path is defined to be the number of edges. By considering the internal vertices of P_i (edges of M) and deleting unnecessary portions, we have two edge-disjoint paths P'_1, P'_2 in M where each P'_i connects $\{u, v\}$ and $\{w, x\}$, with length at most the length of P_i minus one. There is a cycle in M consists only of the edges of P'_1, P'_2 together with $(u, v), (w, x)$. Since $|P'_1| + |P'_2| + 2 \leq k$, it contradicts that M has girth strictly greater than k. □

We prove that MIS-k-g is also hard to approximate by a reduction from MIS-d ($d = \widetilde{\Omega}(k)$), using a slightly different *random matching* idea. Given a degree-d graph with possibly small girth, we replace each vertex by a cloud of B vertices, and replace each edge by a copies of random matching between the two clouds. While maintaining the soundness guarantee, we show that there are only a few small cycles, and by deleting a vertex from each of them and *sparsifying* the graph we obtain a hard instance for MIS-k-g. Note that g does not affect the inapproximability factor but only the runtime of the reduction.

Theorem 13.3.1 (Restatement of Theorem 12.2.2)**.** *For any constants k and g, unless* $\mathsf{NP} \subseteq \mathsf{BPP}$*, no polynomial time algorithm approximates* MIS-k-g *within a factor of* $\Omega(\frac{k}{\log^7 k})$.

Proof. We reduce from MIS-d to MIS-k-g where $k = O(d \log^2 d)$. Given an instance $G_0 = (V_{G_0}, E_{G_0})$ of MIS-d, we construct $G = (V_G, E_G)$ and $G' = (V_{G'}, E_{G'})$ by the following procedure:

- $V_G = V_{G_0} \times [B]$. As usual, let $\mathsf{cloud}(v) = \{v\} \times [B]$.
- For each edge $(u, v) \in E_{G_0}$, for a times, add a *random matching* as follows.
  - Take a random permutation $\pi : [B] \to [B]$.
  - Add an edge $((u, i), (v, \pi(i))$ for all $i \in [B]$.
- Call the resulting graph G. To get the final graph G',
  - For any cycle of length at most g, delete an arbitrary vertex from the cycle. Repeat until there is no cycle of length at most g.

Note that the step of eliminating the small cycles can be implemented trivially in time $O(n^g)$. Let $n = |V_{G_0}|, m = |E_{G_0}|, N = nB = |V_G| \geq |V_{G'}|, M = m \cdot aB = |E_G| \geq$

$|E_{G'}|$. The maximum degree of G and G' is at most ad. By construction, girth of G' is at least $g+1$.

Girth Control. We calculate the expected number of small cycles in G, and argue that the number of these cycles is much smaller than the total number of vertices, so that $|V_G|$ and $|V_{G'}|$ are almost the same. Fix $1 \le k' \le g$. We bound the number of cycles of length k'. Let each pair $((v^1, l^1), (v^2, l^2))$ of vertices a *purported edge* and each k'-tuple $((v^1, l^1), \dots, (v^{k'}, l^{k'}))$ a *purported cycle*. A purported cycle has nonzero probability of being a real cycle in G if

- $v^1 \in V_{G_0}$ can be any vertex.
- For each $1 \le i < k'$, $(v^i, v^{i+1}) \in E_{G_0}$.
- $l^1, \dots, l^{k'} \in B$ can be arbitrary labels.

There are n possible choices for v^1, B choices for each l^i, and d choices for each v^i $(i > 1)$. The number of possibilities to choose such $(v^1, l^1), \dots, (v^{k'}, l^{k'})$ is bounded by $nd^{k'-1}B^{k'}$. Without loss of generality, assume that no vertices appear more than once.

For each edge $e = (u, w) \in G_0$, consider the intersection of the purported cycle $((v^1, l^1), ..., (v^{k'}, l^{k'}))$ and the subgraph induced by $\mathsf{cloud}(u) \cup \mathsf{cloud}(w)$. It is a bipartite graph with the maximum degree 2. Suppose there are q purported edges $e^1, \dots, e^q$ (ordered arbitrarily) in this bipartite graph. By slightly abusing notation, let e^i also denote the event that e^i exists in G. The following claim upper bounds $\Pr[e^i | e^1, \dots, e^{i-1}]$ for each e^i.

Claim 13.3.2. $\Pr[e^i | e^1, \dots, e^{i-1}] \le \frac{a}{B-i}$.

Proof. There are a random matchings between $\mathsf{cloud}(u)$ and $\mathsf{cloud}(w)$, and for each $j < i$, there is at least one random matching including e^j. We calculate the probability e^i is contained by a random matching, conditioned on the fact that it already contains some of $e^1, \dots, e^{i-1}$.

If there is e^j $(j < i)$ that shares a vertex with e^i, e^i cannot be covered by the same random matching with e^j. If a random matching covers p of $e^1, \dots, e^{i-1}$ which are disjoint from e^i, the probability that e^i is covered by that random matching is $\frac{1}{B-p}$, and this is maximized when $p = i - 1$.

By a union bound over the a random matchings, $\Pr[e^i | e^1, \dots, e^{i-1}] \le \frac{a}{B-i}$. $\square$

The probability that all of $e^1, \ldots, e^q$ exist is at most

$$\prod_{i=1}^{q} \frac{a}{B-i} \leq \left(\frac{a}{B-q}\right)^q \leq \left(\frac{a}{B-k'}\right)^q .$$

Since edges of G_0 are processed independently, the probability of success for one fixed purported cycle is $(\frac{a}{B-k'})^{k'}$. The expected number of cycles of length k' is

$$\begin{aligned} & nd^{k'-1}B^{k'} \cdot \left(\frac{a}{B-k'}\right)^{k'} = nd^{k'-1}a^{k'}\left(1+\frac{k'}{B-k'}\right)^{k'} \\ \leq \quad & nd^{k'-1}a^{k'} \exp\left(\frac{k'^2}{B-k'}\right) \\ \leq \quad & en(ad)^{k'} \end{aligned}$$

by taking $B - k' \geq k'^2$. Summing over $k' = 1, \ldots, g$, the expected number of cycles of length up to g, is bounded by $eg(ad)^g n$. Take $B \geq 4d^2 \cdot eg(ad)^g$. Then with probability at least $3/4$, the number of cycles of length at most g is at most $\frac{Bn}{d^2}$. By taking $1/d^2$ fraction of vertices away (one for each short cycle), we have a girth at least $g+1$, which implies $\left(1 - \frac{1}{d^2}\right)|V_G| \leq |V_{G'}| \leq |V_G|$.

Hardness of MIS-d states that it is NP-hard to distinguish the case G_0 has an independent set of measure $c := \Omega(\frac{1}{\log d})$ and the case where the maximum independent set has measure at most $s := O(\frac{\log^3 d}{d})$.

Completeness. Let I_0 be an independent set of G_0 of measure c. Then $I = I_0 \times [B]$ is also an independent set of G of measure c. Let $I' = I \cap V_{G'}$. I' is independent in both G and G', and the measure of I' in G' is at least the measure of I' in G, which is at least $c - 1/d^2 = \Omega(\frac{1}{\log d})$.

Soundness. Suppose that every subset of V_{G_0} of measure at least s contains an edge. Say a graph is (β, α)-*dense* if we take β fraction of vertices, at least α fraction of edges lie within the induced subgraph. We also say a bipartite graph is (β, α)-*bipartite dense* if we take β fraction of vertices from each side, at least α fraction of edges lie within the induced subgraph.

Claim 13.3.3. *For $a = O(\frac{\log(1/s)}{s})$ and $B = O(\frac{\log m}{s})$ the following holds with probability at least $3/4$: For every $(u, w) \in E_{G_0}$, the bipartite graph between $\mathsf{cloud}(u)$ and $\mathsf{cloud}(w)$ is $(\epsilon, \epsilon^2/8)$-bipartite dense for all $\epsilon \geq s$.*

Proof. Fix (u, w), and $\epsilon \in [s, 1]$, and $X \subseteq \mathsf{cloud}(u)$ and $Y \subseteq \mathsf{cloud}(w)$ be such that $|X| = |Y| = \epsilon B$. The possibilities of choosing X and Y is

$$\binom{B}{\epsilon B}^2 \leq \exp(O(\epsilon \log(1/\epsilon) B))$$

Without loss of generality, let $X = \{u\} \times [\epsilon B]$ and $Y = \{v\} \times [\epsilon B]$. In one random matching, let X_i ($i \in [\epsilon B]$) be the random variable indicating whether vertex $(u, i) \in X$ is matched with a vertex in Y or not. $\Pr[X_1 = 1] = \epsilon$, and $\Pr[X_i = 1 | X_1, \ldots, X_{i-1}] \geq \epsilon/2$ for $i \in [\epsilon B/2]$ and any $X_1, \ldots, X_{i-1}$. Therefore, the expected number of edges between X and Y is at least $\epsilon^2 B/4$. With a random matchings, the expected number is at least $a\epsilon^2 B/4$. By Chernoff bound, the probability that it is less than $a\epsilon^2 B/8$ is at most $\exp(\frac{a\epsilon^2 B}{32})$. By union bound over all possibilities of choosing X and Y, the probability that the bipartite graph is not $(\epsilon, \epsilon^2/8)$-bipartite dense is

$$\exp(\epsilon \log(1/\epsilon) B) \cdot \exp\Big(-\frac{a\epsilon^2 B}{32}\Big) \leq \frac{1}{4mB}$$

by taking $a = O(\frac{\log(1/s)}{s})$ and $B = O\big(\frac{\log m}{s}\big)$. A union bound over all possible choices of ϵ (B possibilities) and m edges of E_0 implies the claim. $\square$

Claim 13.3.4. *With the parameters a and B above, G is $(4s \log(1/s), \Omega(\frac{s}{d}))$-dense.*

Proof. Fix a subset S of measure $4s \log(1/s)$. For a vertex v of G_0, let $\mu(v) := \frac{|\mathsf{cloud}(v) \cap S|}{B}$. Note that $\mathbb{E}_v[\mu(v)] = 4s \log(1/s)$. Partition V_{G_0} into $t + 1$ buckets $B_0, \ldots, B_t$ ($t := \lceil \log_2(1/s) \rceil$), such that B_0 contains v such that $\mu(v) \leq s$, and for $i \geq 1$, B_i contains v such that $\mu(v) \in (2^{i-1}s, 2^i s]$. Denote

$$\mu(B_i) := \frac{\sum_{v \in B_i} \mu(v)}{|V_{G_0}|} .$$

Clearly $\mu(B_0) \leq s$. Pick $i \in \{1, \ldots, t\}$ with the largest $\mu(B_i)$. We have $\mu(B_i) \geq 2s$ since $\mathbb{E}_v[\mu(v)] \geq 4s \log(1/s)$. Let $\gamma = 2^{i-1}s$. All vertices of B_i has $\mu(v) \in [\gamma, 2\gamma]$, so $|B_i| \geq (s/\gamma)n$. We use Turán's Theorem.

Theorem 13.3.2 ([Tur41])**.** *Let G be any graph with n vertices such that G is K_{r+1}-free. Then the number of edges in G is at most $\frac{r-1}{r} \cdot \frac{n^2}{2}$.*

Since G_0 has no independent set with more than ns vertices, we can apply Turán's Theorem to the complement of the subgraph of G_0 induced by B_i, so that the subgraph of

G_0 induced by B_i has at least $\frac{|B_i|(|B_i|-1)}{2} - \frac{ns-1}{ns} \cdot \frac{|B_i|^2}{2} = \frac{|B_i|}{2}(\frac{|B_i|}{ns} - 1) = \Omega(\frac{s}{\gamma^2}n)$ edges. This is at least $\Omega(\frac{s}{d\gamma^2})$ fraction of the total number of edges.

For each of these edges, by Claim 13.3.3, at least $\gamma^2/8$ fraction of the edges from the bipartite graph connecting the clouds of its two endpoints, lie in the subgraph induced by S (since $\gamma \geq s$). Overall, we conclude that there are at least $\Omega(\frac{s}{d\gamma^2}) \cdot \frac{\gamma^2}{8} = \Omega(\frac{s}{d})$ fraction of edges inside the subgraph induced by S. $\square$

Sparsification. Recall that G' is obtained from G by deleting at most $\frac{1}{d^2}$ fraction of vertices to have girth greater than g. In the completeness case, G' has an independent set of measure at least $c - 1/d^2 = \Omega(\frac{1}{\log d})$. In the soundness case, G is $(4s\log(1/s), \Omega(\frac{s}{d}))$-dense, so G' is (β, α)-dense where $\beta := \Omega(\frac{\log^4 d}{d}), \alpha := \Omega(\frac{\log^3 d}{d^2})$. Using density of G', we sparsify G' again — keep each edge of G' by probability $\frac{kn}{|E_{G'}|}$ so that the expected total number of edges is kn.

Fix a subset $S \subseteq V_{G'}$ of measure β. Since there are at least α fraction of edges in the subgraph induced by S, the expected number of picked edges in this subgraph is at least αkn. By Chernoff bound, the probability that it is less than $\frac{\alpha kn}{8}$ is at most $\exp(-\frac{\alpha kn}{32})$. By union bound over all sets of measure exactly β (there are at most $\binom{n}{n\beta} \leq \exp(2\beta\log(1/\beta)n)$ of them), and over all possible values of β (there are at most n possible sizes), the desired property fails with probability at most

$$n \cdot \max_{\beta \in [\beta_0, 1]} \left\{ \exp(-\alpha kn/32) \cdot \exp(2\beta \log(1/\beta)n) \right\} \leq n \cdot e^{-n}$$

when $k = O(\frac{\beta\log(1/\beta)}{\alpha}) = O(d\log^2 d)$. In the last step we remove all the vertices of degree more than $10k$. Since the expected degree of each vertex is at most $2k$, the expected fraction of deleted vertices is $\exp(-\Omega(k)) \ll \beta$.

Combining all these results, we have a graph with small degree $10k = O(d\log^2 d)$ and girth strictly greater than g, where it is NP-hard to approximate MIS within a factor of $\frac{c-\frac{1}{d^2}}{\beta} = \Omega(\frac{d}{\log^5 d}) = \Omega(\frac{k}{\log^7 k})$. Therefore, it is NP-hard to approximate MIS-k-g within a factor of $\Omega(\frac{k}{\log^7 k})$. $\square$

13.4 Hardness for Longer Cycles and Connection to FVS

We prove Theorem 12.3.1, which improves Theorem 12.2.1 in the sense that in the completeness case, a small subset $F \subseteq V_G$ intersects not only cycles of length exactly k, but also all cycles of length $3, 4, ..., O(\frac{\log n}{\log\log n})$. The reduction and the soundness analysis are exactly the same. We show the following lemma for the completeness case which is again almost identical to Lemma 13.2.1, but carefully keeps track of parameters to consider cycles of increasing length.

Lemma 13.4.1. *Suppose P has a vertex cover C of measure c. For any $\epsilon > 0$, with probability at least $3/4$, there exists a subset $F \subseteq V_G$ of measure at most $c + \epsilon$ such that the induced subgraph $V_G \setminus F$ has no cycle of length $O(\frac{\log n}{\log\log n})$. The constant hidden in O depends on k, ϵ and the degree d of P.*

Proof. Following the proof of Lemma 13.2.1, the expected number of cycles of length k' that avoid F is bounded by $n(2d)^{k'}B^{k'} \cdot k'(\frac{adk'}{B})^{k'} \leq n(Rk')^{k'}$ where R is a constant depending only on a and d (both are independent of k'). Compared to the general bound $n(kd)^k B^k \cdot \frac{(adk)^{k^2}}{B^k}$ for any 2-connected graph with k vertices, (kd) is improved to $2d$ since every vertex has degree 2, and $(adk)^{k^2}$ is improved to $k'(adk')^{k'}$ since a cycle has exactly k' edges.

With probability at least $3/4$, the number of such cycles of length up to k' is at most $4n(Rk')^{k'+1}$. Let $B \geq \frac{4(Rk')^{k'+1}}{\epsilon}$. Then these cycles can be covered by at most $\epsilon n B = \epsilon N$ vertices. If $k' = \frac{\log n}{\log\log n}$, then $k'^{k'} = \exp(k' \log k')$ is also $o(n)$, we can take B linear in n and $k' \geq \Omega(\frac{\log N}{\log\log N})$. $\square$

13.5 Hardness of k-VERTEX SEPARATOR

In this section, we prove that an f-true approximation algorithm for k-VERTEX SEPARATOR that runs in time $\text{poly}(n, k)$ will result in $2f^2$-approximation algorithm for the DENSEST k-SUBGRAPH, proving Theorem 12.2.4. In particular, $O(\log k)$-true approximation for k-VERTEX SEPARATOR in time $\text{poly}(n, k)$ will lead to $O(\log^2 n)$-approximation for DENSEST k-SUBGRAPH.

Given an undirected graph $G = (V, E)$ and an integer k, DENSEST k-SUBGRAPH asks to find $S \subseteq V$ with $|S| = k$ to maximize the number of edges of $G|_S$. It is one of the notorious problems in approximation algorithms. The current best approximation algorithm achieves $\approx O(n^{1/4})$-approximation [BCC+10]. While only PTAS is ruled out assuming $\mathbf{NP} \not\subseteq \cap_{\epsilon>0}\mathbf{BPTIME}(2^{n^\epsilon})$ [Kho06], there are strong gap instances for Sum-of-Squares hierarchies of convex relaxations ($n^{\Omega(1)}$ gap for $n^{\Omega(1)}$ rounds) [BCV+12], so having a $\text{polylog}(n)$-approximation algorithm for DENSEST k-SUBGRAPH seems unlikely or will lead to a breakthrough. Therefore, it may be the case that achieving $O(\log k)$-approximation for k-VERTEX SEPARATOR requires superpolynomial dependence on k in the running time.

Our reduction is close to that of Drange et al. [DDvH14] who reduced Clique to k-VERTEX SEPARATOR to prove $\mathbf{W}[1]$-hardness. Formally, we introduce another problem called MINIMUM k-EDGE COVERAGE. Given an undirected graph G and an integer k, the problem asks to find the minimum number of vertices whose induced subgraph has at least k edges. This problem can be thought as a *dual* of DENSEST k-SUBGRAPH in a sense that given the same input graph, the optimum of DENSEST a-SUBGRAPH is at least b if and only if the optimum of MINIMUM b-EDGE COVERAGE is at most a. Hajiaghayi and Jain [HJ06] proved the following theorem, relating their approximation ratios.

Theorem 13.5.1 ([HJ06]). *If there is a polynomial time f-approximation algorithm for* MINIMUM k-EDGE COVERAGE, *then there is a polynomial time $2f^2$-approximation algorithm for* DENSEST k-SUBGRAPH.

We introduce a reduction from MINIMUM k-EDGE COVERAGE to k-VERTEX SEPARATOR. Given an instance $G = (V, E)$ and k for MINIMUM k-EDGE COVERAGE, the instance of k-VERTEX SEPARATOR $G' = (V', E')$ and k' is created as follows. Let $n = |V|$, $m = |E|$, and $M = n + 1$.

- $V' = V \cup \{e_i : e \in E, i \in [M]\}$. Note that $|V'| = n + Mm$.
- $E' = \binom{V}{2} \cup \{(u, e_i) : u \in V, e \in E, u \in e, i \in [M]\}$. Intuitively, the subgraph induced by $V \subseteq V'$ forms a clique, and for each $e = (u, v) \in E$ and $i \in [M]$, e_i is connected to u and v in G'.
- $k' = |V'| - Mk$.

Lemma 13.5.2. *Every instance of* k-VERTEX SEPARATOR *produced by the above reduction has an optimal solution $S \subseteq V'$ such that indeed $S \subseteq V$.*

Proof. Take an optimal solution S such that $G'_{V' \setminus S}$ has each connected component with at most k vertices. Suppose S contains e_i for some $e = (u, v) \in E$ and $i \in [M]$. There are three cases.

- $u, v \notin S$: Since there is an edge $(u, v) \in E'$, u and v are in the same connected component in $G'|_{V' \setminus S}$. Removing e_i from S and adding u to S still results in an optimal solution.

- $u \in S$, $v \notin S$: Removing e_i from S and adding u to S decreases the size of the connected component of u by 1, and creates a new singleton component consisting e_i. It is still an optimal solution.

- $u, v \in S$: Removing e_i from S just creates a new singleton component consisting e_i. It is a strictly better solution.

We can repeatedly apply one of these three operations until S is an optimal solution contained in V. $\square$

When $S \subseteq V$, $G'|_{V' \setminus S}$ has the following connected components.

- One component $(V \setminus S) \cup \{e_i : e = (u, v) \in E, \{u, v\} \nsubseteq S, i \in [M]\}$. Call it the *giant component*.

- For each $e = (u, v) \in E$ with $u, v \in S$ and $i \in [M]$, a singleton component $\{e_i\}$. Call them *singleton components*.

Suppose that the instance of MINIMUM k-EDGE COVERAGE admits a solution $T \subseteq V$ such that the induced subgraph $G|_T$ has $l \geq k$ edges. Let $S = T$. Since $|V \setminus S| = n - |T|$ and $|\{(u, v) \in E : \{u, v\} \nsubseteq T\}| = m - l$, In $G'|_{V' \setminus S}$, the giant component will have cardinality

$$n - |T| + M(m - l) \leq n - |T| + M(m - k) \leq n + M(m - k) = |V'| - Mk = k'.$$

On the other hand, suppose that the instance of k-VERTEX SEPARATOR has a solution S. By Lemma 13.5.2, assume that $S \subseteq V$. Let l be the number of edges in $G|_S$. The size of the giant component is at least $n - |S| + M(m - l) \geq M(m - l - 1) + 1$ since $M > n$. Since S is a feasible solution of the k'-VERTEX SEPARATOR, we must have

$$\begin{aligned} & M(m - l - 1) + 1 \leq k' = M - mk \\ \Rightarrow\ & l \geq k. \end{aligned}$$

Therefore, S is also a solution to MINIMUM k-EDGE COVERAGE. This proves that the above reduction is an approximation preserving reduction from MINIMUM k-EDGE COVERAGE to k-VERTEX SEPARATOR, proving Theorem 12.2.4.

Chapter 14

Algorithms for k-Star Transversal

14.1 Approximation Algorithm for k-Star Transversal

In this section, we show that k-Star Transversal admits an $O(\log k)$-approximation algorithm, matching the $\Omega(\log k)$-hardness obtained via a simple reduction from Minimum Dominating Set on degree-$(k-1)$ graphs [CC08], and proving Theorem 12.2.3. Let $G = (V_G, E_G)$ be the instance of k-Star Transversal. This problem has a natural interpretation that it is equivalent to finding the smallest $F \subseteq V_G$ such that the subgraph induced by $V_G \setminus F$ has maximum degree at most $k-2$. Our algorithm consists of two phases.

1. Iteratively solve 2-rounds of Sherali-Adams linear programming (LP) hierarchy and put vertices with a large fractional value in the transversal. If this phase terminates with a partial transversal F, the remaining subgraph induced by $V_G \setminus F$ has small degree (at most $2k$) and the LP solution to the last iteration is *highly fractional.*

2. We reduce the remaining problem to *Constrained Set Multicover* and use the standard greedy algorithm. While the analysis of the greedy algorithm for Constrained Set Multicover is used as a black-box, low degree of the remaining graph and high fractionality of the LP solution imply that the analysis is almost tight for our problem as well.

Iterative Sherali-Adams. Given G, 2-rounds of Sherali-Adams hierarchy of LP relaxation has variables $\{x_v\}_{v \in V_G} \cup \{x_{u,v}\}_{u,v \in V_G}$. An integral solution $y : V_G \mapsto \{0, 1\}$, where

$y(v) = 1$ indicates that v is picked in the transversal, naturally gives a feasible solution to the hierarchy by $x_v = y_v$, $x_{u,v} = y_u y_v$. Consider the following relaxation for k-STAR TRANSVERSAL.

$$
\begin{array}{lll}
\text{minimize} & \displaystyle\sum_{v \in V_G} x_v & \\
\text{subject to} & 0 \leq x_{u,v}, x_v \leq 1 & \forall u, v \in V_G \\
& x_{u,v} \leq x_u & \forall u, v \in V_G \\
& x_u + x_v - x_{u,v} \leq 1 & \forall u, v \in V_G \\
& \displaystyle\sum_{v:(u,v) \in E_G} (x_v - x_{u,v}) \geq (\mathsf{deg}(u) - k + 2)(1 - x_u) & \forall u \in V_G
\end{array}
$$

The first three constraints are common to any 2-rounds of Sherali-Adams hierarchy, and ensure that for any $u, v \in V_G$, the *local distribution* on four assignments $\alpha : \{u, v\} \mapsto \{0, 1\}$ forms a valid distribution. In other words, the following four numbers are nonnegative and sum to 1: $\Pr[\alpha(u) = \alpha(v) = 1] := x_{u,v}$, $\Pr[\alpha(u) = 0, \alpha(v) = 1] := x_v - x_{u,v}$, $\Pr[\alpha(u) = 1, \alpha(v) = 0] := x_u - x_{u,v}$, $\Pr[\alpha(u) = \alpha(v) = 0] := 1 - x_u - x_v + x_{u,v}$.

The last constraint is specific to k-STAR TRANSVERSAL, and it is easy to see that it is a valid relaxation: Given a feasible integral solution $y : V_G \mapsto \{0, 1\}$, the last constraint is vacuously satisfied when $y_u = x_u = 1$, and if not, it requires that at least $\deg(u) - k + 2$ vertices should be picked in the transversal so that there is no copy of k-Star in the induced subgraph centered on u. The first phase proceeds as the following.

- Let $S \leftarrow \emptyset$.
- Repeat the following until the size of S does not increase in one iteration.
  - Solve the above Sherali-Adams hierarchy for $V_G \setminus S$ — it means to solve the above LP with additional constraints $x_v = 1$ for all $v \in S$, which also implies $x_{u,v} = x_u$ for $v \in S, u \in V_G$. Denote this LP by $\mathsf{SA}(S)$.
  - $S \leftarrow \{v : x_v \geq \frac{1}{\alpha}\}$, where $\alpha := 10$.

We need to establish three properties from the first phase:

- The size of S is close to that of the optimal k-STAR TRANSVERSAL.

- Maximum degree of the subgraph induced by $V_G \setminus S$ is small.
- The remaining solution has small fractional values — $x_v < \frac{1}{\alpha}$ for all $v \in V_G \setminus S$.

The final property is satisfied by the procedure. The following two lemmas establish the other two properties.

Lemma 14.1.1. *Let* Frac *be the optimal value of* $\mathsf{SA}(\emptyset)$. *When the above procedure terminates,* $|S| \leq \alpha \,\mathsf{Frac}$.

Proof. Assume that the above loop iterated l times, and for $i = 0, ..., l$, let S_i be S after the ith loop such that $S_0 = \emptyset, ..., S_l = S$. Let Opt_i be the optimal value of the program $\mathsf{SA}(S_i)$ (which is run in the $(i+1)$th iteration of the loop), and let $\mathsf{Frac}_i := \mathsf{Opt}_i - |S_i|$. Note that $\mathsf{Frac} = \mathsf{Frac}_0$. We will prove $|S| = |S_l| \leq \alpha \,\mathsf{Frac}_0 = \alpha \,\mathsf{Frac}$, using induction from the last iteration. For the base case, note that $|S_l| = |S_{l-1}|$ by the termination condition of the loop.

For $i = l-2, l-1, ..., 0$, we show that $|S_l| - |S_i| \leq \alpha \,\mathsf{Frac}_i$. Let x be the optimal fraction solution to $\mathsf{SA}(S_i)$ used to compute S_{i+1}. In particular, $S_{i+1} := \{v : x_v \geq \frac{1}{\alpha}\}$, and $\mathsf{Frac}_i = \sum_{v \in S_{i+1} \setminus S_i} x_v + \sum_{v : x_v < \frac{1}{\alpha}} x_v$. Let x' be the solution obtained by *partially rounding* x in the following way.

- $x'_v = 1$ if $v \in S_{i+1}$. Otherwise, $x'_v = x_v$.
- $x'_{u,v} = x'_u$ ($v \in S_{i+1}$), x'_v ($u \in S_{i+1}$), or $x_{u,v}$ otherwise.

Its value is $|S_{i+1}| + \sum_{v : x_v < \frac{1}{\alpha}} x_v$, and it is easy to check that it is a feasible solution to $\mathsf{SA}(S_{i+1})$ (intuitively, *rounding up* only helps feasibility). Therefore,

$$|S_{i+1}| + \sum_{v : x_v < \frac{1}{\alpha}} x_v \geq \mathsf{Opt}_{i+1} = |S_{i+1}| + \mathsf{Frac}_{i+1},$$

which implies $\sum_{v : x_v < \frac{1}{\alpha}} x_v \geq \mathsf{Frac}_{i+1}$. Therefore,

$$\mathsf{Frac}_i = \sum_{v \in S_{i+1} \setminus S_i} x_v + \sum_{v : x_v < \frac{1}{\alpha}} x_v \geq \frac{1}{\alpha}(|S_{i+1}| - |S_i|) + \mathsf{Frac}_{i+1}\,.$$

Finally, we have

$$\begin{aligned}
& |S_l| - |S_i| \\
= \quad & (|S_l| - |S_{i+1}|) + (|S_{i+1}| - |S_i|) \\
\leq \quad & \alpha \,\mathsf{Frac}_{i+1} + (|S_{i+1}| - |S_i|) \\
\leq \quad & \alpha \,\mathsf{Frac}_i,
\end{aligned}$$

where the first inequality follows from the induction hypothesis. This completes the induction. $\square$

Lemma 14.1.2. *After the termination, every vertex has degree at most $2k$ in the subgraph induced by $V_G \setminus S$.*

Proof. We prove that whenever $G \setminus S$ has a vertex of degree more than $2k$, then $\mathsf{SA}(S)$ contains a vertex $v \notin S$ with $x_v \geq (1/\alpha)$, which implies the lemma. Let $G' := G \setminus S$, $\mathsf{Nbr}(u)$ and $\mathsf{Nbr}'(u)$ be the set of neighbors of u in G and G' respectively, and $\mathsf{deg}(u) = |\mathsf{Nbr}(u)|, \mathsf{deg}'(u) = |\mathsf{Nbr}'(u)|$. Finally, x be the optimal solution of $\mathsf{SA}(S)$.

Suppose that $\mathsf{deg}'(u) > 2k$ for some vertex $u \notin S$. If $x_u \geq \frac{1}{\alpha}$, we are done. Otherwise, consider the constraint

$$\sum_{v \in \mathsf{Nbr}(u)} (x_v - x_{u,v}) \geq (\mathsf{deg}(u) - k + 2)(1 - x_u).$$

For $v \in S$, $x_v = 1$ and $x_{u,v} = x_u$, so the above constraint is equivalent to

$$\sum_{v \in \mathsf{Nbr}'(u)} (x_v - x_{u,v}) \geq (\mathsf{deg}'(u) - k + 2)(1 - x_u).$$

Finally, using the fact that $x_u < \frac{1}{\alpha}$, we conclude that

$$\sum_{v \in \mathsf{Nbr}'(u)} (x_v - x_{u,v}) \geq (1 - \frac{1}{\alpha}) \mathsf{deg}'(u)(1 - \frac{k}{\mathsf{deg}'(u)}) \geq \frac{1}{\alpha} \cdot \mathsf{deg}'(u).$$

Therefore, there exists $v \in \mathsf{Nbr}'(u)$ such that $x_v \geq \frac{1}{\alpha}$. $\square$

Constrained Set Multicover. The first phase returns a set S whose size is at most α times the optimal solution and the subgraph induced by $V_G \setminus S$ has maximum degree at most $2k$. As above, let G' be the subgraph induced by $V_G \setminus S$, $\mathsf{Nbr}(u), \mathsf{Nbr}'(u)$ be the neighbors of u in G and G' respectively, and $\mathsf{deg}(u) = |\mathsf{Nbr}(u)|, \mathsf{deg}'(u) = |\mathsf{Nbr}'(u)|$. The remaining task is to find a small subset $F \subseteq V_G \setminus S$ such that the subgraph of G' (and G) induced by $V_G \setminus (S \cup F)$ has no vertex of degree at least $k - 1$. We reduce the remaining problem to the *Constrained Set Multicover* problem defined below.

Definition 14.1.3. *Given an set system $U = \{e_1, ..., e_n\}$, a collection of subsets $\mathcal{C} = \{C_1, ..., C_m\}$, and a positive integer r_e for each $e \in U$, the Constrained Set Multicover problem asks to find the smallest subcollection (each set must be used at most once) such that each element e is covered by at least r_e times.*

Probably the most natural greedy algorithm does the following:

- Pick a set C with the largest cardinality (ties broken arbitrarily).
- Set $r_e \leftarrow r_e - 1$ for $e \in C$. If $r_e = 0$, remove it from U. For each $C \in \mathcal{C}$, let $C \leftarrow C \cap U$.
- Repeat while U is nonempty.

Constrained Set Cover has the following standard LP relaxation, and Rajagopalan and Vazirani [RV98] showed that the greedy algorithm gives an integral solution whose value is at most H_d (i.e. the dth harmonic number) times the optimal solution to the LP, where d is the maximum set size.

$$\begin{array}{lll} \text{minimize} & \displaystyle\sum_{C \in \mathcal{C}} z_C & \\ \text{subject to} & \displaystyle\sum_{C : e \in C} z_C \geq r_e & e \in U \\ & 0 \leq z_C \leq 1 & C \in \mathcal{C} \end{array}$$

Our remaining problem, k-Star Transversal on G', can be thought as an instance of Constrained Set Cover in the following way: $U := \{u \in V_G \setminus S : \mathsf{deg}'(u) \geq k-1\}$ with $r_u := \mathsf{deg}'(u) - k + 2$, and for each $v \in V_G \setminus S$, add $\mathsf{Nbr}'(v) \cap U$ to $\mathcal{C}$. Intuitively, this formulation requires at least r_u neighbors be picked in the transversal whether u is picked or not. This is not a valid reduction because the optimal solution of the above formulation can be much more than the optimal solution of our problem. However, at least one direction is clear (any feasible solution to the above formulation is feasible for our problem), and it suffices to show that the above LP admits a solution whose value is close to the optimum of our problem. The LP relaxation of the above special case of Constrained Set Cover is the following:

$$\begin{array}{lll} \text{minimize} & \displaystyle\sum_{v \in V_G \setminus S} z_v & \\ \text{subject to} & \displaystyle\sum_{v : v \in \mathsf{Nbr}'(u)} z_v \geq \mathsf{deg}'(u) - k + 2 & u \in U \\ & 0 \leq z_v \leq 1 & v \in V_G \setminus S \end{array}$$

Consider the last iteration of the first phase where we solved $\mathsf{SA}(S)$. Let x be the optimal solution to $\mathsf{SA}(S)$ and $\mathsf{Frac} := \sum_v x_v - |S|$. Note that $x_v < \frac{1}{\alpha}$ when $v \notin S$. Define $\{y_v\}_{v \in V \setminus S}$ such that $y_v := 2x_v$.

Lemma 14.1.4. $\{y_v\}$ *is a feasible solution to the above LP for Constrained Set Cover.*

Proof. By construction $0 \leq y_v < \frac{2}{\alpha}$, so it suffices to check for each $u \in U$,

$$\sum_{v:v \in \mathsf{Nbr}'(u)} y_v \geq \mathsf{deg}'(u) - k + 2.$$

Fix $u \in U$. Recall that Sherali-Adams constraints on x imply that

$$\begin{aligned} &\sum_{v:\mathsf{Nbr}'(u)} (x_v - x_{u,v}) \geq (\mathsf{deg}'(u) - k + 2)(1 - x_u) \\ \Rightarrow &\sum_{v:\mathsf{Nbr}'(u)} x_v \geq (\mathsf{deg}'(u) - k + 2)(1 - x_u) \\ \Rightarrow &\sum_{v:\mathsf{Nbr}'(u)} 2x_v \geq \mathsf{deg}'(u) - k + 2, \end{aligned}$$

where the last line follows from the fact that $1 - \frac{1}{\alpha} > \frac{1}{2}$. $\square$

Therefore, Constrained Set Cover LP admits a feasible solution of value $2\,\mathsf{Frac}$, and the greedy algorithm gives a k-STAR TRANSVERSAL F with $|F| \leq 2 \cdot \mathsf{Frac} \cdot H_{2k}$. Since Frac is at most the size of the optimal k-STAR TRANSVERSAL for G' (and clearly G), $|S \cup F|$ is at most $O(\log k)$ times the size of the smallest k-STAR TRANSVERSAL of G.

Chapter 15

Algorithms for k-VERTEX SEPARATOR and k-PATH TRANSVERSAL

15.1 Techniques

Our algorithms for k-VERTEX SEPARATOR and k-EDGE SEPARATOR consist of the following three steps. We give a simple overview of our techniques for k-VERTEX SEPARATOR.

1. Spreading Metrics. *Spreading metrics* were introduced in Even et al. [ENRS00] and subsequently used for ρ-separator [ENRS99].[1] They assign lengths to vertices such that any subset S of vertices with $|S| \geq k$ that induce a single connected component are *spread apart*.

Given lengths x_v to each vertex $v \in V$, define $d_{u,v}$ to be the length of the shortest path from u to v, including the lengths of both u and v (so that $d_{u,u} = x_u$). Given a feasible solution $S \subseteq V$ for k-VERTEX SEPARATOR, let $x_v = 1$ if $v \in S$, and $x_v = 0$ if $v \notin S$. It is easy to see that two vertices u and v lie on the same component of $G|_{V \setminus S}$ if and only if $d_{u,v} = 0$. Otherwise, $d_{u,v} \geq 1$. Therefore, for every vertex v, the number of vertices that have distance strictly less than 1 from v must be at most k.

Spreading metrics are a continuous relaxation of the above integer program. We relax each distance x_v to have value in $[0, 1]$, and let $d_{u,v}$ still be the length of the shortest path from u to v. Let $f_{u,v} = \max(1 - d_{u,v}, 0)$. In the integral solution, it indicates whether

[1]The conference version of [ENRS00] precedes that of [ENRS99].

$d_{u,v} < 1$ or not. The constraint $\sum_u f_{v,u} \leq k$ for all $v \in V$ is a relaxation of the requirement that the number of vertices that have distance strictly less than 1 from v must be at most k.

Even though this relaxation does not exactly capture the integer problem, one crucial property of this relaxation is that for every $v \in V$ and $\epsilon \in (0, \frac{1}{2})$, the number of vertices that have distance at most ϵ from v can be at most $\frac{k}{1-\epsilon}$. This can proved via a simple averaging argument.

2. Low-Diameter Decomposition. Before we introduce our rounding algorithm, we briefly discuss why the previous algorithms based on the same (or stronger) relaxation has the approximation ratio depending on n.

The current best algorithm by Krauthgamer et al. [KNS09] further strengthened the above spreading metrics by requiring that they also form an ℓ_2^2 metric, and transformed them to an ℓ_2 metric. This black-box transformation of an n-points ℓ_2^2 metric incurs distortion of $\Omega(\sqrt{\log n})$, so the approximation ratio must depend on n.

The older work of Even et al. [ENRS99] used the rounding algorithm of Garg et al. [GVY96] that iterative takes a ball of small radius from the graph. More specifically, they defined $\mathsf{vol}(v, r)$ to be the total sum of lengths in the ball of radius r centered at v, and grow r until the boundary-volume ratio becomes $O(\log(\frac{\mathsf{vol}(v,\frac{1}{2})}{\mathsf{vol}(v,0)}))$. To make $\mathsf{vol}(v, 0)$ nonzero, a *seed value* of $\epsilon \cdot \mathsf{Opt}$ must be added to the the definition of $\mathsf{vol}(v, r)$. But when $k = O(1)$ so that the number of balls we need to remove from the graph is $\Omega(n)$, this incurs extra cost of $\Omega(\epsilon n \mathsf{Opt})$, forcing ϵ to depend on n.

We apply another standard technique for the *low-diameter decomposition* to our spreading metrics. In particular, our algorithm is similar to that of Carlinescu et al. [CKR05], preceded by a simple rounding algorithm that removes every vertex with large x_v. One simple but crucial observation is that the performance of this algorithm only depends on the size of the ball around each vertex, which is exactly what spreading metrics is designed for! Since the size of each ball of radius $\frac{1}{2}$ is at most $O(k)$, we can guarantee that we can delete at most $O(\log k) \cdot \mathsf{Opt}$ vertices so that each connected component has at most $O(k)$ vertices.

When $k = O(1)$, to the best of our knowledge, this is a rare example where the number of partitions (i.e., the number of balls taken) is $\Omega(n)$ but the approximation ratio is much smaller than that. The original rounding algorithm of Carlinescu et al. [CKR05] is applied to 0-*Extension* with k terminals to achieve $O(\log k)$-approximation, where only k balls are needed to be taken. The famous $O(\log k)$-approximation for Multicut with k source-sink pairs [GVY96] also required only k partitions.

3. Cleanup. After running the bicriteria approximation algorithm to make sure that each connected component has size at most $O(k)$, for k-VERTEX SEPARATOR, we run the exhaustive search for each component to have the true approximation. This incurs the extra running time of $2^{O(k)}n$, but our hardness result implies that the superpolynomial dependence on k may be necessary.

For k-EDGE SEPARATOR, essentially the same bicriteria approximation algorithm works. After that, for each component, we use (a variant of) Racke's $O(\log n)$-true approximation algorithm for MINIMUM BISECTION to each component to make sure that each component has at most k vertices. The existence of *true approximation* for MINIMUM BISECTION is a key difference between the vertex version and the edge version. Even $O(\sqrt{\log n})$-bicriteria approximation is known for the vertex version of MINIMUM BISECTION [FHL08], but our hardness result for the vertex version suggests that this algorithm is not likely to be applicable. While MINIMUM BISECTION asks to partition the graph into two pieces while k-EDGE SEPARATOR may need to partition it into many pieces, we prove that as long as each connected component has size at most $\frac{3k}{2}$, a simple trick makes the two problems equivalent.

15.2 Algorithm for k-VERTEX SEPARATOR

15.2.1 Spreading Metrics

Our relaxation is close to *spreading metrics* used for ρ-separator [ENRS99]. While their relaxation involves an exponential number of constraints and is solved by the ellipsoid algorithm, we present a simpler relaxation where the total number of variables and constraints is polynomial. Our relaxation has the following variables.

- x_v for $v \in V$: It indicates whether v is removed or not.
- $d_{u,v}$ for $(u, v) \in V \times V$: Given $\{x_v\}_{v \in V}$ as lengths on vertices, $d_{u,v}$ is supposed to be the minimum distance between u and v. Let $\mathcal{P}_{u,v}$ be the set of simple paths from u to v, and given $P = (u_0 := u, u_1, \ldots, u_p := v) \in \mathcal{P}_{u,v}$, let $d(P) = x_{u_0} + \cdots + x_{u_p}$. Formally, we want
$$d_{u,v} = \min_{P \in \mathcal{P}_{u,v}} d(P).$$
Note that $d_{u,v} = d_{v,u}$ and $d_{u,u} = x_u$.
- $f_{u,v}$ for all $(u, v) \in V \times V$: It indicates whether u and v belong to the same connected component or not.

Our LP is written as follows.

$$\begin{array}{lll}
\text{minimize} & \sum_{v \in V} x_v & \\
\text{subject to} & d_{u,v} \le \min_{P \in \mathcal{P}_{u,v}} d(P) & \forall (u,v) \in V \times V \qquad (15.1) \\
& f_{u,v} \ge 1 - d_{u,v} & \forall (u,v) \in V \times V \\
& f_{u,v} \ge 0 & \forall (u,v) \in V \times V \\
& \sum_{u \in V} f_{v,u} \le k-1 & \forall v \in V \qquad (15.2) \\
& x_v \ge 0 & \forall v \in V \qquad (15.3)
\end{array}$$

(15.1) can be formally written as

$$\begin{array}{ll}
d_{u,u} = x_u & \forall u \in V \\
d_{u,w} \le d_{u,v} + x_w & \forall (u,v) \in V \times V, (v,w) \in E
\end{array}$$

Therefore, the size of our LP is polynomial in n. It is easy to verify that our LP is a relaxation — given a subset $S \subseteq V$ such that each connected component of $G|_{V \setminus S}$ has at most $k-1$ vertices, the following is a feasible solution with $\sum_v x_v = |S|$.

- $x_v = 1$ if $v \in S$. $x_v = 0$ if $v \notin S$.
- $d_{u,v} = \min_{P \in \mathcal{P}_{u,v}} d(P)$.
- $f_{u,v} = 1$ if u and v are in the same component of $G|_{V \setminus S}$. Otherwise $f_{u,v} = 0$.

Fix an optimal solution $\{x_v\}_v, \{d_{u,v}, f_{u,v}\}_{u,v}$ for the above LP. It only ensures that $d_{u,v} \le \min_{P \in \mathcal{P}_{u,v}} d(P)$, so a priori $d_{u,v}$ can be strictly less than $\min_{P \in \mathcal{P}_{u,v}} d(P)$. However, in that case increasing $d_{u,v}$ still maintains feasibility, since larger $d_{u,v}$ provides a looser lower bound of $f_{u,v}$ and lower $f_{u,v}$ helps to satisfy (15.2). For the subsequent sections, we assume that $d_{u,v} = \min_{P \in \mathcal{P}_{u,v}} d(P)$, and $f_{u,v} = \max(1 - d_{u,v}, 0)$ for all u, v.

15.2.2 Low-diameter Decomposition

Given the above spreading metrics, we show how to decompose a graph such that each connected component has small number of vertices. Our algorithm is based on that of Calinescu et al. [CKR05]. One major difference is to bound the size of each *ball* by $O(k)$ in the analysis, and simple algorithmic steps to ensure this fact.

Fix $\epsilon \in (0, \frac{1}{2})$. Given an optimal solution $\{x_v\}_{v \in V}$, the first step of the rounding algorithm is to remove every vertex $v \in V$ with $x_v \geq \epsilon$. This simple step is crucial in bounding the size of the ball around each vertex. It removes at most $\frac{\mathsf{Opt}}{\epsilon}$ vertices. Let $V' := V \setminus \{v : x_v \geq \epsilon\}$, and $G' = (V', E')$ be the subgraph of G induced by V'. Let $d'_{u,v}$ be the minimum distance between u and v in G', and let $f'_{u,v} := \max(1 - d'_{u,v}, 0)$. Since removing vertices only increases distances, $d'_{u,v} \geq d_{u,v}$ and $f'_{u,v} \leq f_{u,v}$ for all $(u, v) \in V' \times V'$.

Our low-diameter decomposition removes at most $O(\frac{\log k}{\epsilon}) \cdot \sum_{v \in V'} x_v$ vertices so that each resulting connected component has at most $\frac{k}{1-2\epsilon}$ vertices. It proceeds as follows.

- Pick $X \in [\epsilon/2, \epsilon]$ uniformly at random.
- Choose a random permutation $\pi : V' \mapsto V'$ uniformly at random.
- Consider the vertices one by one, in the order given by π. Let w be the considered vertex (we consider every vertex whether it was previously disconnected, removed or not).
  - For each vertex $v \in V'$ with $d'_{w,v} - x_v \leq X \leq d'_{w,v}$, we remove v when it was neither removed nor *disconnected* previously.
  - The vertices in $\{v : d'_{w,v} < X\}$ are now disconnected from the rest of the graph. Say these vertices are *disconnected*.

For each vertex w, let $B(w) := \{v \in V' : d'_{w,v} \leq 2\epsilon\}$. A simple averaging argument bounds $|B(w)|$.

Lemma 15.2.1. *For each vertex w, $|B(w)| \leq \frac{k}{1-2\epsilon}$.*

Proof. Assume towards contradiction that $|B(w)| > \frac{k}{1-2\epsilon}$. For all $u \in B(w)$,

$$f_{w,u} \geq f'_{w,u} \geq 1 - d'_{w,u} \geq 1 - 2\epsilon.$$

Furthermore, even for $u \notin B(w)$, our LP ensures that $f_{w,u} \geq 0$. Therefore,

$$\sum_{u \in V} f_{w,u} \geq \sum_{u \in B(w)} f_{w,u} \geq (1 - 2\epsilon)|B(w)| > k,$$

contradicting (15.2) of our LP. □

Note that at the end of the algorithm, every vertex is removed or disconnected, since every $w \in V'$ becomes removed or disconnected after being considered. Moreover, each connected component is a subset of $\{v : d'_{w,v} < X\}$ for some $w \in V'$ and $X \leq \epsilon$, which is a subset of $B(w)$. Therefore, each connected component has at most $\frac{k}{1-2\epsilon}$ vertices. We finally analyze the probability that a vertex v is removed.

Lemma 15.2.2. *The probability that $v \in V'$ is removed is at most $O(\frac{\log k}{\epsilon}) \cdot x_v$.*

Proof. Fix a vertex $v \in V'$. When $w \in V'$ is considered, v can be possibly removed only if

$$\begin{aligned} & d'_{v,w} - x_v \leq \epsilon \\ \Rightarrow\ & d'_{v,w} \leq 2\epsilon && \text{(since } x_v \leq \epsilon) \\ \Rightarrow\ & w \in B(v). \end{aligned}$$

Let $W = \{w_1, \ldots, w_p\}$ be such vertices such that $d'_{v,w_1} \leq \cdots \leq d'_{v,w_p} \leq 2\epsilon$. By Lemma 15.2.1, $p \leq \frac{k}{1-2\epsilon}$.

Fix i and consider the event that v is removed when w_i is considered. This happens only if $d'_{v,w_i} - x_v \leq X \leq d'_{v,w_i}$. For fixed such X, a crucial observation is that if w_j with $j < i$ is considered before w_i, since $d'_{v,w_j} - x_v \leq X$, v will be either removed or disconnected when w_j is considered. In particular, v will not be removed by w_i. Given these observations, the probability that v is removed is bounded by

$$\begin{aligned} \Pr[v \text{ is removed}] &= \sum_{i=1}^{p} \Pr[v \text{ is removed when } w_i \text{ is considered}] \\ &= \sum_{i=1}^{p} \Pr[X \in [d'_{v,w_i} - x_v, d'_{v,w_i}] \text{ and } w_i \text{ comes before } w_1, \ldots, w_{i-1} \text{ in } \pi] \\ &\leq \sum_{i=1}^{p} \frac{2x_v}{\epsilon i} = x_v \cdot O(\frac{\log p}{\epsilon}) = x_v \cdot O(\frac{\log k}{\epsilon}). \end{aligned}$$

□

Therefore, the low-diameter decomposition removes at most $O(\frac{\log k}{\epsilon}) \cdot \sum_v x_v \leq O(\frac{\log k}{\epsilon}) \cdot$ Opt vertices so that each resulting connected component has at most $\frac{k}{1-2\epsilon}$ vertices. This gives a bicriteria approximation algorithm that runs in time $\mathrm{poly}(n, k)$, proving Theorem 12.2.2.

15.3 k-PATH TRANSVERSAL

Let $G = (V, E)$ and $k \in \mathbb{N}$ be an instance of k-PATH TRANSVERSAL, where we want to find the smallest $S \subseteq V$ such that the length of the longest path in $G|_{V \setminus S}$ (denoted by $l(G|_{V \setminus S})$) is strictly less than k. Recall that the length here denotes the number of vertices in a path. Call a path *l-path* if it has l vertices.

Let $\mathcal{P}_k$ be the set of all simple paths of length k. Our algorithm starts by solving the following naive LP.

$$\begin{aligned}
\text{minimize} \quad & \sum_{v \in V} x_v & & \\
\text{subject to} \quad & \sum_{i=1}^{k} x_{v_i} \geq 1 & & \forall P = (v_1, \ldots, v_k) \in \mathcal{P}_k \\
& x \geq 0 & & \forall v \in V \times V
\end{aligned} \tag{15.4}$$

When G is a clique with n vertices, any feasible solution needs to remove at least $n - k + 1$ vertices while the above LP has the optimum at most $\frac{n}{k}$ by giving $\frac{1}{k}$ to every x_v. Therefore, it has an integrality gap close to k, but our algorithm bypasses this gap.

Lemma 15.3.1. *The above LP can be solved in time $k^{O(k)} n^{O(1)}$.*

Proof. Given the current solution $\{x_v\}_v$, we show how to check (15.4) in FPT time, so that the LP can be solved efficiently via the ellipsoid algorithm. In particular, it suffices to compute

$$\min_{P=(v_1,\ldots,v_k) \in \mathcal{P}_k} \sum_{i=1}^{k} x_{v_i}.$$

Our algorithm is a simple variant of an algorithm for the k-Path problem. Our presentation follows Williams [Wil13].

Call a set of functions $F = \{f_i\}_i$ with $f_i : [n] \mapsto [k]$ a *k-perfect hash family* if for any subset $S \subseteq [n]$ with $|S| = k$, there exists $f_i \in F$ such that $f_i(S) = [k]$. Naor et al. [NSS95] show that efficiently computable such F exists with $|F| = 2^{O(k)} \log n$.

For each $f_i \in F$ and a permutation $\pi \in \mathcal{S}_k$, we construct a directed acyclic graph (DAG) $D_{f_i,\pi}$, where for each edge $(u, v) \in E$, we add an arc from u to v if $\pi(f_i(u)) < \pi(f_i(v))$. Finding the k-directed path that minimizes $\sum_{i=1}^{k} x_{v_i}$ in a DAG can be done via dynamic programming. For $v \in V$ and $l \in [k]$, let $T[v, l]$ be the minimum weighted length of l-path that ends at v, and compute T in topological order.

Let $P^* = (v_1^*, \ldots, v_k^*)$ be the path that minimizes $\min_{P=(v_1,\ldots,v_k)\in\mathcal{P}} \sum_{i=1}^k x_{v_i}$. There must be $f_i \in F$ and $\pi \in \mathcal{S}_k$ such that $\pi(f_i(v_i^*)) < \pi(f_i(v_{i+1}^*))$ and arc (v_i^*, v_{i+1}^*) exists for $1 \le i < k$. For this f_i and π, the above dynamic programming algorithm for $D_{f_i,\pi}$ finds P^*.

The dynamic programming takes $n^{O(1)}$ time, and we try $2^{O(k)}k! \log n = k^{O(k)} \log n$ different pairs (f_i, π), so the separation oracle runs in time $k^{O(k)}n^{O(1)}$. Our LP has only n variables, so the total LP runs in time $k^{O(k)}n^{O(1)}$. □

Solve the above LP to get an optimal solution $\{x_v\}_{v\in V}$. Let $\mathsf{FRAC} := \sum_v x_v$. Call a vertex $v \in V$ *red* if $x_v \ge \frac{1}{k}$. Let R be the set of red vertices. One simple but crucial observation is that every k-path must contain at least one red vertices, since all non-red vertices have $x_v < \frac{1}{k}$.

Let S^* be the optimal solution of k-PATH TRANSVERSAL. Let $V^* := V \setminus S^*, R^* := R \setminus S^*$ and $G^* = G|_{V\setminus S^*}$. The result for k-PATH TRANSVERSAL requires the following lemma.

Lemma 15.3.2. *There exists $S' \subseteq V^*$ with $|S'| \le \frac{|R^*|}{k}$ vertices so that in the induced subgraph $G^*_{V^*\setminus S'}$, each connected component has at most k^3 red vertices.*

Proof. We prove the lemma by the following (possibly exponential time) algorithm: For each connected component C that has more than k^3 red vertices, take an arbitrary longest path, remove all vertices in it (i.e., add them to S') and charge its cost to all red vertices in C uniformly. Since the length of any longest path should not exceed k and C has more than k^3 red vertices, each red vertex in C gets charged at most $\frac{1}{k^2}$ in each iteration.

We argue that each vertex in G^* is charged at most k times. This is based on the following simple observation.

Fact 15.3.3. *In a connected component C, any two longest paths should intersect.*

Proof. Let $P_1 = (v_1, \ldots, v_p)$ and $P_2 = (u_1, \ldots, u_p)$ be two vertex-disjoint longest paths in the same connected component. Since they are in the same component, there exist $i, j \in [k]$ and another path $P_3 = (v_i, w_1, \ldots, w_q, u_j)$ such that $w_1, \ldots, w_q$ are disjoint from v's and u's (q may be 0). By reversing the order of P_1 or P_2, we can assume that $i, j \ge \frac{p+1}{2}$. Then $(v_1, \ldots, v_i, w_1, \ldots, w_q, u_j, \ldots, u_1)$ is a path with length at least $p+1$, contradicting the fact that P_1 and P_2 are longest paths. □

Therefore, if we remove one longest path from C, whether the remaining graph is still connected or divided into several connected components, the length of the longest path in each resulting connected component should be strictly less than the length of the longest path in C. Therefore, each vertex in G^* can be charged at most k times, and the total amount of charge is $k \cdot \frac{1}{k^2} = \frac{1}{k}$. $\square$

Consider $S^* \cup S'$. Its size is at most $\mathsf{Opt} + \frac{|R^*|}{k} \leq \mathsf{Opt} + \mathsf{FRAC} \leq 2\mathsf{Opt}$, since every red vertex has $x_v \geq \frac{1}{k}$, and each component of $G_{S^* \cup S'}$ has at most k^3 red vertices. We formally define the following generalization of k-VERTEX SEPARATOR.

> **k-SUBSET VERTEX SEPARATOR**
>
> **Input**: An undirected graph $G = (V, E)$, a subset $R \subseteq V$ and $k \in \mathbb{N}$.
>
> **Output**: Subset $S \subseteq V$ such that in the subgraph induced on $V \setminus S$ (denoted by $G|_{V \setminus S}$), each connected component has strictly less than k vertices from R.
>
> **Goal**: Minimize $|S|$.

Even though it seems a nontrivial generalization of k-VERTEX SEPARATOR, the analogous bicriteria approximation algorithm also exists. It is proved in Section 15.4.

Theorem 15.3.4. *For any $\epsilon \in (0, 1/2)$, there is a polynomial time $(\frac{1}{1-2\epsilon}, O(\frac{\log k}{\epsilon}))$-bicriteria approximation algorithm for k-SUBSET VERTEX SEPARATOR.*

For k-PATH TRANSVERSAL, run the above bicriteria approximation algorithm for k-SUBSET VERTEX SEPARATOR with $k \leftarrow k^3$ and $\epsilon \leftarrow \frac{1}{4}$. This returns a subset $S \subseteq V$ such that $|S| \leq O(\log k) \cdot \mathsf{Opt}$ and each connected component of $G_{V \setminus S}$ has at most $2k^3$ red vertices.

Now we solve for each connected component C. Since every k-path has to have at least one red vertex, removing every red vertex destroys every k-path. In particular, the optimal solution has at most $2k^3$ vertices in C. We run the following simple recursive algorithm.

- Find a k-path $P = (v_1, \ldots, v_k)$ if exists.
  - Otherwise, we found a solution — compare with the current best one and return.
- If the depth of the recursion is more than $2k^3$, return.
- For each $1 \leq i \leq k$,

– Remove v_i from the graph and recurse.

Finding a path takes time $2^{O(k)}n^{O(1)}$. In each stage the algorithm makes k branches, but the depth of the recursion is at most $2k^3$ and the algorithm is guaranteed to find the optimal solution. Therefore, it runs in time $2^{O(k)}n^{O(1)} \cdot k^{2k^3} = 2^{O(k^3 \log k)}n^{O(1)}$. This proves Theorem 12.2.1.

15.4 k-SUBSET VERTEX SEPARATOR

Given a graph $G = (V, E)$ and $k \in \mathbb{N}$. There is a subset $R \subseteq V$ of *red* vertices. Our relaxation has the following variables.

- x_v for $v \in V$: It indicates whether v is removed or not.
- $d_{u,v}$ for $(u, v) \in V \times V$: Given $\{x_v\}_{v \in V}$ as lengths on vertices, $d_{u,v}$ is supposed to be the minimum distance between u and v. Let $\mathcal{P}_{u,v}$ be the set of simple paths from u to v, and given $P = (u_0 := u, u_1, \ldots, u_p := v) \in \mathcal{P}_{u,v}$, let $d(P) = x_{u_0} + \cdots + x_{u_p}$. Formally, we want
$$d_{u,v} = \min_{P \in \mathcal{P}_{u,v}} d(P).$$
Note that $d_{u,v} = d_{v,u}$ and $d_{u,u} = x_u$.
- $f_{u,v}$ for all $(u, v) \in V \times V$: It indicates whether u and v belong to the same connected component or not.

Our LP is written as follows.

$$\begin{array}{lllr}
\text{minimize} & \displaystyle\sum_{v \in V} x_v & & \\
\text{subject to} & d_{u,v} \le \displaystyle\min_{P \in \mathcal{P}_{u,v}} d(P) & \forall (u, v) \in V \times V & (15.5) \\
& f_{u,v} \ge 1 - d_{u,v} & \forall (u, v) \in V \times V & \\
& f_{u,v} \ge 0 & \forall (u, v) \in V \times V & \\
& \displaystyle\sum_{u \in R} f_{v,u} \le k - 1 & \forall v \in V & (15.6)
\end{array}$$

(15.5) can be formally written as

$$\begin{aligned} d_{u,u} &= x_u && \forall u \in V \\ d_{u,w} &\le d_{u,v} + x_w && \forall (u,v) \in V \times V, (v,w) \in E \end{aligned}$$

The only change is that in (15.6), $f_{v,u}$ is summed over $u \in R$ instead of $u \in V$. It is clearly a relaxation.

Fix an optimal solution $\{x_v\}_v, \{d_{u,v}, f_{u,v}\}_{u,v}$ for the above LP. As usual, assume without loss of generality that $d_{u,v} = \min_{P \in \mathcal{P}_{u,v}} d(P)$, and $f_{u,v} = \max(1 - d_{u,v}, 0)$ for all u, v.

15.4.1 Low-diameter Decomposition

Fix $\epsilon \in (0, \frac{1}{2})$. Given an optimal solution $\{x_v\}_{v \in V}$, the first step of the rounding algorithm is to remove every vertex $v \in V$ with $x_v \ge \epsilon$. It removes at most $\frac{\mathsf{Opt}}{\epsilon}$ vertices.

Let $V' := V \setminus \{v : x_v \ge \epsilon\}$, and $G' = (V', E')$ be the subgraph of G induced by V'. Let $R' = V' \cap R$. Let $d'_{u,v}$ be the minimum distance between u and v in G', and let $f'_{u,v} := \max(1 - d'_{u,v}, 0)$. Since removing vertices only increases distances, $d'_{u,v} \ge d_{u,v}$ and $f'_{u,v} \le f_{u,v}$ for all $(u,v) \in V' \times V'$.

Our low-diameter decomposition removes at most $O(\frac{\log k}{\epsilon}) \cdot \sum_{v \in V'} x_v$ vertices so that each resulting connected component has at most $\frac{k}{1-2\epsilon}$ red vertices. It proceeds as follows.

- Pick $X \in [\epsilon/2, \epsilon]$ uniformly at random.
- Choose a random permutation $\pi : R' \mapsto R'$ uniformly at random.
- Consider the red vertices one by one, in the order given by π. Let w be the considered vertex (we consider every vertex whether it was previously disconnected, removed or not).
  - For each vertex $v \in V'$ with $d'_{w,v} - x_v \le X \le d'_{w,v}$, we remove v when it was neither removed nor *disconnected* previously.
  - The vertices in $\{v : d'_{w,v} < X\}$ are now disconnected from the rest of the graph. Say these vertices are *disconnected*.

For each vertex $w \in V'$, let $B(w) := \{v \in R' : d'_{w,v} \le 2\epsilon\}$. A simple averaging argument bounds $|B(w)|$.

Lemma 15.4.1. *For each vertex $w \in V'$, $|B(w)| \le \frac{k}{1-2\epsilon}$.*

Proof. Assume towards contradiction that $|B(w)| > \frac{k}{1-2\epsilon}$. For all $u \in B(w)$,

$$f_{w,u} \ge f'_{w,u} \ge 1 - d'_{w,u} \ge 1 - 2\epsilon.$$

Furthermore, even for $u \notin B(w)$, our LP ensures that $f_{w,u} \ge 0$. Therefore,

$$\sum_{u \in R} f_{w,u} \ge \sum_{u \in B(w)} f_{w,u} \ge (1-2\epsilon)|B(w)| > k,$$

contradicting (15.6) of our LP. □

Note that at the end of the algorithm, every red vertex is removed or disconnected, since every $w \in V'$ becomes removed or disconnected after being considered. Moreover, each connected component is a subset of $\{v : d'_{w,v} < X\}$ for some $w \in V'$ and $X \le \epsilon$, which is a subset of $B(w)$. Therefore, each connected component has at most $\frac{k}{1-2\epsilon}$ red vertices. We finally analyze the probability that a vertex $v \in V'$ is removed.

Lemma 15.4.2. *The probability that $v \in V'$ is removed is at most $O(\frac{\log k}{\epsilon}) \cdot x_v$.*

Proof. Fix a vertex $v \in V'$. When $w \in R'$ is considered, v can be possibly removed only if

$$\begin{aligned} & d'_{v,w} - x_v \le \epsilon \\ \Rightarrow\ & d'_{v,w} \le 2\epsilon && \text{(since } x_v \le \epsilon\text{)} \\ \Rightarrow\ & w \in B(v). \end{aligned}$$

Let $W = \{w_1, \dots, w_p\}$ be such vertices such that $d'_{v,w_1} \le \cdots \le d'_{v,w_p} \le 2\epsilon$. By Lemma 15.4.1, $p \le \frac{k}{1-2\epsilon}$.

Fix i and consider the event that v is removed when w_i is considered. This happens only if $d'_{v,w_i} - x_v \le X \le d'_{v,w_i}$. For fixed such X, a crucial observation is that if w_j with $j < i$ is considered before w_i, since $d'_{v,w_j} - x_v \le X$, v will be either removed or disconnected when w_j is considered. In particular, v will not be removed by w_i. Given these observations, the probability that v is removed is bounded by

$$\begin{aligned}
\Pr[v \text{ is removed}] &= \sum_{i=1}^{p} \Pr[v \text{ is removed when } w_i \text{ is considered}] \\
&= \sum_{i=1}^{p} \Pr[X \in [d'_{v,w_i} - x_v, d'_{v,w_i}] \text{ and } w_i \text{ comes before } w_1, \ldots, w_{i-1} \text{ in } \pi] \\
&\le \sum_{i=1}^{p} \frac{2x_v}{\epsilon i} = x_v \cdot O(\frac{\log p}{\epsilon}) = x_v \cdot O(\frac{\log k}{\epsilon}).
\end{aligned}$$

□

Therefore, the low-diameter decomposition removes at most $O(\frac{\log k}{\epsilon}) \cdot \sum_v x_v \le O(\frac{\log k}{\epsilon}) \cdot \mathsf{Opt}$ vertices so that each resulting connected component has at most $\frac{k}{1-2\epsilon}$ red vertices. This gives a bicriteria approximation algorithm that runs in time $\text{poly}(n, k)$, proving Theorem 15.3.4.

15.5 Algorithm for k-EDGE SEPARATOR

We present an $O(\log k)$-true approximation algorithm for k-EDGE SEPARATOR, proving Theorem 12.2.5. Except the cleanup step, the algorithm is almost identical to that of k-VERTEX SEPARATOR.

15.5.1 Spreading Metrics

Our relaxation for the edge version is very close to that of the vertex version. It has the following variables.

- x_e for $e \in E$: It indicates whether e is removed or not.
- $d_{u,v}$ for $(u, v) \in V \times V$: Given $\{x_e\}_{e \in E}$ as lengths on vertices, $d_{u,v}$ is supposed to be the minimum distance between u and v. Let $\mathcal{P}_{u,v}$ be the set of simple paths from u to v, and given $P = (u_0 := u, u_1, \ldots, u_p := v) \in \mathcal{P}_{u,v}$, let $d(P) = x_{u_0,u_1} + \cdots + x_{u_{p-1},u_p}$. Formally, we want
$$d_{u,v} = \min_{P \in \mathcal{P}_{u,v}} d(P).$$
Note that $d_{u,v} = d_{v,u}$ and $d_{u,u} = 0$.

- $f_{u,v}$ for all $(u,v) \in V \times V$: It indicates whether u and v belong to the same connected component or not.

Our LP is written as follows.

$$
\begin{array}{lllr}
\text{minimize} & \sum_{e \in E} x_e & & \\
\text{subject to} & d_{u,v} \le \min_{P \in \mathcal{P}_{u,v}} d(P) & \forall (u,v) \in V \times V & (15.7)\\
& f_{u,v} \ge 1 - d_{u,v} & \forall (u,v) \in V \times V & \\
& f_{u,v} \ge 0 & \forall (u,v) \in V \times V & \\
& \sum_{u \in V} f_{v,u} \le k-1 & \forall v \in V & (15.8)\\
& x_e \ge 0 & \forall e \in E & (15.9)
\end{array}
$$

(15.7) can be formally written as

$$
\begin{array}{ll}
d_{u,u} = 0 & \forall u \in V \\
d_{u,w} \le d_{u,v} + x_{v,w} & \forall (u,v) \in V \times V, (v,w) \in E
\end{array}
$$

Therefore, the size of our LP is polynomial in n. It is easy to verify that our LP is a relaxation — given a subset $S \subseteq E$ such that each connected component of $(V, E \setminus S)$ has at most k vertices, the following is a feasible solution with $\sum_e x_e = |S|$.

- $x_e = 1$ if $e \in S$. $x_e = 0$ if $v \notin S$.
- $d_{u,v} = \min_{P \in \mathcal{P}_{u,v}} d(P)$.
- $f_{u,v} = 1$ if u and v are in the same component of $(V, E \setminus S)$. Otherwise $f_{u,v} = 0$.

Fix an optimal solution $\{x_e\}_e, \{d_{u,v}, f_{u,v}\}_{u,v}$ for the above LP. It only ensures that $d_{u,v} \le \min_{P \in \mathcal{P}_{u,v}} d(P)$, so a priori $d_{u,v}$ can be strictly less than $\min_{P \in \mathcal{P}_{u,v}} d(P)$. However, in that case increasing $d_{u,v}$ still maintains feasibility, since larger $d_{u,v}$ provides a looser lower bound of $f_{u,v}$. For the subsequent sections, we assume that $d_{u,v} = \min_{P \in \mathcal{P}_{u,v}} d(P)$ for all u, v.

15.5.2 Low-diameter Decomposition

Fix $\epsilon \in (0, \frac{1}{2}]$. Given an optimal solution $\{x_e\}_{e \in E}$ and $\{d_{u,v}\}_{u,v \in V \times V}$ to the above LP, our low-diameter decomposition removes at most $O(\frac{\log k}{\epsilon}) \cdot \sum_e x_e$ edges so that each resulting connected component has at most $\frac{k}{1-\epsilon}$ vertices. It proceeds as follows.

- Pick $X \in [\epsilon/2, \epsilon]$ uniformly at random.
- Choose a random permutation $\pi : V \mapsto V$ uniformly at random.
- Consider the vertices one by one, in the order given by π. Let w be the considered vertex (we consider every vertex whether it was previously disconnected or not).
  - Let $W \leftarrow \emptyset$.
  - For each vertex $v \in V$ with $d_{w,v} \leq X$, if it is not disconnected yet, add it to W.
  - *Disconnect* W from the rest of the graph (i.e., remove every edge that has exactly one endpoint in W).

For each vertex w, let $B(w) := \{v : d_{w,v} \leq \epsilon\}$. A simple averaging argument bounds $|B(w)|$.

Lemma 15.5.1. *For each vertex w, $|B(w)| \leq \frac{k}{1-\epsilon}$.*

Proof. Assume towards contradiction that $|B(w)| > \frac{k}{1-\epsilon}$. For all $u \in B(w)$, $f_{w,u} \geq 1 - d_{w,u} \geq 1 - \epsilon$. Furthermore, even for $u \notin B(w)$, our LP ensures that $f_{w,u} \geq 0$. Therefore,

$$\sum_{u \in V} f_{w,u} \geq \sum_{u \in B(w)} f_{w,u} \geq (1-\epsilon)|B(w)| > k,$$

contradicting (15.8) of our LP. □

Note that at the end of the algorithm, every vertex is disconnected, since every $w \in V$ becomes disconnected after being considered. Moreover, each connected component is a subset of $\{v : d_{w,v} \leq X\}$ for some $w \in V$ and $X \leq \epsilon$, which is a subset of $B(w)$. Therefore, each connected component has at most $\frac{k}{1-\epsilon}$ vertices. We finally analyze the probability that an edge e is removed.

Lemma 15.5.2. *The probability that $e \in E$ is removed is at most $O(\frac{\log k}{\epsilon}) \cdot x_e$.*

Proof. Fix an edge $e = (u, v) \in E$. For a vertex $v \in W$, let $d^{\mathsf{near}}_{w,e} = \min(d_{w,u}, d_{w,v})$ and $d^{\mathsf{far}}_{w,e} = \max(d_{w,u}, d_{w,v})$. When $w \in V$ is considered, e can be possibly removed only if $d^{\mathsf{near}}_{w,e} \leq \epsilon \Rightarrow w \in B(v) \cup B(u)$. Let $W = \{w_1, \ldots, w_p\}$ be such vertices such that $d^{\mathsf{near}}_{w_1,e} \leq \cdots \leq d^{\mathsf{near}}_{w_p,e} \leq \epsilon$. By Lemma 15.5.1, $p \leq 2 \cdot \frac{k}{1-\epsilon}$.

Fix i and consider the event that e is removed when w_i is considered. This happens only if $d^{\mathsf{near}}_{w_i,e} \leq X \leq d^{\mathsf{far}}_{w_i,e}$. For fixed such X, a crucial observation is that if w_j with $j < i$

is considered before w_i, since $d^{\mathsf{near}}_{w_j,e} \le X$, e will be either removed (exactly one of u and v is disconnected) or disconnected (both u and v are disconnected) when w_j is considered. In particular, e will not be removed by w_i. Given these observations, the probability that e is removed is bounded by

$$\begin{aligned}
\Pr[e \text{ is removed}] &= \sum_{i=1}^{p} \Pr[e \text{ is removed when } w_i \text{ is considered}] \\
&= \sum_{i=1}^{p} \Pr[X \in [d^{\mathsf{near}}_{v,w_i}, d^{\mathsf{far}}_{v,w_i}] \text{ and } w_i \text{ comes before } w_1, \ldots, w_{i-1} \text{ in } \pi] \\
&\le \sum_{i=1}^{p} \frac{2x_e}{\epsilon i} = x_e \cdot O(\frac{\log p}{\epsilon}) = x_v \cdot O(\frac{\log k}{\epsilon}).
\end{aligned}$$

□

Therefore, the low-diameter decomposition removes at most $O(\frac{\log k}{\epsilon}) \cdot \sum_v x_v \le O(\frac{\log k}{\epsilon}) \cdot$ Opt edges so that each resulting connected component has at most $\frac{k}{1-\epsilon}$ vertices.

To get true approximation, we use the algorithm for BALANCED b-CUT. For an undirected graph $G = (V, E)$ with n vertices and a real $b \in (0, 1/2]$, the BALANCED b-CUT problem asks to find a subset $S \subseteq V$ with $bn \le |S| \le (1-b)n$ such that the number of edges that have exactly one endpoint in S is minimized. Racke [Rac08] gave an $O(\log n)$-true approximation algorithm for BALANCED b-CUT.[2]

We set $\epsilon = \frac{1}{3}$ such that each connected component after the low-diameter decomposition, each connected component has at most $\frac{3k}{2}$ vertices. Fix a component of size k'. If $k' \le k$, we are done. Otherwise, we use the $O(\log k') = O(\log k)$-approximation algorithm for BALANCED b-CUTg within the component. Usually k-EDGE SEPARATOR (requires many connected components) and BALANCED b-CUT (requires 2 connected components) behave very differently, but given $k' \le \frac{3k}{2}$, we show that they are equivalent.

Lemma 15.5.3. *In a graph $G = (V, E)$ with at most $k' \in (k, \frac{2}{3}k]$ vertices, the optimum solution of k-EDGE SEPARATOR and b-BALANCED CUT with $b = \frac{k'-k}{k'}$ are the same.*

[2]His algorithm is originally stated for MINIMUM BISECTION, the special case with $b = \frac{1}{2}$. For any $c \in [0, 1-2b]$, adding a disjoint clique with cn vertices and infinite-weight edges (his algorithm works in weighted version), forces the MINIMUM BISECTION algorithm to output a cut in the original graph where the smaller side contains exactly $\frac{(1-c)n}{2} \in [bn, \frac{n}{2}]$ vertices. Trying every value of $c \in [0, 1-2b]$ that makes cn an integer and taking the best cut gives the desired $O(\log n)$-true approximation for BALANCED b-CUT. We thank Anupam Gupta for this idea.

Proof. Any cut $(S, V \setminus S)$ feasible for b-BALANCED CUT ensures that $\max(|S|, |V \setminus S|)$ is at most $(1-b)k' = k$, so it is feasible for k-EDGE SEPARATOR.

For the other direction, given a feasible solution of k-EDGE SEPARATOR where V is partitioned into $S_1, \ldots, S_l$ (assume $k \geq |S_1| \geq \cdots \geq |S_l|$), if $l = 2$, (S_1, S_2) is a feasible solution for b-BALANCED CUT and we are done. If $l \geq 3$, merge S_{l-1}, S_l into one set (one S_i may contain multiple connected components). This reduces l by 1, and since $|S_{l-1}| + |S_l| \leq \frac{2}{l} \cdot k' \leq \frac{2}{3}k' \leq k$, maintains the invariant that $|S_i| \leq k$ for all i. Iterating until $l = 2$ gives a feasible solution for b-BALANCED CUT with the same number of edges cut. $\square$

Therefore, running the approximation algorithm b-BALANCED CUT for each component guarantees that we remove $O(\log k) \cdot \mathsf{Opt}$ additional edges and each component has at most k vertices. This proves Theorem 12.2.5.

Part V

Cut Problems

Chapter 16

Cut Problems Overview

16.1 Introduction

One of the most important implications of the Unique Games Conjecture (UGC, [Kho02b]) is the results of Khot et al. [KKMO07] and Raghavendra [Rag08], which say that for any maximum constraint satisfaction problem (MAX CSP), an integrality gap instance of the standard semidefinite programming (SDP) relaxation can be converted to the NP-hardness result with the same gap. These results initiated the study of beautiful connections between power of convex relaxations and hardness of approximation, from which surprising results for both subjects have been discovered.

While their results hold for problems in MAX CSP, the framework of *converting an integrality gap instance to hardness* has been successfully applied to covering and graph cut problems. For graph cut problems, Manokaran et al. [MNRS08] showed that for UNDIRECTED MULITWAY CUT and its generalizations, an integrality gap of the standard linear programming (LP) relaxation implies the hardness result assuming the UGC. Their result is further generalized by Ene et al. [EVW13] by formulating them as MIN CSP. In addition, Kumar et al. [KMTV11] studied STRICT CSP and showed the same phenomenon for the standard LP relaxation.

One of the limitations of the previous CSP-based transformations from LP gap instances to hard instances is based on the fact that they do not usually preserve the desired structure of the constraint hypergraph.[1] For example, consider the LENGTH-BOUNDED

[1] One of notable exceptions we are aware is the result of Guruswami et al. [GSS15], using Kumar et al. [KMTV11] to show that k-Uniform k-Partite Hypergraph Vertex Cover is hard to approximate within a factor $\frac{k}{2} - \epsilon$ for any $\epsilon > 0$.

EDGE CUT problem where the input consists of a graph $G = (V, E)$, two vertices $s, t \in V$, and a constant $l \in \mathbb{N}$, and the goal is to remove the fewest edges to ensure there is no path from s to t of length less than l. This problem can be viewed as a special case of HYPERGRAPH VERTEX COVER (HVC) by viewing each edge as a vertex of a hypergraph and creating a hyperedge for every s-t path of length less than l. While HVC is in turn a STRICT CSP, its integrality gap instance cannot be converted to hardness using Kumar et al. [KMTV11] as a black-box, since the set of hyperedges created in the resulting hard instance is not guaranteed to correspond to the set of short s-t paths of some graph.

For UNDIRECTED MULITWAY CUT, Manokaran et al. [MNRS08] bypassed this difficulty by using 2-ary constraints so that the resulting constraint hypergraph becomes a graph again. For UNDIRECTED NODE-WEIGHTED MULTIWAY CUT, Ene et al. [EVW13] used the equivalence to HYPERGRAPH MULTIWAY CUT [OFN12] so that the resulting hypergraph does not need to satisfy additional structure. These problems are then formulated as a MIN CSP by using many labels which are supposed to represent different connected components. However, these MIN CSP based techniques often require nontrivial problem-specific ideas and do not seem to be easily generalized to many other cut problems.

We study variants of the classical s-t cut problem in both directed and undirected graphs that have been actively studied. We prove the optimal hardness or the first superconstant hardness for them. See Section 16.2 for the definitions of the problems and our results. All our results are based on the general framework of converting an integrality gap instance to a *length-control dictatorship test*. The structure of our length-control dictatorship tests allows us to naturally convert an integrality gap instance for the basic LP for various cut problems to hardness based on the UGC. Section 16.3 provides more detailed intuition of this framework. We believe that our framework is general and will be useful to prove tight inapproximability of other cut problems.

16.2 Problems and Results

DIRECTED MULTICUT and DIRECTED MULTIWAY CUT. Given a directed graph and two vertices s and t, one of the most natural variants of s-t cut is to remove the fewest edges to ensure that there is no directed path from s to t and no directed path from t to s. This problem is known as s-t BICUT and admits the trivial 2-approximation algorithm by computing the minimum s-t cut and t-s cut.

DIRECTED MULTIWAY CUT is a generalization of s-t BICUT that has been actively studied. Given a directed graph with k terminals $s_1, \dots, s_k$, the goal is to remove the fewest number of edges such that there is no path from s_i to s_j for any $i \neq j$. DIRECTED MULTIWAY CUT also admits 2-approximation [NZ01, CM16]. If k is allowed to increase polynomially with n, there is a simple reduction from Vertex Cover that shows $(2-\epsilon)$-approximation is hard under the UGC [GVY94, KR08].

DIRECTED MULTIWAY CUT can be further generalized to DIRECTED MULTICUT. Given a directed graph with k source-sink pairs $(s_1, t_1), \ldots, (s_k, t_k)$, the goal is to remove the fewest number of edges such that there is no path from s_i to t_i for any i. Computing the minimum s_i-t_i cut for all i separately gives the trivial k-approximation algorithm. Chuzhoy and Khanna [CK09] showed DIRECTED MULTICUT is hard to approximate within a factor $2^{\Omega(\log^{1-\epsilon} n)} = 2^{\Omega(\log^{1-\epsilon} k)}$ when k is polynomially growing with n. Agarwal et al. [AAC07] showed $\tilde{O}(n^{\frac{11}{23}})$-approximation algorithm, which improves the trivial k-approximation when k is large.

Very recently, Chekuri and Madan [CM16] showed simple approximation-preserving reductions from DIRECTED MULTICUT with $k = 2$ to s-t BICUT (the other direction is trivially true), and (UNDIRECTED) NODE-WEIGHTED MULTIWAY CUT with $k = 4$ to s-t BICUT. Since NODE-WEIGHTED MULTIWAY CUT with $k = 4$ is hard to approximate within a factor $1.5 - \epsilon$ under the UGC [EVW13] (matching the algorithm of Garg et al. [GVY94]), the same hardness holds for s-t BICUT, DIRECTED MULTIWAY CUT, and DIRECTED MULTICUT for constant k. To the best of our knowledge, $1.5 - \epsilon$ is the best hardness factor for constant k even assuming the UGC. In the same paper, Chekuri and Madan [CM16] asked whether a factor $2 - \epsilon$ hardness holds for s-t BICUT under the UGC.

We prove that for any constant $k \geq 2$, the trivial k-approximation for DIRECTED MULTICUT might be optimal. Our result for $k = 2$ gives the optimal hardness result for s-t BICUT, answering the question of Chekuri and Madan.

Theorem 16.2.1. *Assuming the Unique Games Conjecture, for every $k \geq 2$ and $\epsilon > 0$,* DIRECTED MULTICUT *with k source-sink pairs is NP-hard to approximate within a factor $k - \epsilon$.*

Corollary 16.2.2. *Assuming the Unique Games Conjecture, for any $\epsilon > 0$, s-t* BICUT *is hard to approximate within a factor $2 - \epsilon$.*

Remark 16.2.3. *Chekuri and Madan [CM17] obtained an independent and different proof of Theorem 16.2.1. Indeed, they studied the approximability of* DIRECTED MULTICUT*(H) for a fixed* demand graph *H, and proved that when H is directed bipartite, an LP gap instance implies hardness based on the UGC, which proves Theorem 16.2.1 as a corollary. While their ideas are specialized for* DIRECTED MULTICUT *as the previous CSP-based approaches, our length-control dictatorship framework can be directly applied to their* *more general setting and give a simpler proof of their result. Section 17.2.1 proves their main result, Theorem 17.2.4.*

Bicuts. The hardness of s-t BICUT suggests that it may be hard to outperform a simple approximation algorithm that outputs the union of the min s-t cut and the min t-s cut. This strong hardness result also motivates the following question: *Can an algorithm do better if it can choose s and t?* Formally, in the global version of bicut, denoted EDGE BICUT, the goal is to find the smallest number of edges whose deletion ensures that there exist two distinct nodes s and t such that s cannot reach t and t cannot reach s in the resulting digraph.

The dichotomy between global cut problems and fixed-terminal cut problems in undirected graphs is well-known. For concreteness, recall EDGE 3-CUT and EDGE 3-WAY CUT. In EDGE 3-CUT, the input is an undirected graph and the goal is to find the smallest number of edges to delete so that the resulting graph has at least 3 connected components. In EDGE 3-WAY CUT, the input is an undirected graph with 3 specified nodes and the goal is to find the smallest number of edges to delete so that the resulting graph has at least 3 connected components with at most one of the 3 specified nodes in each component. While EDGE 3-WAY CUT is NP-hard [DJP+94], EDGE 3-CUT is solvable efficiently [GH94]. Similarly, while s-t BICUT is inapproximable to a factor better than 2 assuming UGC, EDGE BICUT is approximable within a factor of $2 - 1/448$ [BCK+17].

We also consider the problem between s-t BICUT and EDGE BICUT, denoted s-$*$ EDGE BICUT: Given a directed graph with a specified node s, find the smallest number of edges to delete so that there exists a node t such that s cannot reach t and t cannot reach s in the resulting graph. s-$*$ EDGE BICUT admits a 2-approximation by guessing the terminal t and then using the 2-approximation for s-t BICUT. We show the following inapproximability results for s-$*$ EDGE BICUT:

Theorem 16.2.1. s-$*$ EDGE BICUT *has no efficient* $(4/3-\epsilon)$*-approximation for any* $\epsilon > 0$ *assuming the Unique Games Conjecture.*

Furthermore, we consider the node-weighted variant of bicut, denoted NODE BICUT: Given a directed graph, find the smallest number of nodes whose deletion ensures that there exist nodes s and t such that s cannot reach t and t cannot reach s in the resulting graph. Every directed graph that is not a tournament has a feasible solution to NODE BICUT. NODE BICUT admits a 2-approximation by a simple reduction to s-t BICUT. We show the following inapproximability results.

Theorem 16.2.2. NODE BICUT *has no efficient* $(3/2 - \epsilon)$*-approximation for any* $\epsilon > 0$ *assuming the Unique Games Conjecture.*

Double Cuts. Recall that an arborescence in a directed graph $D = (V, E)$ is a minimal subset $F \subseteq E$ of arcs such that there exists a node $r \in V$ with every node $u \in V$ having a unique path from r to u in the subgraph (V, F) (e.g., see [Sch03]).

The input to the NODE DOUBLE CUT problem is a directed graph and the goal is to find the smallest number of nodes whose deletion ensures that the remaining graph has no arborescence. This problem is key to understanding fault tolerant consensus in networks [TV15].

A directed graph $D = (V, E)$ has no arborescence if and only if there exist two distinct nodes $s, t \in V$ such that every node $u \in V$ can reach at most one node in $\{s, t\}$ [BP13]. By this characterization, every directed graph that is not a tournament has a feasible solution to NODE DOUBLE CUT. This characterization motivates the following fixed-terminal version, denoted s-t NODE DOUBLE CUT: Given a directed graph with two specified nodes s and t, find the smallest number of nodes whose deletion ensures that every remaining node u can reach at most one node in $\{s, t\}$ in the resulting graph. An instance of s-t NODE DOUBLE CUT has a feasible solution provided that the instance has no edge between s and t. An efficient algorithm to solve/approximate s-t NODE DOUBLE CUT immediately gives an efficient algorithm to solve/approximate NODE DOUBLE CUT.

In the edge-weighted variation of two-terminal double cut, namely s-t EDGE DOUBLE CUT, the goal is to delete the smallest number of edges to ensure that every node in the graph can reach at most one node in $\{s, t\}$. Similarly, in the global variant, denoted EDGE DOUBLE CUT, the goal is to delete the smallest number of edges to ensure that there exist nodes s, t such that every node u can reach at most one node in $\{s, t\}$. Thus, EDGE DOUBLE CUT is equivalent to deleting the smallest number of edges to ensure that the graph has no arborescence. The fixed-terminal variant s-t EDGE DOUBLE CUT is solvable in polynomial time using maximum flow and, consequently, EDGE DOUBLE CUT is also solvable in polynomial time [BP13].

We show the following inapproximability results for s-t NODE DOUBLE CUT.

Theorem 16.2.3. s-t NODE DOUBLE CUT *has no efficient* $(2 - \epsilon)$*-approximation for any* $\epsilon > 0$ *assuming the Unique Games Conjecture.*

This matches a 2-approximation algorithm for s-t NODE DOUBLE CUT [BCK$^+$17], which also leads to a 2-approximation for NODE DOUBLE CUT. Note that the inapproximability results for s-t NODE DOUBLE CUT do not imply the hardness of NODE DOUBLE CUT. We also have the following inapproximability of NODE DOUBLE CUT.

Theorem 16.2.4. NODE DOUBLE CUT *has no efficient* $(3/2 - \epsilon)$*-approximation for any* $\epsilon > 0$ *assuming the Unique Games Conjecture.*

NODE k-CUT and VERTEX COVER ON k-PARTITE GRAPHS. Another way to show hardness of NODE DOUBLE CUT is a reduction from the node-weighted 3-cut problem in undirected graphs, though Theorem 16.2.4 shows a better hardness using length-control dictatorship tests and we do not show this reduction in this thesis ([BCK+17] presents this reduction to show an inapproximability result only assuming $\mathbf{P} \neq \mathbf{NP}$).

In the node weighted k-cut problem, denoted NODE 3-CUT, the input is an undirected graph and the goal is to find the smallest subset of nodes whose deletion leads to at least k connected components in the remaining graph. A classic result of Goldschmidt and Hochbaum [GH94] showed that the edge-weighted variant, denoted EDGE k-CUT (more commonly known as k-CUT)—namely find a smallest subset of edges of a given undirected graph whose deletion leads to at least k connected components—is solvable in polynomial time when k is a constant. Surprisingly, the complexity of NODE k-CUT for $k = 3$ is open. NODE k-CUT admits a $2(k-1)/k$-approximation algorithm [GVY04], and there is a simple approximation preserving reduction from VERTEX COVER ON k-PARTITE GRAPHS to NODE k-CUT. We prove that VERTEX COVER ON k-PARTITE GRAPHS is hard to approximate within a factor $2(k-1)/k$ assuming the UGC, so that $2(k-1)/k$ may be the optimal approximation factor for both VERTEX COVER ON k-PARTITE GRAPHS and NODE k-CUT.

Theorem 16.2.5. VERTEX COVER ON k-PARTITE GRAPHS *has no efficient* $(2(k-1)/k - \epsilon)$*-approximation algorithm for any* $\epsilon > 0$ *assuming the Unique Games Conjecture.*

Theorem 16.2.1 for s-$*$ EDGE BICUT and Theorem 16.2.2 for NODE BICUT follow from the above theorem as they are as hard to approximate as VERTEX COVER ON k-PARTITE GRAPHS for $k = 3$ and $k = 4$ respectively (see Section 16.5). We finally note that Theorem 16.2.5 is the only UG-hardness result in this part that does not require a length-control dictatorship test.

LENGTH-BOUNDED CUT and SHORTEST PATH INTERDICTION. The LENGTH-BOUNDED CUT problem is a natural variant of s-t cut, where given a graph (directed or undirected), $s, t \in V$, and an integer l, we only want to cut s-t paths of length strictly less than l.[2] Its practical motivation is based on the fact that in most communication / transportation networks, short paths are preferred to be used to long paths [MM10].

Lovász et al. [LNLP78] gave an exact algorithm for LENGTH-BOUNDED VERTEX CUT ($l \leq 5$) in undirected graphs. Mahjoub and McCormick [MM10] proved that LENGTH-

[2] It is more conventional to cut s-t paths of length *at most* l. We use this slightly nonconventional way to be more consistent with SHORTEST PATH INTERDICTION.

BOUNDED EDGE CUT admits an exact polynomial time algorithm for $l \leq 4$ in undirected graphs. Baier et al. [BEH+10] showed that both LENGTH-BOUNDED VERTEX CUT ($l > 5$) and LENGTH-BOUNDED EDGE CUT ($l > 4$) are NP-hard to approximate within a factor 1.1377. They presented $O(\min(l, \frac{n}{l})) = O(\sqrt{n})$-approximation algorithm for LENGTH-BOUNDED VERTEX CUT and $O(\min(l, \frac{n^2}{l^2}, \sqrt{m})) = O(n^{2/3})$-approximation algorithm for LENGTH-BOUNDED EDGE CUT, with matching LP gaps. LENGTH-BOUNDED CUT problems have been also actively studied in terms of their fixed parameter tractability [GT11, DK15, BNN15, FHNN15].

If we exchange the roles of the objective k and the length bound l, the problem becomes SHORTEST PATH INTERDICTION, where we want to maximize the length of the shortest s-t path after removing at most k vertices or edges. It is also one of the central problems in a broader class of *interdiction* problems, where an *attacker* tries to remove some edges or vertices to destroy a desirable property (e.g., short s-t distance, large s-t flow, cheap MST) of a network (see the survey of [SPG13]). The study of SHORTEST PATH INTERDICTION started in 1980's when the problem was called as the k MOST VITAL ARCS problem [CD82, MMG89, BGV89] and proved to be NP-hard [BGV89]. Khachiyan et al. [KBB+07] proved that it is NP-hard to approximate within a factor less than 2. While many heuristic algorithms were proposed [IW02, BB08, Mor11] and hardness in planar graphs [PS13] was shown, whether the general version admits a constant factor approximation was still unknown.

Given a graph $G = (V, E)$ and $s, t \in V$, let $\mathsf{dist}(G)$ be the length of the shortest s-t path. For $V' \subseteq V$, let $G \setminus V'$ be the subgraph induced by $V \setminus V'$. For $E' \subseteq E$, we use the same notation $G \setminus E'$ to denote the subgraph $(V, E \setminus E')$. We primarily study undirected graphs. We first present our results for the vertex version of both problems (collectively called as SHORT PATH VERTEX CUT onwards).

Theorem 16.2.4. *Assuming the Unique Games Conjecture, for infinitely many values of constant $l \in \mathbb{N}$, the following three tasks are NP-hard: Given an undirected graph $G = (V, E)$ and $s, t \in V$ where there exists $C^* \subseteq V \setminus \{s, t\}$ such that* $\mathsf{dist}(G \setminus C^*) \geq l$,

1. *Find $C \subseteq V \setminus \{s, t\}$ such that $|C| \leq \Omega(l) \cdot |C^*|$ and* $\mathsf{dist}(G \setminus C) \geq l$.
2. *Find $C \subseteq V \setminus \{s, t\}$ such that $|C| \leq |C^*|$ and* $\mathsf{dist}(G \setminus C) \geq O(\sqrt{l})$.
3. *Find $C \subseteq V \setminus \{s, t\}$ such that $|C| \leq \Omega(l^{\frac{\epsilon}{2}}) \cdot |C^*|$ and* $\mathsf{dist}(G \setminus C) \geq O(l^{\frac{1+\epsilon}{2}})$ *for some $0 < \epsilon < 1$.*

The first result shows that LENGTH-BOUNDED VERTEX CUT is hard to approximate within a factor $\Omega(l)$. This matches the best $O(l)$-approximation [BEH+10] when l is a constant. The second result shows that SHORTEST PATH VERTEX INTERDICTION is hard to approximate with in a factor $\Omega(\sqrt{\mathsf{Opt}})$, and the third result rules out *bicriteria approximation* — for any constant c, it is hard to approximate both l and $|C^*|$ within a factor of c.

The above results hold for directed graphs by definition. Our hard instances will have a natural *layered* structure, so it can be easily checked that the same results (up to a constant) hold for directed acyclic graphs. Since one vertex can be split as one directed edge, the same results hold for the edge version in directed acyclic graphs.

For LENGTH-BOUNDED EDGE CUT and SHORTEST PATH EDGE INTERDICTION in undirected graphs (collectively called SHORTEST PATH EDGE CUT onwards), we prove the following theorems.

Theorem 16.2.5. *Assuming the Unique Games Conjecture, for infinitely many values of constant $l \in \mathbb{N}$, the following three tasks are NP-hard: Given an undirected graph $G = (V, E)$ and $s, t \in V$ where there exists $C^* \subseteq E$ such that* $\mathsf{dist}(V \setminus C^*) \geq l$,

1. *Find $C \subseteq E$ such that $|C| \leq \Omega(\sqrt{l}) \cdot |C^*|$ and* $\mathsf{dist}(G \setminus C) \geq l$.
2. *Find $C \subseteq E$ such that $|C| \leq |C^*|$ and* $\mathsf{dist}(G \setminus C) \geq l^{\frac{2}{3}}$.
3. *Find $C \subseteq E$ such that $|C| \leq \Omega(l^{\frac{2\epsilon}{3}}) \cdot |C^*|$ and* $\mathsf{dist}(G \setminus C) \geq O(l^{\frac{2+2\epsilon}{3}})$ *for some $0 < \epsilon < \frac{1}{2}$.*

Our hardness factors for the undirected edge versions, $\Omega(\sqrt{l})$ for LENGTH-BOUNDED EDGE CUT and $\Omega(\sqrt[3]{\mathsf{Opt}})$ for SHORTEST PATH EDGE INTERDICTION, are slightly weaker than those for their vertex counterparts, but we are not aware of any approximation algorithm specialized for the undirected edge versions. It is an interesting open problem whether there exist better approximation algorithms for the undirected edge versions.

RMFC. RESOURCE MINIMIZATION FOR FIRE CONTAINMENT (RMFC) is a problem closely related to LENGTH-BOUNDED CUT with the additional notion of time. Given a graph G, a vertex s, and a subset T of vertices, consider the situation where fire starts at s on Day 0. For each Day i ($i \geq 1$), we can *save* at most k vertices, and the fire spreads from currently burning vertices to its unsaved neighbors. Once a vertex is burning or saved, it remains so from then onwards. The process is terminated when the fire cannot spread anymore. RMFC asks to find a strategy to save k vertices each day with the minimum k so that no vertex in T is burnt. These problems model the spread of epidemics or ideas through a social network, and have been actively studied recently [CC10, ACHS12, ABZ16, CV16].

RMFC, along with other variants, is first introduced by Hartnell [Har95]. Another well-studied variant is called the FIREFIGHTER problem, where we are only given $s \in V$ and want to maximize the number of vertices that are not burnt at the end. It is known to be NP-hard to approximate within a factor $n^{1-\epsilon}$ for any $\epsilon > 0$ [ACHS12]. King and MacGillivray [KM10] proved that RMFC is hard to approximate within a factor less than 2. Anshelevich et al. [ACHS12] presented an $O(\sqrt{n})$-approximation algorithm for general graphs, and Chalermsook and Chuzhoy [CC10] showed that RMFC admits $O(\log^* n)$-approximation in trees. Very recently, the approximation ratio in trees has been improved to $O(1)$ [ABZ16]. Both Anshelevich et al. [ACHS12] and Chalermsook and Chuzhoy [CC10] independently studied directed layer graphs with b layers, showing $O(\log b)$-approximation.

Our final result on RMFC assumes Conjecture 3.2.4, a variant of the Unique Games Conjecture which is not known to be equivalent to the original UGC. Given a bipartite graph as an instance of Unique Games, it states that in the completeness case, all constraints incident on $(1-\epsilon)$ fraction of vertices in one side are satisfied, and in the soundness case, in addition to having a low value, every $\frac{1}{10}$ fraction of vertices on one side have at least a $\frac{9}{10}$ fraction of vertices on the other side as neighbors. Our conjecture is implied by the conjecture of Bansal and Khot [BK09] that is used to prove the hardness of MINIMIZING WEIGHTED COMPLETION TIME WITH PRECEDENCE CONSTRAINTS and requires a more strict expansion condition.

Theorem 16.2.6. *Assuming Conjecture 3.2.4, it is NP-hard to approximate* RMFC *in undirected graphs within any constant factor.*

Again, our reduction has a natural layered structure and the result holds for directed layered graphs. With b layers, we prove that it is hard to approximate with in a factor $\Omega(\log b)$, matching the best approximation algorithms [CC10, ACHS12].

Table 16.1 summarizes our results.

16.3 Techniques

All our results are based on a general method of *converting an integrality gap instance to a dictatorship test*. This method has been successfully applied by Raghavendra [Rag08] for MAX CSP, Manokaran et al. [MNRS08] and Ene et al. [EVW13] for Multiway Cut and MIN CSP, and Kumar et al. [KMTV11] for STRICT CSP, and by Guruswami et al. [GSS15] for k-uniform k-partite HYPERGRAPH VERTEX COVER, and Chekuri and

Promises	Algorithm	Hardness
DIRECTED MULTICUT	k	$k - \epsilon$
s-t BICUT	2	$2 - \epsilon$
s-$*$ EDGE BICUT	2	$4/3 - \epsilon$
NODE BICUT	2	$3/2 - \epsilon$
s-t NODE DOUBLE CUT	2	$2 - \epsilon$
NODE DOUBLE CUT	2	$3/2 - \epsilon$
LENGTH-BOUNDED VERTEX CUT	$O(l)$	$\Omega(l)$
SHORTEST PATH VERTEX INTERDICTION		$\Omega(\sqrt{\mathsf{Opt}})$
LENGTH-BOUNDED EDGE CUT	$O(l)$	$\Omega(\sqrt{l})$
SHORTEST PATH EDGE INTERDICTION		$\Omega(\sqrt[3]{\mathsf{Opt}})$
NODE k-CUT/ VERTEX COVER ON k-PARTITE GRAPHS	$2(k-1)/k$	$2(k-1)/k - \epsilon$
RMFC	$\sqrt{n}$	$\omega(1)$

Table 16.1: Summary of our hardness results for various cut problems.

Madan [CM17] for DIRECTED MULTICUT. As mentioned in the introduction, the previous CSP-based results do not generally preserve the structure of constraint hypergraphs or use ingenious and specialized tricks to reduce the problem to a CSP.

We bypass this difficulty by constructing a special class of dictatorship tests that we call *length-control dictatorship tests*. Consider a meta-problem where given a directed graph $G = (V, E)$, some terminal vertices, and a set $\mathcal{P}$ of *desired paths* between terminals, we want to remove the fewest number of non-terminal vertices to cut every path in $\mathcal{P}$. The integrality gap instances we use in this part [SSZ04, BEH+10, MM10, CC10] share the common feature that every $p \in \mathcal{P}$ is of length at least r, and the fractional solution cuts $\frac{1}{r}$ fraction of each non-terminal vertex so that each path $p \in \mathcal{P}$ is cut. This gives a good LP value, and additional arguments are required to ensure that there is no efficient integral cut.

Given such an integrality gap instance, we construct our dictatorship test instance as follows. We replace every non-terminal vertex by a *hypercube* $\mathbb{Z}_r^R$ and put edges such that for two vertices (v, x) and (w, y) where $v, w \in V$ and $x, y \in \mathbb{Z}_r^R$, there is an edge from (v, x) to (w, y) if (1) $(v, w) \in E$ and (2) $y_j = x_j + 1$ for all $j \in [R]$. The set of desired paths $\mathcal{P}'$ is defined to be $\{(s, (v_1, x_1), \ldots, (v_l, x_l), t) : (s, v_1, \ldots, v_l, t) \in \mathcal{P}\}$ (s, t denote some terminals). Note that each path in $\mathcal{P}'$ is also of length at least r. We want to ensure that in the *completeness case* (i.e., every hypercube reveals the same *influential coordinate*), there is a very efficient cut, while in the *soundness case* (i.e., no hypercube reveals an influential coordinate), there is no such efficient cut.

In the completeness case, let $q \in [R]$ be an influential coordinate. For each vertex (v, x) where $v \in V, x \in \mathbb{Z}_r^R$, remove (v, x) if $x_q = 0$. Consider a desired path $p = (s, (v_1, x_1), \ldots, (v_l, x_l), t) \in \mathcal{P}'$ for some terminals s, t and some $v_j \in V, x_j \in \mathbb{Z}_r^R$ $(1 \leq j \leq l)$, and let $y_j = (x_j)_q$. By our construction, $y_{j+1} = y_j + 1$ for $0 \leq j < l$. Since p is desirable, $l \geq r$, so there exists j such that $y_j = (x_j)_q = 0$, but (v_j, x_j) is already removed by our previous definition. Therefore, every desired path is cut by this vertex cut. Note that this cut is integral and cuts exactly $\frac{1}{r}$ fraction of non-terminal vertices. This corresponds to the fractional solution to the gap instance that cuts $\frac{1}{r}$ fraction of every vertex.

For the soundness analysis, our final dictatorship test has additional *noise* vertices and edges to the test defined above. If no hypercube reveals an influential coordinate, the standard application of the invariance principle [Mos10] proves that we can always take an edge between two hypercubes unless we almost completely cut one hypercube. We can then invoke the proof for the integrality gap instance to show that there is no efficient cut.

This idea is implicitly introduced by the work of Svensson [Sve13] for Feedback Vertex Set (FVS) and DAG Vertex Deletion (DVD) by applying the *It ain't over till it's over theorem* to ingeniously constructed dictatorship tests with auxiliary vertices. Guruswami and Lee [GL16c] gave a simpler construction and a new proof using the invariance principle instead of the It ain't over till it's over theorem. Our results are based on the observation that length-control dictatorship tests and LP gap instances *fool* algorithms in a similar way for various cut problems as mentioned above, so that the previous LP gap instances can be plugged into our framework to prove matching hardness results.

This method for the above meta-problem can be almost directly applied to DIRECTED MULTICUT. For LENGTH-BOUNDED CUT and RMFC in undirected graphs, we use the fact that the known integrality gap instances have a natural layered structure with s in the first layer and t in the last layer. Every edge is given a natural orientation, and the similar analysis can be applied. For LENGTH-BOUNDED CUT, another set of edges called *long edges* are added to the dictatorship test. More technical work is required for edge cut versions in undirected graphs (SHORT PATH EDGE CUT), and the notion of time (RMFC).

Our framework seems general enough so that they can be applied to integrality gap instances to give strong hardness results. It would be interesting to further abstract this method of converting integrality gap instances to length-bounded dictatorship tests, as well as to apply it to other problems whose approximability is not well-understood.

16.4 Organization.

Section 16.5 shows that s-$*$ EDGE BICUT and NODE BICUT are as hard to approximate as VERTEX COVER ON k-PARTITE GRAPHS with $k = 3$ and $k = 4$ respectively. Chapter 17 presents our dictatorship tests for the problem mentioned in this overview. Except for VERTEX COVER ON k-PARTITE GRAPHS, all dictatorship tests are length-control dictatorship tests. Chapter 18 shows how to use these tests to prove hardness results based on the UGC.

16.5 Combinatorial Reductions

Lemma 16.5.1. *There is an approximation-preserving reduction from* VERTEX COVER ON 4-PARTITE GRAPHS *to* NODE BICUT.

Proof. Given a 4-partite graph $G = (V_1 \cup V_2 \cup V_3 \cup V_4, E)$, we construct an instance $D = (V_D, A_D)$ for NODE BICUT as follows: Let $V_D := V_1 \cup V_2 \cup V_3 \cup V_4 \cup \{s, t\}$. The set of arcs A_D are obtained as follows:

1. For every $u, v \in V_i$ for some $i \in [4]$, we add a bidirected arc between u and v.

2. For every $(u, v) \in E$, we add a bidirected arc between u and v.

3. For every $u \in V_1$, we add a bidirected arc between s and u.

4. For every $u \in V_2$, we add an arc from s to u and an arc from t to u.

5. For every $u \in V_3$, we add an arc from u to s and an arc from u to t.

6. For every $u \in V_4$, we add a bidirected arc between t and u.

We now show the completeness of the reduction. Suppose $R \subseteq V_1 \cup V_2 \cup V_3 \cup V_4$ is a vertex cover in G. Then $D - R$ has no $s \to t$ path, since s can only reach vertices in V_1 and V_2, only vertices in V_3 and V_4 can reach t, and there is no arc between V_i and V_j for any $i \neq j$. Similarly, there is no $t \to s$ path. Therefore, R is a feasible solution to NODE BICUT in D.

Next we show soundness of the reduction. Suppose $R \subseteq V_1 \cup V_2 \cup V_3 \cup V_4$ is a feasible solution to NODE BICUT in D. There exists two vertices $u, v \in V_D \setminus R$ such that there is no $u \to v$ path and no $v \to u$ path in the subgraph of D induced by $V_D \setminus R$. We note that

v and u cannot be in the same V_i since V_i is a clique in V_D. We also rule out the following cases:

1. If $v \in V_1, u \in V_2$, then (v, s, u) is a path from v to u, a contradiction.
2. If $v \in V_1, u \in V_3$, then (u, s, v) is a path from u to v, a contradiction.
3. If $v \in V_2, u \in V_3$, then (u, s, v) is a path from u to v, a contradiction.
4. If $v \in V_2, u \in V_4$, then (u, t, v) is a path from u to v, a contradiction.
5. If $v \in V_3, u \in V_4$, then (v, t, u) is a path from v to u, a contradiction.

Thus, $v \in V_1$ and $u \in V_4$. We will show that if R is not a vertex cover, then there is a $v \to u$ path or $u \to v$ path, a contradiction. Suppose there exists $\{a, b\} \in E$ such that $a, b \notin R$.

1. If $a \in V_1, b \in V_2$, then (u, t, b, a, v) is a path from u to v, a contradiction.
2. If $a \in V_1, b \in V_3$, then (v, a, b, t, u) is a path from v to u, a contradiction.
3. If $a \in V_1, b \in V_4$, then (v, a, b, u) is a path from v to u, a contradiction.
4. If $a \in V_2, b \in V_3$, then (v, s, a, b, t, u) is a path from v to u, a contradiction.
5. If $a \in V_2, b \in V_4$, then (v, s, a, b, u) is a path from v to u, a contradiction.
6. If $a \in V_3, b \in V_4$, then (u, b, a, s, v) is a path from u to v, a contradiction.

Therefore, R must be a vertex cover. This establishes the soundness of the reduction and completes the proof. $\square$

Lemma 16.5.2. *There is an approximation-preserving reduction from* VERTEX COVER ON 3-PARTITE GRAPHS *to* s-$*$ EDGE BICUT.

Proof. Given a 3-partite graph $G = (A \cup B \cup C, E)$, we construct an instance $D = (V_D, A_D)$ for s-$*$ EDGE BICUT as follows: Let $V_D := A_1 \cup A_2 \cup B_1 \cup B_2 \cup C_1 \cup C_2 \cup \{s, t\}$. For a vertex $v \in A \cup B \cup C$ and $i \in \{1, 2\}$, let v_i denote the corresponding vertex in V (e.g., if $v \in A$, then $v_1 \in A_1$ and $v_2 \in A_2$). We introduce three types of arcs in A_D.

1. Vertex arcs: For every $v \in A \cup B \cup C$, create an arc (v_1, v_2) with weight 1.

2. Forward arcs: Create arcs with weight ∞

   (a) (s, a_1) for all $a \in A$, (s, b_1) for all $b \in B$, (b_2, s) for all $b \in B$, (c_2, s) for all $c \in C$.

   (b) (t, a_1) for all $a \in A$, (c_2, t) for all $c \in C$.

   (c) (a_2, b_1) for every $\{a, b\} \in E$, $a \in A, b \in B$ (call them AB arcs), (a_2, c_1) for every $\{a, c\} \in E$, $a \in A, c \in C$ (call them AC arcs), (b_2, c_1) for every $\{b, c\} \in E$, $b \in B, c \in C$ (call them BC arcs).

3. Backward arcs: Create arcs with weight ∞

   (a) (v_2, u_1) for all $u, v \in A$ (call them AA arcs), (v_2, u_1) for all $u, v \in C$ (call them CC arcs), (c_1, a_1) for all $a \in A, c \in C$ (call them CA_1 arcs), (c_2, a_2) for all $a \in A, c \in C$ (call them CA_2 arcs).

We first show completeness of the reduction. Suppose $R \subseteq A \cup B \cup C$ is a vertex cover in G. Let $F = \{(v_1, v_2) : v \in R\}$. We will show that there is no $s \to t$ path and no $t \to s$ path in $D - F$.

1. Suppose there is a $t \to s$ path in $D - F$. Fix the shortest such $t \to s$ path P. Then, the path P has the following properties:

   (a) Path P does not contain AA arcs or CA_1 arcs, since t has direct arcs to vertices in A_1. Similarly, P does not contain CC arcs or CA_2 arcs, since vertices in C_2 have direct arcs to s. So, P does not contain any backward arcs.

   (b) Path P does not contain BC arcs, since vertices in B_2 have direct arcs to s.

   Thus, the only possibility for the path P is $P = (t, a_1, a_2, v_1, v_2, s)$ for $a \in A$, $v \in B \cup C$, and $\{a, v\} \in E$. This contradicts that R is a vertex cover.

2. Suppose there is a $s \to t$ path in $D - F$. Fix the shortest such $t \to s$ path P. Then, the path P has the following properties:

   (a) Path P does not contain AA arcs or CA_1 arcs, since s has direct arcs to vertices in A_1. Similarly, P does not contain CC arcs or CA_2 arcs since vertices in C_2 have direct arcs t. So, P does not contain any backward arcs.

   (b) Path P does not contain AB arcs, since s has direct arcs to vertices in B_1.

   Thus, the only possibility for the path P is $P = (t, v_1, v_2, c_1, c_2, s)$ for $v \in A \cup B$, $c \in C$, and $\{v, c\} \in E$. This contradicts that R is a vertex cover.

Therefore, s and t cannot reach each other in $D - F$. Consequently, the existence of a vertex cover R in G implies the existence of a feasible solution to s-$*$ EDGE BICUT in D of the same size.

Next we show soundness of the reduction. Suppose $F \subseteq E_D$ is a feasible solution to s-$*$ EDGE BICUT in D. Let $R \subseteq A \cup B \cup C$ be the set of vertices whose vertex arcs are in F. We will show that if R is not a vertex cover in G, then every vertex $v \in V_D$ has either a path to s or a path from s. Since vertices in A_1, B_1, B_2, C_2 have a direct arc either from or to s, we only need to check vertices in A_2, C_1 and t. We verify these cases below:

1. Suppose there exist $a \in A \setminus R, b \in B \setminus R$ such that $\{a, b\} \in E$.
   - (i) Considering t, we have $(t, a_1, a_2, b_1, b_2, s)$ as a path from t to s.
   - (ii) For every $a' \in A_2$, we have $(a', a_1, a_2, b_1, b_2, s)$ as a path from a' to s.
   - (iii) For every $c' \in C_1$, we have $(c', a_1, a_2, b_1, b_2, s)$ as a path from c' to s.
2. Suppose there exist $a \in A \setminus R, c \in C \setminus R$ such that $\{a, c\} \in E$.
   - (i) Considering t, we have $(t, a_1, a_2, c_1, c_2, s)$ as a path from t to s.
   - (ii) For every $a' \in A_2$, we have $(a', a_1, a_2, c_1, c_2, s)$ as a path from a' to s.
   - (iii) For every $c' \in C_1$, we have $(c', a_1, a_2, c_1, c_2, s)$ as a path from c' to s.
3. Suppose there exist $b \in B \setminus R, c \in C \setminus R$ such that $\{b, c\} \in E$.
   - (i) Considering t, we have $(s, b_1, b_2, c_1, c_2, t)$ as a path from s to t.
   - (ii) For every $a' \in A_2$, we have $(s, b_1, b_2, c_1, c_2, a')$ as a path from s to a'.
   - (iii) For every $c' \in C_1$, we have $(s, b_1, b_2, c_1, c_2, c')$ as a path from s to c'.

Therefore, the existence of a feasible solution to s-$*$ EDGE BICUT in D implies the existence of a vertex cover in G of the same size. This establishes the soundness of the reduction, and proves the lemma. $\square$

Chapter 17

Dictatorship Tests for Cut Problems

17.1 Preliminaries

Graph Terminologies. Depending on whether we cut vertices or edges, we introduce weight $\mathsf{wt}(v)$ for each vertex v, or weight $\mathsf{wt}(e)$ for each edge e. Some weights can be ∞, which means that some vertices or edges cannot be cut. For vertex-weighted graphs, we naturally have $\mathsf{wt}(s) = \mathsf{wt}(t) = \infty$. To reduce the vertex-weighted version to the unweighted version, we duplicate each vertex according to its weight and replace each edge by a complete bipartite graph between corresponding copies. To reduce the edge-weighted version to the unweighted version, we replace a single edge with parallel edges according to its weight. To reduce to simple graphs, we split each parallel into two edges by introducing a new vertex.

For the LENGTH-BOUNDED CUT problems, we also introduce length $\mathsf{len}(e)$ for each edge e. It can be also dealt with serially splitting an edge according to its weight. We allow weights to be rational numbers, but as our hardness results are stated in terms of the length, all lengths in this chapter will be a positive integer.

For a path p, depending on the context, we abuse notation and interpret it as a set of edges or a set of vertices. The length of p is always defined to be the number of edges.

17.2 DIRECTED MULTICUT

We propose our dictatorship test for DIRECTED VERTEX MULTICUT that will be used for proving Unique Games hardness. Note that hardness of DIRECTED EDGE MULTICUT easily follows from that of the vertex version by splitting each vertex. Our dictatorship test is inspired by the integrality gap for the standard LP constructed by Saks et al. [SSZ04], and parameterized by positive integers r, k, R and small $\epsilon > 0$, where k in this section denotes the number of (s_i, t_i) pairs for DIRECTED MULTICUT. All graphs in this section are directed.

For positive integers r, k, R, and $\epsilon > 0$, define $\mathcal{D}^{\mathsf{M}}_{r,k,R,\epsilon} = (V, E)$ be the graph defined as follows. Consider the probability space (Ω, μ) where $\Omega := \{0, \ldots, r-1, *\}$, and $\mu : \Omega \mapsto [0, 1]$ with $\mu(*) = \epsilon$ and $\mu(x) = \frac{1-\epsilon}{r}$ for $x \neq *$.

- $V = \{s_i, t_i\}_{1 \leq i \leq k} \cup \{v^\alpha_x\}_{\alpha \in [r]^k, x \in \Omega^R}$. Let v^α denote the set of vertices $\{v^\alpha_x\}_{x \in \Omega^R}$.
- For $\alpha \in [r]^k$ and $x \in \Omega^R$, $\mathsf{wt}(v^\alpha_x) = \mu^{\otimes R}(x)$. Note that the sum of weights is r^k.
- For any $i \in [k]$, there are edges from s_i to $\{v^\alpha_x : \alpha \in [r]^k, \alpha_i = 1, x \in \Omega^R\}$, and edges from $\{v^\alpha_x : \alpha \in [r]^k, \alpha_i = r, x \in \Omega^R\}$ to t_i.
- For $\alpha, \beta \in [r]^k$ and $x, y \in \Omega^R$, we have an edge from v^α_x to v^β_y if $\alpha \neq \beta$ and
  - For any $1 \leq i \leq r$: $\alpha_i - \beta_i \in \{-1, 0, +1\}$.
  - For any $1 \leq j \leq R$: $[y_j = (x_j + 1) \mod r]$ or $[y_j = *]$ or $[x_j = *]$.

Completeness. We first prove that vertex cuts that correspond to *dictators* behave the same as the fractional solution that gives $\frac{1}{r}$ to every vertex. For any $q \in [R]$, let $V_q := \{v^\alpha_x : \alpha \in [r]^k, x_q = * \text{ or } 0\}$. Note that the total weight of V_q is $r^k(\epsilon + \frac{1-\epsilon}{r}) \leq r^{k-1}(1 + \epsilon r)$.

Lemma 17.2.1. *After removing vertices in V_q, there is no path from s_i to t_i for any i.*

Proof. Fix i and let $p = (s_i, v^{\alpha_1}_{x^1}, \ldots, v^{\alpha_z}_{x^z}, t_i)$ be a path from s_i to t_i where $\alpha_j \in [r]^k$ and $x^j \in \Omega^R$ for each $1 \leq j \leq z$. Let $y_j := (x^j)_q$ for each $1 \leq j \leq z$. The construction ensures that $y_{j+1} = (y_j + 1) \mod r$, so after removing vertices in V_q, z must be strictly less than r. Since any path from s_i to t_i must contain at least r non-terminal vertices, there must be no path from s_i to t_i. $\square$

Soundness. To analyze soundness, we define a correlated probability space $(\Omega_1 \times \Omega_2, \nu)$ where both Ω_1, Ω_2 are copies of $\Omega = \{0, \ldots, r-1, *\}$. It is defined by the following process to sample $(x, y) \in \Omega^2$.

- Sample $x \in \{0, \ldots, r-1\}$. Let $y = (x + 1) \mod r$.
- Change x to $*$ with probability ϵ. Do the same for y independently.

Note that the marginal distribution of both x and y is equal to μ. Assuming $\epsilon < \frac{1}{2r}$, the minimum probability of any atom in $\Omega_1 \times \Omega_2$ is ϵ^2. In our correlated space, the bipartite graph on $\Omega_1 \cup \Omega_2$ is connected since every $x \in \Omega_1$ is connected to $* \in \Omega_2$ and vice versa. Therefore, we can apply Lemma 3.3.3 to conclude that $\rho(\Omega_1, \Omega_2; \nu) \leq \rho := 1 - \frac{\epsilon^4}{2}$.

Apply Theorem 3.3.10 ($\rho \leftarrow \rho, \alpha \leftarrow \epsilon^2, \epsilon \leftarrow \frac{\Gamma_\rho(\frac{\epsilon}{3}, \frac{\epsilon}{3})}{2}$) to get τ and d. We will later apply this theorem with the parameters obtained here. Fix an arbitrary subset $C \subseteq V$, and let $C_\alpha := C \cap v^\alpha$. For $\alpha \in [r]^k$, call v^α *blocked* if $\mu^{\otimes R}(C_\alpha) \geq 1 - \epsilon$. The number of blocked v^α's is at most $\frac{\mathsf{wt}(C)}{1-\epsilon}$.

Consider the following graph $D = (V_D, E_D)$, which is the original integrality gap instance constructed by Saks et al. [SSZ04].

- $V_D = \{s_i, t_i\}_{i \in [k]} \cup \{v^\alpha\}_{\alpha \in [r]^k}$.
- For any $i \in [k]$, there are edges from s_i to $\{v^\alpha : \alpha \in [r]^k, \alpha_i = 1\}$, and edges from $\{v^\alpha : \alpha \in [r]^k, \alpha_i = r\}$ to t_i.
- For $\alpha, \beta \in [r]^k$, we have an edge from v^α to v^β if $\alpha \neq \beta$ and $1 \leq i \leq r$: $\alpha_i - \beta_i \in \{-1, 0, +1\}$.

Saks et al. [SSZ04] proved the following theorem in their analysis of their integrality gap.

Theorem 17.2.2. *Let C' be a set of less than $k(r-1)^{k-1}$ vertices. There exists a path $(s_i, v^{\alpha_1}, \ldots, v^{\alpha_z}, t_i)$ for some i that does not intersect C'.*

Setting $C' := \{v^\alpha \in V_D : v^\alpha \text{ is not blocked.}\}$, and applying Theorem 17.2.2 concludes that unless $\mathsf{wt}(C) \geq (1-\epsilon) \cdot k \cdot (r-1)^{k-1}$, there exists a path $(s_i, v^{\alpha_1}, \ldots, v^{\alpha_z}, t_i)$ where each v^{α_i} is unblocked for each $i \in [k]$.

For $1 \leq j \leq z$, let $S_j \subseteq v^{\alpha_j}$ be such that $x \in S_j$ if there exists a path $(s_i, v^{\alpha_1}_{x^1}, \ldots, v^{\alpha_{j-1}}_{x^{j-1}}, v^{\alpha_j}_x)$ for some $x^1, \ldots, x^{j-1}$. For $1 \leq j \leq z$, let $f_j : \Omega^R \mapsto \{0, 1\}$ be the indicator function of S_j. We prove that if none of f_j reveals any *influential coordinate*, $\mu^{\otimes R}(S_z) > 0$, which shows that there exists a s_i-t_i path even after removing vertices in C.

Lemma 17.2.3. *Suppose that for any $1 \leq j \leq z$ and $1 \leq i \leq R$, $\mathsf{Inf}_i^{\leq d}[f_j] \leq \tau$. Then $\mu^{\otimes R}(S_z) > 0$.*

Proof. We prove by induction that $\mu^{\otimes R}(S_j) \geq \frac{\epsilon}{3}$. It holds when $j = 1$ since v^{α_1} is unblocked. Assuming $\mu^{\otimes R}(S_j) \geq \frac{\epsilon}{3}$, since S_j does not reveal any influential coordinate, Theorem 3.3.10 shows that for any subset $T_{j+1} \subseteq v^{\alpha_{j+1}}$ with $\mu^{\otimes R}(T_{j+1}) \geq \frac{\epsilon}{3}$, there exists an edge from S_j and T_{j+1}. If $S'_{j+1} \subseteq v^{\alpha_{j+1}}$ is the set of out-neighbors of S_j, we have $\mu^{\otimes R}(S'_{j+1}) \geq 1 - \frac{\epsilon}{3}$. Since $v^{\alpha_{j+1}}$ is unblocked, $\mu^{\otimes R}(S'_{j+1} \setminus C) \geq \frac{2\epsilon}{3}$, completing the induction. □

In summary, in the completeness case, if we cut vertices of total weight $r^{k-1}(1+\epsilon r)$, we cut every s_i-t_i pair. In the soundness case, unless we cut vertices of total weight at least $(1-\epsilon) \cdot k \cdot (r-1)^{k-1}$, we cannot cut every s_i-t_i pair. The gap is $\frac{k(1-\epsilon)(r-1)^{k-1}}{(1+\epsilon r)r^{k-1}}$. For a fixed k, increasing r and decreasing ϵ faster makes the gap arbitrarily close to k.

17.2.1 DIRECTED MULTICUT with a Fixed Demand Graph

Let $H = (V_H, E_H)$ a fixed directed graph. In DIRECTED MULTICUT(H), we are given directed supply graph $G = (V_G, E_G)$ and an injective map $h : V_H \to V_G$, and the goal is to remove the smallest number of edges in E_G such that for every $(u, v) \in E_H$, there is no path from $h(u)$ to $h(v)$ in G. Chekuri and Madan [CM17] studied the relationship between approximation algorithms and the LP gaps of DIRECTED MULTICUT(H) for each fixed graph H. Their main result is the following theorem. Let α_H be the worst LP gap over all instances with demand graph H.

Theorem 17.2.4 ([CM17]). *Assuming the UGC, for any fixed directed bipartite graph H, and for any fixed $\epsilon > 0$, there is no polynomial-time $(\alpha_H - \epsilon)$ approximation for* DIRECTED MULTICUT*(H).*

Since every directed graph on k vertices can be decomposed into the union of at most $2\lceil \log k \rceil$ directed bipartite graphs, they also showed that for every fixed demand graph H with k vertices, DIRECTED MULTICUT(H) does not admit $(\frac{\alpha_H}{2\lceil \log k \rceil} - \epsilon)$-approximation under the UGC.

We provide a simpler proof of Theorem 17.2.4 using our length-control dictatorship framework. Let $H = (V_H, E_H)$ be a fixed directed bipartite graph. Let V_H^S and V_H^T be the set of source vertices and sink vertices respectively ($V_H = V_H^S \cup V_H^T$). Let DIRECTED VERTEX MULTICUT(H) be the problem where given $G = (V_G, E_G)$ and injective $h : V_H \to V_G$ with the promise that all vertices in $h(V_H^S)$ (resp. $h(V_H^T)$) are source (resp. sink) vertices in G, the goal is to remove the smallest number of non-terminal vertices (i.e., vertices in $V_G \setminus h(V_H)$) such that for every $(u, v) \in E_H$, there is no path from $h(u)$ to $h(v)$ in G. We prove the following two lemmas, based on well-known reductions between the edge and vertex versions of DIRECTED MULTICUT, to show that Theorem 17.2.4 for DIRECTED VERTEX MULTICUT(H) implies Theorem 17.2.4 for DIRECTED MULTICUT(H).

Lemma 17.2.5. *An LP gap of α for* DIRECTED MULTICUT*(H) implies that an LP gap of* DIRECTED VERTEX MULTICUT*(H) is at least α.*

Proof. Let $G = (V_G, E_G)$ and $h : V_H \to V_G$ be an LP gap instance for DIRECTED MULTICUT(H), and Opt and FRAC be the integral and fractional optimal value respectively for the instance (G, h). We construct an instance $G' = (V'_G, E'_G)$ and $h' : V_H \to V'_G$ of DIRECTED VERTEX MULTICUT(H) as follows.

- $V'_G := h(V_H) \cup E_G$ and $h' := h$.
- For each edge $(u, v) \in G$ and $(v, w) \in G$ with $u \neq w$, add an edge from (u, v) to (v, w) in G'.
- For each $s \in h(V_H^S)$ and an edge (s, u), we add an edge from s to (s, u) in G'.
- For each $t \in h(V_H^T)$ and an edge (u, t), we add an edge from (u, t) to t in G'.

Let Opt' and FRAC' be the integral and fractional optimal value respectively for the reduced instance (G', h') of DIRECTED VERTEX MULTICUT(H). Note that in G', vertices in $h'(V_H^S)$ are sources and vertices in $h'(V_H^T)$ are sinks. There is an one-to-one correspondence between edges of G and non-terminal vertices of G'.

We first claim $\mathsf{Opt}' \leq \mathsf{Opt}$. Let $F \subseteq E_G$ be the optimal solution for (G, h). Then F as a subset of V'_G is a feasible solution for (G', h'), since for every edge $(u, v) \in E_H$ and a path $(u, w_1, \ldots, w_p, v)$ in G', each of $w_1, \ldots, w_p$ corresponds to an edge of G (i.e., none of them is in $h'(V_H)$ since all vertices in $h'(V_H)$ is either a source or a sink), and at least one of them must be in F.

We finally claim $\mathsf{FRAC}' \geq \mathsf{FRAC}$, finishing the proof of the lemma. Let $\ell : V'_G \setminus h'(V_H) \to [0, 1]$ be a feasible fractional solution for (G', h'): for every edge $(u, v) \in E_H$ and a path $(u, w_1, \ldots, w_p, v)$ in G', all $\ell(w_1) + \cdots + \ell(w_p) \geq 1$. Since $V'_G \setminus h'(V_H) = E_G$, ℓ can be considered as a function from E_G to $[0, 1]$. Then ℓ is a feasible fractional solution for (G, h) as well, since for every edge $(u, v) \in E_H$ and a path $(u, w_1, \ldots, w_p, v)$ in G, $(u, (u, w_1), (w_1, w_2), \ldots, (w_p, v), v)$ is a path in G', implying that $\ell(u, w_1) + \cdots + \ell(w_p, v) \geq 1$. $\square$

Lemma 17.2.6. *There is an approximation-preserving reduction from* DIRECTED VERTEX MULTICUT*(H) to* DIRECTED MULTICUT*(H).*

Proof. Given an instance $G = (V_G, E_G)$ and $h : V_H \to V_G$ of DIRECTED VERTEX MULTICUT(H), let Opt be the integral optimal value for the instance (G, h) as follows. We construct an instance $G' = (V'_G, E'_G)$ and $h' : V_H \to V'_G$ of DIRECTED MULTICUT(H).

- $V'_G := h(V_H) \cup \{v_{in} : v \in V_G \setminus h(V_H)\} \cup \{v_{out} : v \in V_G \setminus h(V_H)\}$. Let $h' := h$.
- For each $v \in V_G \setminus h(V_H)$, add an edge (v_{in}, v_{out}) of weight 1 to G'.
- For each $(u, v) \in E_G$, add an edge (u_{out}, v_{in}) of weight ∞ to G' (when $v \in h(V_H)$, let $v_{in} = v_{out} = v$).

There is a natural one-to-one correspondence between non-terminal vertices of G and edges of G of weight 1. It is easy to see that $F \subseteq V_G \setminus h(V_H)$ is a feasible integral solution of (G', h') if and only if it is feasible in (G, h), proving the lemma. □

Therefore, it suffices to prove Theorem 17.2.4 for DIRECTED VERTEX MULTICUT(H). For the rest of the paper, we use DIRECTED MULTICUT(H) to denote this version. We propose our dictatorship test for DIRECTED MULTICUT(H) that will be used for proving Unique Games hardness. Let $G = (V_G, E_G)$ and $h : V_H \to V_G$ be an instance of DIRECTED MULTICUT(H) such that

- The optimal value of DIRECTED MULTICUT(H) is Opt.
- LP value of the instance is FRAC: there exists a map $\ell : V_G \to [0, 1]$ such that
  - $\ell(v) = 0$ if $v \in h(V_H)$.
  - $\sum_{v \in V_G} \ell(v) = \mathsf{FRAC}$.
  - For each $(u, v) \in E_H$ and every path $(h(u), w_1, \ldots, w_p, h(v))$ in G, $\sum_{i=1}^{p} \ell(h(w_i)) \geq 1$.

Our dictatorship test is a directed generalization of the previous section, and parameterized by G, H, h, ℓ, and $R \in \mathbb{N}$, and small $\epsilon > 0$. All graphs in this section are directed.

Take r to be a large integer to be determined later, and assume that for every $v \in V_G$, $\ell(v)$ is an integer multiple of $\frac{1}{r}$. This assumption still satisfies the property that $\sum_{v \in V_G} \ell(v) \leq \mathsf{FRAC} + \frac{|V_G|}{r}$. Define $\mathcal{D}^{\mathsf{M}}_{G,H,h,\ell,R,\epsilon} = (V, E)$ be the graph defined as follows. Consider the probability space (Ω, μ) where $\Omega := \{0, \ldots, r-1, *\}$, and $\mu : \Omega \mapsto [0, 1]$ with $\mu(*) = \epsilon$ and $\mu(x) = \frac{1-\epsilon}{r}$ for $x \neq *$. Let $V_G^N := V_G \setminus h(V_H)$ be the set of non-terminal vertices of G.

- $V = h(V_H) \cup \{v_x^\alpha\}_{\alpha \in V_G^N, x \in \Omega^R}$. Let v^α denote the set of vertices $\{v_x^\alpha\}_{x \in \Omega^R}$.
- For $\alpha \in V_G^N$ and $x \in \Omega^R$, $\mathsf{wt}(v_x^\alpha) = \mu^{\otimes R}(x)$. Note that the sum of weights is $|V_G^N|$.

- For each $(s, \alpha) \in E_G$ with $s \in h(V_H^S)$ and $\alpha \in V_G^N$, add edges from s to every vertex in v_x^α.

- For each $(\alpha, t) \in E_G$ with $t \in h(V_H^T)$ and $\alpha \in V_G^N$, add edges from every vertex in v_x^α to t.

- For $(\alpha, \beta) \in E_G$ with $\alpha, \beta \in V_G^N$ and $x, y \in \Omega^R$, we have an edge from v_x^α to v_y^β if
  - For any $1 \leq j \leq R$: $[y_j = (x_j + r \cdot \ell(\beta)) \mod r]$ or $[y_j = *]$ or $[x_j = *]$.

Completeness. We first prove that vertex cuts that correspond to *dictators* behave the same as ℓ. For any $q \in [R]$, let $V_q := \{v_x^\alpha : \alpha \in V_G^N, x_q = * \text{ or } x_q < r\ell(\alpha)\}$. Note that the total weight of V_q is

$$\sum_{\alpha \in V_G^N} (\epsilon + \frac{r\ell(\alpha)(1-\epsilon)}{r}) = \sum_{\alpha \in V_G^N} (\ell(\alpha) + \epsilon) \leq \mathsf{FRAC} + \frac{|V_G^N|}{r} + \epsilon |V_G^N|.$$

Lemma 17.2.7. *After removing vertices in V_q, for each $(s, t) \in E_H$, there is no path from $h(s)$ to $h(t)$.*

Proof. Fix i and let $p = (h(s), v_{x^1}^{\alpha_1}, \ldots, v_{x^z}^{\alpha_z}, h(t))$ be a path from $h(s)$ to $h(t)$ where $\alpha_j \in V_G^N$ and $x^j \in \Omega^R$ for each $1 \leq j \leq z$. Let $y_j := (x^j)_q$ for each $1 \leq j \leq z$. The construction ensures that $y_{j+1} = (y_j + r\ell(\alpha_{j+1})) \mod r$. Since we removed vertces from $v^{\alpha_{j+1}}$ whose qth coordinate is less than $r\ell(\alpha_{j+1})$, it means that $y_j < r - r\ell(\alpha_{j+1})$ and $y_{j+1} = (y_j + r\ell(\alpha_{j+1}))$. Moreover, $y_1 \geq r\ell(\alpha_1)$. Since any path from $h(s)$ to $h(t)$ satisfies $\sum_{i=1}^z r\ell(\alpha_i) \geq r$, there must be no path from $h(s)$ to $h(t)$. □

Soundness. To analyze soundness, we define a correlated probability space $(\Omega_1 \times \Omega_2, \nu)$ where both Ω_1, Ω_2 are copies of $\Omega = \{0, \ldots, r-1, *\}$. It is defined by the following process to sample $(x, y) \in \Omega^2$.

- Sample $x \in \{0, \ldots, r-1\}$. Let $y = (x + 1) \mod r$.

- Change x to $*$ with probability ϵ. Do the same for y independently.

Note that the marginal distribution of both x and y is equal to μ. Assuming $\epsilon < \frac{1}{2r}$, the minimum probability of any atom in $\Omega_1 \times \Omega_2$ is ϵ^2. For any $(x, y) \in \Omega^2$ with nonzero probability, $(*, y)$ and $(*, *)$ also have nonzero probabilities, so we can apply Lemma 3.3.3 to have $\rho(\Omega_1, \Omega_2; \nu) \leq \rho := 1 - \epsilon^4/2$.

Apply Theorem 3.3.10 ($\rho \leftarrow \rho, \alpha \leftarrow \epsilon^2, \epsilon \leftarrow \frac{\Gamma_\rho(\frac{\epsilon}{3},\frac{\epsilon}{3})}{2}$) to get τ and d. We will later apply this theorem with the parameters obtained here. Fix an arbitrary subset $C \subseteq V$, and let $C_\alpha := C \cap v^\alpha$. For $\alpha \in V_G^N$, call v^α *blocked* if $\mu^{\otimes R}(C_\alpha) \geq 1 - \epsilon$. The number of blocked v^α's is at most $\frac{\mathsf{wt}(C)}{1-\epsilon}$.

Setting $C' := \{v^\alpha \in V_D : v^\alpha \text{ is not blocked.}\}$, and applying Theorem 17.2.2 concludes that unless $\mathsf{wt}(C) > (1-\epsilon)\mathsf{Opt}$, there exists $(s,t) \in E_H$ and a path $(h(s), \alpha_1, \ldots, \alpha_z, h(t))$ in G where each v^{α_i} is unblocked for each $i \in [k]$.

For $1 \leq j \leq z$, let $S_j \subseteq v^{\alpha_j}$ be such that $x \in S_j$ if there exists a path $(s, v^{\alpha_1}_{x^1}, \ldots, v^{\alpha_{j-1}}_{x^{j-1}}, v^{\alpha_j}_x)$ for some $x^1, \ldots, x^{j-1}$. For $1 \leq j \leq z$, let $f_j : \Omega^R \mapsto \{0,1\}$ be the indicator function of S_j. We prove that if none of f_j reveals any *influential coordinate*, $\mu^{\otimes R}(S_z) > 0$, which shows that there exists a path from $h(s)$ to $h(t)$ even after removing vertices in C.

Lemma 17.2.8. *Suppose that for any* $1 \leq j \leq z$ *and* $1 \leq i \leq R$, $\mathsf{Inf}_i^{\leq d}[f_j] \leq \tau$. *Then* $\mu^{\otimes R}(S_z) > 0$.

Proof. We prove by induction that $\mu^{\otimes R}(S_j) \geq \frac{\epsilon}{3}$. It holds when $j = 1$ since v^{α_1} is unblocked. Assuming $\mu^{\otimes R}(S_j) \geq \frac{\epsilon}{3}$, since S_j does not reveal any influential coordinate, Theorem 3.3.10 shows that for any subset $T_{j+1} \subseteq v^{\alpha_{j+1}}$ with $\mu^{\otimes R}(T_{j+1}) \geq \frac{\epsilon}{3}$, there exists an edge from S_j and T_{j+1}. If $S'_{j+1} \subseteq v^{\alpha_{j+1}}$ is the set of out-neighbors of S_j, we have $\mu^{\otimes R}(S'_{j+1}) \geq 1 - \frac{\epsilon}{3}$. Since $v^{\alpha_{j+1}}$ is unblocked, $\mu^{\otimes R}(S'_{j+1} \setminus C) \geq \frac{2\epsilon}{3}$, completing the induction. □

In summary, in the completeness case, if we cut vertices of total weight $\mathsf{FRAC} + \epsilon|V_G^N|$, we cut every $(h(s), h(t))$ pair for each $(s,t) \in E_H$. In the soundness case, unless we cut vertices of total weight larger than $(1-\epsilon)\mathsf{Opt}$, some $(h(s), h(t))$ wih $(s,t) \in E_H)$ is not cut. The gap is $\frac{\mathsf{Opt}(1-\epsilon)}{\mathsf{FRAC}+|V_G^N|/r+\epsilon|V_G^N|}$. For fixed G and H, increasing r and decreasing ϵ faster makes the gap arbitrarily close to $\frac{\mathsf{Opt}}{\mathsf{FRAC}}$.

17.3 s-t NODE DOUBLE CUT

17.3.1 LP Gap

This section studies s-t NODE DOUBLE CUT. Bérczi et al. [BCK+17] gave a 2-approximation algorithm for s-t NODE DOUBLE CUT. It is based on the following natural LP relaxation, where we have a variable d_u for every node $u \in V$:

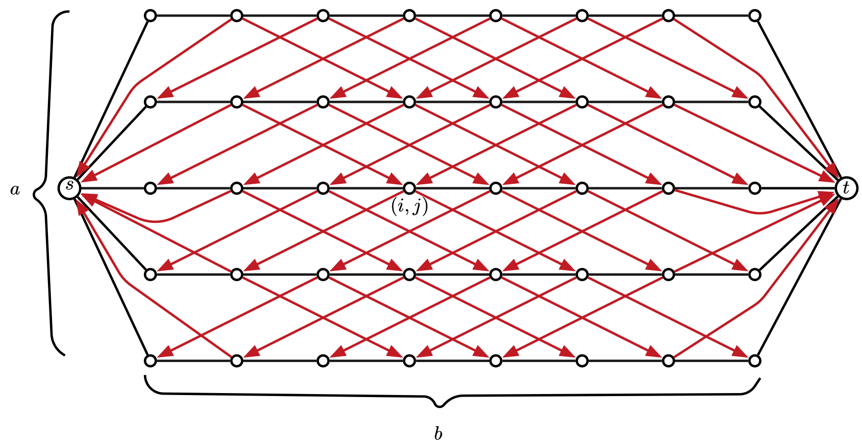

Figure 17.1: $D_{a,b}$ in the proof of Lemma 17.3.1 and $(2-\epsilon)$-inapproximability of s-t NODE DOUBLE CUT.

$$\min \sum_{v \in V \setminus \{s,t\}} c_v d_v \qquad \text{(Path-Blocking-LP)}$$

$$\sum_{v\in P} d_v + \sum_{v \in Q} d_v - d_u \geq 1 \; \forall\, P \in \mathcal{P}^{u\to s}, Q \in \mathcal{P}^{u \to t}, \; \forall\, u \in V$$

$$d_s, d_t = 0$$

$$d_v \geq 0 \; \forall\, v \in V$$

The integrality gap of this LP is at most 2 as proved in [BCK+17]. We prove that it is at least $2 - o(1)$. Consider the following graph in Figure 17.1. Our next lemma shows a lower bound on the integrality gap that nearly matches the approximation factor achieved by our rounding algorithm.

Lemma 17.3.1. *The integrality gap of the Path-Blocking-LP for directed graphs containing n nodes is at least $2 - 7/n^{1/3}$.*

For two integers $a, b \in \mathbb{N}$, consider the directed graph $D_{a,b} = (V_D, A_D)$ obtained as follows (see Figure 17.1): Let $V_D := \{s,t\} \cup ([a] \times [b])$. There are $ab + 2$ nodes. Let $I_D := [a] \times [b]$ and call them as the *internal nodes*. The set of arcs A_D are as follows:

1. For each $1 \leq i \leq a$, there is a bidirected arc between s and $(i, 1)$, and a bidirected arc between (i, b) and t.

2. For each $1 \leq i \leq a$ and $1 \leq j < b$, there is a bidirected arc between (i, j) and $(i, j + 1)$.

3. For each $1 \leq i < a$ and $2 \leq j \leq b - 1$, there is an arc from (i, j) to $(i + 1, j - 2)$, and an arc from (i, j) to $(i + 1, j + 2)$ (let $(i, 0) := s$ and $(i, b + 1) := t$ for every i). Call them *jumping arcs*.

Lemma 17.3.2. *$D_{a,b}$ has the following properties:*

1. *For each internal node $\alpha = (\alpha_1, \alpha_2) \in I_D$, each $\alpha \to s$ path has at least $\alpha_2 - a$ internal nodes other than α. Similarly, each $\alpha \to t$ path has at least $b - \alpha_2 - a + 1$ internal nodes other than α.*

2. *If $S \subseteq I_D$ is such that the subgraph induced by $V_D \setminus S$ has no node v that has paths to both s and t, then $|S| \geq 2a - 1$.*

Proof. 1. Jumping arcs are the only arcs that change α_2 by 2 while all other arcs change α_2 by 1. However, a path to s can use at most $a-1$ jumping arcs because they strictly increase α_1. The first property follows from these observations.

2. Suppose that $S \subseteq I_D$ is such that the subgraph induced by $V_D \setminus S$ has no node v that has paths to both s and t. For $i = 1, \ldots, a$, let $s_i := |S \cap \{\{i\} \times [b]\}|$. We note that $s_i \geq 1$ for each i, otherwise s can reach t and t can reach s.

   Suppose $s_i = 1$ for some $1 < i \leq a$ and let j be such that $S \cap \{\{i\} \times [b]\} = (i, j)$. If $j = 1$, then $(i, 2) \in V_D \setminus S$ and $(i, 2)$ can reach both s and t. If $j = b$, then $(i, b - 1) \in V_D \setminus S$ and $(i, b - 1)$ can reach both s and t. Therefore, we have $1 < j < b$. Then $s_{i-1} \geq 3$ because $(i - 1, j - 1), (i - 1, j), (i - 1, j + 1)$ can reach both s and t using one jumping arc followed by regular arcs in the ith row.

   Therefore, $|S| = \sum_{i=1}^{a} s_i \geq 1 + 2(a - 1) = 2a - 1$.

□

Proof of Lemma 17.3.1. The integer optimum of Path-Blocking-LP on $D_{a,b}$ is at least $2a-1$ by the second property of Lemma 17.3.2. Let $r := b-2a+1$. We set $d_v := 1/r$ for every internal node v. The resulting solution is feasible to Path-Blocking-LP: Indeed, consider $\alpha = (\alpha_1, \alpha_2)$. By the first property of Lemma 17.3.2, any $\alpha \to s$ path and $\alpha \to t$ path have to together traverse at least $\alpha_2 - a + (b - \alpha_2 - a + 1) = r$ internal nodes.

Setting $b = a^2$, the integrality gap is at least $(2a-1)/(a^3/r) = 2-1/a^3+4/a^2-5/a \geq 2 - 6/a$ for $a \geq 2$. Using the fact that $a = (|V(D_{a,b})| - 2)^{1/3}$, we get the desired bound on the integrality gap. □

17.3.2 Dictatorship Test

Consider the digraph $D_{a,b}$ introduced in Section 17.3.1. Let $r = b-2a+1$. Based on $D_{a,b}$, we define the *dictatorship test* graph $\mathcal{D}^{\mathsf{st}}_{a,b,R,\epsilon} = (V, A)$ as follows, for a positive integer R and $\epsilon > 0$. It will be used to show hardness results under the Unique Games Conjecture in Chapter 18. Consider the probability space (Ω, μ) where $\Omega := \{0, \ldots, r-1, *\}$, and $\mu : \Omega \mapsto [0, 1]$ with $\mu(*) = \epsilon$ and $\mu(x) = (1-\epsilon)/r$ for $x \neq *$.

1. $V = \{s, t\} \cup \{v^{\alpha}_{x}\}_{\alpha \in I_D, x \in \Omega^R}$. Let v^{α} denote the set of vertices $\{v^{\alpha}_{x}\}_{x \in \Omega^R}$.
2. For $\alpha \in I_D$ and $x \in \Omega^R$, define the weight as $\mathsf{wt}(v^{\alpha}_{x}) = \mu^{\otimes R}(x)$. We note that the sum of weights is ab. The terminals s and t have infinite weight.
3. For each arc between s and $\alpha \in I_D$, for each $x \in \Omega^R$, add an arc with the same direction betweem s and v^{α}_{x}. Do the same for each arc between t and $\alpha \in I_D$.
4. For each arc $(\alpha, \beta) \in A_D$ with $\alpha = (\alpha_1, \alpha_2), \beta = (\beta_1, \beta_2) \in I_D$ and $x, y \in \Omega^R$, we have an arc from v^{α}_{x} to v^{β}_{y} according to the following rule (note that $\alpha_2 \neq \beta_2$).
   (a) $\alpha_2 < \beta_2$: add an arc if for any $1 \leq j \leq R$: $[y_j = (x_j + 1) \mod r]$ or $[y_j = *]$ or $[x_j = *]$. Call them *forward* arcs.
   (b) $\alpha_2 > \beta_2$: add an arc if for any $1 \leq j \leq R$: $[y_j = (x_j - 1) \mod r]$ or $[y_j = *]$ or $[x_j = *]$. Call them *backward* arcs.
   (c) If $(\alpha, \beta) \in A_D$ is a jumping arc, call $(v^{\alpha}_{x}, v^{\beta}_{y})$ also a jumping arc.

Completeness. We first prove that removing a set of vertices that correspond to *dictators* behaves the same as the fractional solution that gives $1/r$ to every vertex. For any $q \in [R]$, let $V_q := \{v^{\alpha}_{x} : \alpha \in I_D, x_q = * \text{ or } 0\}$. We note that the total weight of V_q is

$$ab\left(\epsilon + \frac{1-\epsilon}{r}\right) \leq ab\epsilon + \frac{ab}{b-2a}.$$

Lemma 17.3.3. *After removing vertices in V_q, no vertex in V can reach both s and t.*

Proof. Suppose towards contradiction that there exists a vertex that can reach both s and t. First, assume that this vertex is $v^{\alpha_0}_{x_0}$ for some $\alpha_0 = ((\alpha_0)_1, (\alpha_0)_2) \in I_D$ and $x_0 \in \Omega^R$. Let $p_1 = (v^{\alpha_0}_{x_0}, v^{\beta_1}_{y_1}, \ldots, v^{\beta_l}_{y_l}, s)$ be a $v^{\alpha_0}_{x_0} \to s$ path and $p_2 = (v^{\alpha_0}_{x_0}, v^{\alpha_1}_{x_1}, \ldots, v^{\alpha_k}_{x_k}, t)$ be a $v^{\alpha_0}_{x_0} \to t$ path in $\mathcal{D}^{\mathsf{st}}_{R,\epsilon} - V_q$ for some $k, l \in \mathbb{N}$, and $\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_l \in I_D$, and $x_1, \ldots, x_k, y_1, \ldots, y_l \in \Omega^R$.

Proposition 17.3.4. $(x_k)_q \geq (x_0)_q + b - (\alpha_0)_2 - a + 1$.

Proof. Consider the two sequences $(\alpha_0)_q, \ldots, (\alpha_k)_q$ and $(x_0)_q, \ldots, (x_k)_q$. Since we removed V_q, $(\alpha_{i+1})_2 > (\alpha_i)_2$ if and only if $(x_{i+1})_q > (x_i)_q$. Let n_{jf}, n_{jb}, n_{rf}, n_{rb} be the number forward jumping arcs, backward jumping arcs, forward non-jumping arcs, backward non-jumping arcs in p_2 respectively. Jumping forward arcs, jumping backward arcs, non-jumping forward arcs, and non-jumping backward arcs change $(\alpha_i)_2$ by $+2, -2, +1$, and -1 respectively. By considering $(\alpha_0)_q, \ldots, (\alpha_k)_q$,

$$2n_{\mathsf{jf}} + n_{\mathsf{rf}} - 2n_{\mathsf{jb}} - n_{\mathsf{rb}} = (b+1) - (\alpha_0)_2.$$

Since using a jumping arc increases $(\alpha_i)_1$ by 1,

$$n_{\mathsf{jf}} + n_{\mathsf{jb}} \leq a - 1.$$

Forward arcs (whether they are jumping or not) increase $(x_i)_q$ by 1 and backward arc decrease it by 1. Consider $(x_0)_q, \ldots, (x_k)_q$,

$$\begin{aligned}(x_k)_q - (x_0)_q &\geq n_{\mathsf{jf}} + n_{\mathsf{rf}} - n_{\mathsf{jb}} - n_{\mathsf{rb}} - 1 \\ &\geq (2n_{\mathsf{jf}} + n_{\mathsf{rf}} - 2n_{\mathsf{jb}} - 2n_{\mathsf{rb}}) - (n_{\mathsf{jf}} - n_{\mathsf{jb}}) - 1 \\ &\geq b - (\alpha_0)_2 - a + 1,\end{aligned}$$

as claimed. □

The same proof for p_1 shows that $(x_0)_q \geq (y_l)_q + (\alpha_0)_2 - a$. Therefore, $(x_k)_q \geq (y_l)_q + b - 2a + 1$ and $(y_l)_q \geq 1$. This implies $(x_k)_q > b - 2a + 1 = r$, leading to contradiction. □

Soundness. Suppose that we removed some vertices C such that no vertex $w \in V \setminus C$ can reach both s and t. We show this happens only if C reveals an influential coordinate or $\mathsf{wt}(C) \geq 2a(1-\epsilon)$.

To analyze soundness, we define a correlated probability space $(\Omega_1 \times \Omega_2, \nu)$ where both Ω_1, Ω_2 are copies of $\Omega = \{0, \ldots, r-1, *\}$. It is defined by the following process to sample $(x, y) \in \Omega^2$.

1. Sample $x \in \{0, \ldots, r-1\}$. Let $y = (x+1) \mod r$.

2. Change x to $*$ with probability ϵ. Do the same for y independently.

We note that the marginal distribution of both x and y is equal to μ. Assuming $\epsilon < 1/2r$, the minimum probability of any atom in $\Omega_1 \times \Omega_2$ is ϵ^2. For any $(x, y) \in \Omega^2$ with nonzero probability, $(*, y)$ and $(*, *)$ also have nonzero probabilities, so we can apply Lemma 3.3.3 to have $\rho(\Omega_1, \Omega_2; \nu) \leq \rho := 1 - \epsilon^4/2$.

Apply Theorem 3.3.10 $\rho \leftarrow \rho, \alpha \leftarrow \epsilon^2, \epsilon \leftarrow \underline{\Gamma}_\rho(\epsilon/3, \epsilon/3)/2$ to get τ and d. We will later apply this theorem with the parameters obtained here. Fix an arbitrary subset $C \subseteq V$, and let $C_\alpha := C \cap v^\alpha$. For $\alpha \in I_D$, call v^α *blocked* if $\mu^{\otimes R}(C_\alpha) \geq 1 - \epsilon$. The number of blocked v^α's is at most $\mathsf{wt}(C)/(1-\epsilon)$.

By Property 2. of Lemma 17.3.2, unless $\mathsf{wt}(C) \geq (2a-1)(1-\epsilon)$ (i.e., unless $2a-1$ vertices are blocked), there exists $\alpha_0 \in I_D$ and a path $(v^{\alpha_0}, v^{\alpha_{-1}}, \ldots, v^{\alpha_{-k}}, s)$ and $(v^{\alpha_0}, v^{\alpha_1}, \ldots, v^{\alpha_l}, t)$ where each v^{α_i} is unblocked for $-k \leq i \leq l$.

For $-k \leq j \leq -1$, let $S_j \subseteq v^{\alpha_j}$ be such that $x \in S_j$ if there exists a path $(v_x^{\alpha_j}$, $v_{x^{j-1}}^{\alpha_{j-1}}$, $\ldots$, $v_{x^{-k}}^{\alpha_{-k}}$, $s)$ for some $x^{j-1}, \ldots, x^{-k}$. Similarly, For $1 \leq j \leq l$, let $S_j \subseteq v^{\alpha_j}$ be such that $x \in S_j$ if there exists a path $(v_x^{\alpha_j}, v_{x^{j+1}}^{\alpha_{j+1}}, \ldots, v_{x^l}^{\alpha_l}, t)$ for some $x^{j+1}, \ldots, x^l$. Let $f_j : \Omega^R \mapsto \{0, 1\}$ be the indicator function of S_j. We prove that if none of f_j reveals any *influential coordinate*, that there exists a $x^0 \in \Omega^R$ such that $v_{x^0}^{\alpha_0}$ can reach both s and t even after removing vertices in C.

Lemma 17.3.5. *Suppose that for any $j \in \{-k, \ldots, -1\} \cup \{1, \ldots, l\}$ and $1 \leq i \leq R$, $\mathsf{Inf}_i^{\leq d}[f_j] \leq \tau$. Then there exists a $x^0 \in \Omega^R$ such that $v_{x^0}^{\alpha_0}$ can reach both s and t*

Proof. We prove that $\mu^{\otimes R}(S_1) \geq \epsilon/3$ by induction on $j = l, \ldots, 1$. It holds when $j = l$ since v^{α_l} is unblocked. Assuming $\mu^{\otimes R}(S_j) \geq \epsilon/3$, since S_j does not reveal any influential coordinate, Theorem 3.3.10 shows that for any subset $T_{j-1} \subseteq v^{\alpha_{j-1}}$ with $\mu^{\otimes R}(T_{j-1}) \geq \epsilon/3$, there exists an arc from S_j and T_{j-1}. If $S'_{j-1} \subseteq v^{\alpha_{j-1}}$ is the set of in-neighbors of S_j, we have $\mu^{\otimes R}(S'_{j-1}) \geq 1 - \epsilon/3$. Since $v^{\alpha_{j-1}}$ is unblocked, $\mu^{\otimes R}(S'_{j-1} \setminus C) \geq 2\epsilon/3$, completing the induction.

The same argument also proves that $\mu^{\otimes R}(S_{-1}) \geq \epsilon/3$ by induction on $j = -k, \ldots, -1$. The total weight of the in-neighbors of S_{-1} in v^{α_0} is at least $1 - \epsilon/3$, and the total weight of the in-neighbors of S_1 in v^{α_0} is at least $1-\epsilon/3$. Therefore, the total weight of vertices in v^{α_0} that has outgoing arcs to both S_{-1} and S_1 is at least $1 - 2\epsilon/3$. Since α_0 is not blocked, there exists a vertex $v_{x^0}^{\alpha_0}$ that has outgoing arcs to both S_1 and S_{-1}, and is not contained C. This vertex can reach both s and t. □

In summary, in the completeness case, if we remove vertices of total weight at most $ab\epsilon + ab/(b-2a)$, no vertex can reach both s and t. In the soundness case, unless we reveal an influential coordinate or we remove vertices of total weight at least $(2a-1)(1-\epsilon)$,

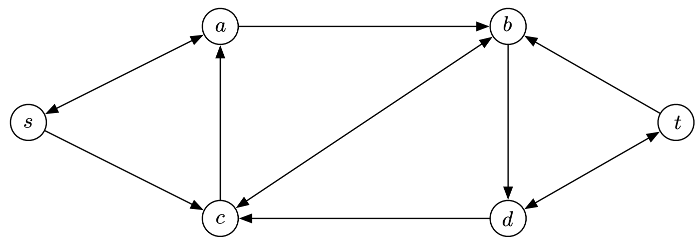

Figure 17.2: D in the proof of $(3/2-\epsilon)$-inapproximability of NODE DOUBLE CUT.

there exists a vertex that can reach both s and t. The gap between the two cases is at least

$$\frac{(2a-1)(1-\epsilon)}{ab\epsilon + ab/(b-2a)},$$

which approaches to 2 as a increases, $b = a^2$ and $\epsilon = 1/a^4$.

17.4 NODE DOUBLE CUT

Consider the directed graph $D = (V_D, A_D)$ (see Fig. 17.2) defined by

$$\begin{aligned}
&V_D := \{s,t,a,b,c,d\},\\
&A_D := \{(a,s),(s,a),(s,c),(c,a),(a,b),(b,c),(c,b),(d,c),(b,d),(d,t),(t,d),(t,b)\}.
\end{aligned}$$

Let $I_D := \{a,b,c,d\}$ be the set of internal vertices.

We summarize the properties of D that can be verified easily.

Proposition 17.4.1. *D has the following three properties.*

(i) For any vertex $v \in V$, there exists a vertex $u \in \{s,t\}$ such that every $v \to u$ path has at least three internal vertices.

(ii) Every $v \in I_D$ has an incoming arc from either s or t.

(iii) Even after deleting one vertex from I_D, there exists a $s \to t$ path or a $t \to s$ path with exactly three remaining internal vertices.

Based on D, we define the *dictatorship test* graph $\mathcal{D}^{\mathsf{global}}_{R,\epsilon} = (V, A)$ as follows, for a positive integer R and $\epsilon > 0$. It will be used to show hardness results under the Unique Games Conjecture in Chapter 18. Let $r = 3$. Consider the probability space (Ω, μ) where $\Omega := \{0, \ldots, r-1, *\}$, and $\mu : \Omega \mapsto [0, 1]$ with $\mu(*) = \epsilon$ and $\mu(x) = (1-\epsilon)/r$ for $x \neq *$.

1. We take $V := \{s, t\} \cup \{v^\alpha_x\}_{\alpha \in I_D, x \in \Omega^R}$. Let v^α denote the set of vertices $\{v^\alpha_x\}_{x \in \Omega^R}$.
2. For $\alpha \in I_D$ and $x \in \Omega^R$, we define the weight as $\mathsf{wt}(v^\alpha_x) := \mu^{\otimes R}(x)$. We note that the sum of weights is 4. The terminals s and t have infinite weight.
3. There are arcs from s to all vertices in v^c, from v^a to s, s to v^a, from v^d to t, from t to v^d, from t to v^b.
4. For each $(\alpha, \beta) \in \{(c, a), (a, b), (b, c), (c, b), (d, c), (b, d)\}$ and $x, y \in \Omega^R$, we have an arc from v^α_x to v^β_y if there exists $1 \leq j \leq R$ such that $y_j = (x_j + 1) \mod r$ or $y_j = *$ or $x_j = *$.

Completeness. We first prove that removing a set of vertices that correspond to *dictators* behaves the same as the fractional solution that gives $1/r$ to every internal vertex. For any $q \in [R]$, let $V_q := \{v^\alpha_x : \alpha \in I_D, x_q = * \text{ or } 0\}$. We note that the total weight of V_q is $4(\epsilon + (1-\epsilon)/r) \leq 4(1+\epsilon)/3$

Lemma 17.4.2. *After removing vertices in V_q, no vertex in V can reach both s and t.*

Proof. Suppose towards contradiction that there exists a vertex that can reach both s and t. First, assume that this vertex is $v^{\alpha_0}_{x_0}$ for some $\alpha_0 \in I_D$ and $x_0 \in \Omega^R$. By Property (i) of Proposition 17.4.1, there exists $u \in \{s, t\}$ such that every $\alpha_0 \to u$ path has at least three internal vertices in D. Let $(v^{\alpha_0}_{x_0}, v^{\alpha_1}_{x_1}, \ldots, v^{\alpha_k}_{x_k}, u)$ be a path from v^α_x to u in $\mathcal{D}^{\mathsf{global}}_{R,\epsilon} - V_q$. Note that $k \geq 2$.

Consider the sequence $((x_0)_q, (x_1)_q, \ldots, (x_k)_q)$. Recall that v^α_x has an arc to v^β_y for some α, β, x, y only if $y_q = (x_q + 1) \mod r$ or $y_q = *$ or $x_q = *$. Since we removed V_q, $(x_i)_q \notin \{0, *\}$, $(x_i)_q = (x_{i-1})_q + 1$. This forces $k \leq 1$, leading to contradiction.

Finally, assume that s can reach t, and let $(s, v^{\alpha_0}_{x_0}, v^{\alpha_1}_{x_1}, \ldots, v^{\alpha_k}_{x_k}, t)$ be a $s \to t$ path for some $\alpha_i \in I_D$, $x_i \in \Omega^R$. Every $s \to t$ path in D has to have at least three internal vertices, which forces $k \geq 2$, but considering the sequence $((x_0)_q, (x_1)_q, \ldots, (x_k)_q)$ forces $k \leq 1$, which leads to contradiction. Paths from t to s can be ruled out in the same way. $\square$

Soundness. Suppose that we removed some vertices C such that there exist two vertices $u, v \in V \setminus C$ where no vertex $w \in V \setminus C$ can reach both u and v. This implies that no vertex $w \in V \setminus C$ can reach both s and t, since both u and v have an incoming arc from either s or t. Therefore, it suffices to show that unless C reveals an influential coordinate or $\mathsf{wt}(C) \geq 2(1-\epsilon)$, either s can reach t or t can reach s.

To analyze soundness, we define a correlated probability space $(\Omega_1 \times \Omega_2, \nu)$ where both Ω_1, Ω_2 are copies of $\Omega = \{0, \ldots, r-1, *\}$. It is defined by the following process to sample $(x, y) \in \Omega^2$.

1. Sample $x \in \{0, \ldots, r-1\}$. Let $y = (x+1) \mod r$.
2. Change x to $*$ with probability ϵ. Do the same for y independently.

We note that the marginal distribution of both x and y is equal to μ. Assuming $\epsilon < 1/2r$, the minimum probability of any atom in $\Omega_1 \times \Omega_2$ is ϵ^2. We use the following lemma to bound the correlation $\rho(\Omega_1, \Omega_2; \nu)$. For any $(x, y) \in \Omega^2$ with nonzero probability, $(*, y)$ and $(*, *)$ also have nonzero probabilities, so we can apply Lemma 3.3.3 to have $\rho(\Omega_1, \Omega_2; \nu) \leq \rho := 1 - \epsilon^4/2$.

Apply Theorem 3.3.10 by setting $\rho \leftarrow \rho, \alpha \leftarrow \epsilon^2, \epsilon \leftarrow \underline{\Gamma}_\rho(\epsilon/3, \epsilon/3)/2$ to get τ and d. We will later apply this theorem with the parameters obtained here. Fix an arbitrary subset $C \subseteq V$, and let $C_\alpha := C \cap v^\alpha$. For $\alpha \in I_D$, call v^α *blocked* if $\mu^{\otimes R}(C_\alpha) \geq 1 - \epsilon$. The number of blocked v^α's is at most $\mathsf{wt}(C)/(1-\epsilon)$.

By Property (iii) of Proposition 17.4.1, unless $\mathsf{wt}(C) \geq 2(1-\epsilon)$ (i.e., unless two vertices are blocked), there exists a path $(s, v^{\alpha_1}, v^{\alpha_2}, v^{\alpha_3}, t)$ or $(t, v^{\alpha_1}, v^{\alpha_2}, v^{\alpha_3}, s)$ where each v^{α_i} is unblocked. Without loss of generality, suppose we have a path $(s, v^{\alpha_1}, v^{\alpha_2}, v^{\alpha_3}, t)$.

For $1 \leq j \leq 3$, let $S_j \subseteq v^{\alpha_j}$ be such that $x \in S_j$ if there exists a path $(v_x^{\alpha_j}, v_{x^{j+1}}^{\alpha_{j+1}}, \ldots, v_{x^3}^{\alpha_3}, t)$ for some $x^{j+1}, \ldots, x^3$. For $1 \leq j \leq 3$, let $f_j : \Omega^R \mapsto \{0, 1\}$ be the indicator function of S_j. We prove that if none of f_j reveals any *influential coordinate*, then $\mu^{\otimes R}(S_1) > 0$, which shows that there exists a $s \to t$ path even after removing vertices in C.

Lemma 17.4.3. *Suppose that for any* $1 \leq j \leq 3$ *and* $1 \leq i \leq R$, $\mathsf{Inf}_i^{\leq d}[f_j] \leq \tau$. *Then* $\mu^{\otimes R}(S_1) > 0$.

Proof. We prove by induction that $\mu^{\otimes R}(S_j) \geq \epsilon/3$ for $j = 3, 2, 1$. It holds when $j = 3$ since v^{α_3} is unblocked. Assuming $\mu^{\otimes R}(S_j) \geq \epsilon/3$, since S_j does not reveal any influential coordinate, Theorem 3.3.10 shows that for any subset $T_{j-1} \subseteq v^{\alpha_{j-1}}$ with $\mu^{\otimes R}(T_{j-1}) \geq \epsilon/3$, there exists an arc from S_j and T_{j-1}. If $S'_{j-1} \subseteq v^{\alpha_{j-1}}$ is the set of in-neighbors of S_j, we have $\mu^{\otimes R}(S'_{j-1}) \geq 1 - \epsilon/3$. Since $v^{\alpha_{j-1}}$ is unblocked, $\mu^{\otimes R}(S'_{j-1} \setminus C) \geq 2\epsilon/3$, completing the induction. $\square$

In summary, in the completeness case, if we remove vertices of total weight $\leq 4(1+\epsilon)/3$, no vertex can reach both s and t. In the soundness case, unless we reveal an influential coordinate or we remove vertices of total weight at least $2(1-\epsilon)$, there is a $s \to t$ path or $t \to s$ path, which means that either s or t can reach every vertex. The gap between the two cases is at least

$$\frac{2(1-\epsilon)}{4(1+\epsilon)/3},$$

which approaches to $3/2$ as $\epsilon \to 0$.

17.5 VERTEX COVER ON k-PARTITE GRAPHS

Fix $k \geq 3$ and $\epsilon > 0$. Let $\Omega := \{*, 0, 1\}$. Let $R \in \mathbb{N}$ be another parameter. Our dictatorship test $\mathcal{D}^{\mathsf{vc}}_{k,R,\epsilon} = ([k] \times \Omega^R, E)$ is defined as follows. Each vertex is represented by v^i_x where $i \in [k]$ and $x \in \Omega^R$ is a R-dimensional vector. Let $v^i := \{v^i_x\}_{x \in \Omega^R}$. There will be no edge within each v^i, so $\mathcal{D}^{\mathsf{vc}}_{k,R,\epsilon}$ will be k-partite. Consider the probability space (Ω, μ) where $\Omega := \{0, 1, *\}$, and $\mu : \Omega \mapsto [0,1]$ with $\mu(*) = \epsilon$ and $\mu(x) = (1-\epsilon)/2$ for $x \neq *$. We define the weight $\mathsf{wt}(v^i_x) := \mu^{\otimes R}(x) = \prod_{i=1}^{R} \mu(x_i)$. The sum of weights is k. The edges are constructed as follows.

1. There is an edge between v^i_x with $x = (x_1, \dots, x_R)$ and v^j_y with $y = (y_1, \dots, y_R)$ if and only if
   (a) $i \neq j$.
   (b) For any $1 \leq l \leq R$: $[x_l \neq y_l]$ or $[y_l = *]$ or $[x_l = *]$.

Completeness. Fix $q \in [R]$ and let $U_q := \{v^i_x : x_q = 0 \text{ or } *\}$. The weight of U_q is $\mathsf{wt}(U_q) = k(1+\epsilon)/2$.

Lemma 17.5.1. *U_q is a vertex cover.*

Proof. Let $\{v^i_x, v^j_y\}$ be an edge of $\mathcal{D}^{\mathsf{vc}}_{k,R,\epsilon}$. If both endpoints do not belong to U_q, it implies $x_q = y_q = 1$. It contradicts our construction. □

Soundness. To analyze soundness, we define a correlated probability space $(\Omega_1 \times \Omega_2, \nu)$ where both Ω_1, Ω_2 are copies of Ω. It is defined by the following process to sample $(x, y) \in \Omega^2$.

1. Sample $x \in \{0,1\}$ uniformly at random. Let $y = 1 - x$.
2. Change x to $*$ with probability ϵ. Do the same for y independently.

We note that the marginal distribution of both x and y is equal to μ. Assuming $\epsilon < 1/3$, the minimum probability of any atom in $\Omega_1 \times \Omega_2$ is ϵ^2. For any $(x,y) \in \Omega^2$ with nonzero probability, $(*, y)$ and $(*, *)$ also have nonzero probabilities, so we can apply Lemma 3.3.3 to have $\rho(\Omega_1, \Omega_2; \nu) \leq \rho := 1 - \epsilon^4/2$. Apply Theorem 3.3.10 ($\rho \leftarrow \rho, \alpha \leftarrow \epsilon^2, \epsilon \leftarrow \underline{\Gamma}_\rho(\epsilon, \epsilon)/2$) to get τ and d. We will later apply this theorem with the parameters obtained here.

Fix an arbitrary vertex cover $U \subseteq V$, and let $U_i := U \cap v^i$ for $i \in [k]$. Let $f_i : \Omega^R \mapsto \{0,1\}$ be the indicator function of U_i. Call v^i *blocked* if $\mathbb{E}[f_i] = \mu^{\otimes R}(U_i) \geq 1 - \epsilon$. The number of blocked v^i's is at most $\mathsf{wt}(U)/(1-\epsilon)$. We prove that if none of f_i reveals any *influential coordinate*, all but one v^i's must be blocked.

Lemma 17.5.2. *Suppose that for any* $1 \leq i \leq k$ *and* $1 \leq j \leq R$, $\mathsf{Inf}_j^{\leq d}[f_i] \leq \tau$. *Then at least* $k-1$ v^i*'s must be blocked.*

Proof. Assume towards contradiction that there exist $i_1 \neq i_2 \in [k]$ such that v^{i_1} and v^{i_2} are unblocked. Since both f_{i_1} and f_{i_2} do not reveal influential coordinates and $\mathbb{E}[f_{i_1}], \mathbb{E}[f_{i_2}] \leq 1 - \epsilon$, Theorem 3.3.10 ($f \leftarrow 1 - f_1, g \leftarrow 1 - f_2$) shows that $\mathbb{E}_{(x,y)\sim\nu^{\otimes R}}[(1-f_1)(x) \cdot (1 - f_2)(y)]$ is strictly greater than 0. This implies that there exists x, y such that there is an edge between $v_x^{i_1}$ and $v_y^{i_2}$ but neither $v_x^{i_1}$ nor $v_y^{i_2}$ is contained in U. This contradicts that U is a vertex cover. $\square$

Therefore, if U does not reveal any influential coordinate, then $\mathsf{wt}(U) \geq (k-1)(1-\epsilon)$. In summary, in the completeness case, there exists a vertex cover of weight $k(1+\epsilon)/2$. In the soundness case, unless we reveal an influential coordinate, every vertex cover has weight at least $(k-1)(1-\epsilon)$. The gap between the two cases is at least

$$\frac{2(k-1)(1-\epsilon)}{k(1+\epsilon)},$$

which approaches to $2(k-1)/k$ as $\epsilon \to 0$.

17.6 SHORTEST PATH EDGE CUT

We propose our dictatorship test for SHORT PATH EDGE CUT that will be used for proving Unique Games hardness. It is parameterized by positive integers a, b, r, R. It is inspired by the integrality gap instances by Baier et al. [BEH+10] Mahjoub and and McCormick [MM10], and made such that the edge cuts that correspond to *dictators* behave the same as the fractional solution that cuts $\frac{1}{r}$ fraction of every edge. All graphs in this section are undirected.

For positive integers a, b, r, R, we construct $\mathcal{D}^{\mathsf{E}}_{a,b,r,R} = (V, E)$. Let $\Omega = \{0, \dots, r-1\}$, and $\mu : \Omega \mapsto [0, 1]$ with $\mu(x) = \frac{1}{r}$ for each $x \in \Omega$. We also define a correlated probability space $(\Omega_1 \times \Omega_2, \nu)$ where both Ω_1, Ω_2 are copies of Ω. It is defined by the following process to sample $(x, y) \in \Omega^2$.

- Sample $x \in \{0, \dots, r-1\}$. Let $y = (x + 1) \mod r$.
- With probability $1 - \frac{1}{r}$, output (x, y). Otherwise, resample $x, y \in \Omega$ independently and output (x, y).

Note that the marginal distribution of both x and y is equal to μ. Given $x = (x_1, \dots, x_R) \in \Omega^R$ and $y = (y_1, \dots, y_R) \in \Omega^R$, let $\nu^{\otimes R}(x, y) = \prod_{i=1}^{R} \nu(x_i, y_i)$. We define $\mathcal{D}^{\mathsf{E}}_{a,b,r,R} = (V, E)$ as follows.

- $V = \{s, t\} \cup \{v^i_x\}_{0 \le i \le b, x \in \Omega^R}$. Let v^i denote the set of vertices $\{v^i_x\}_{x \in \Omega^R}$.
- For any $x \in \Omega^R$, there is an edge from s to v^0_x and an edge from v^b_x to t, both with weight ∞ and length 1.
- For $0 \le i < b, x \in \Omega^R$, there is an edge (v^i_x, v^{i+1}_x) of length a and weight ∞. Call it a *long edge*.
- For any $0 \le i < b\ x, y \in \Omega^R$, there is an edge (v^i_x, v^{i+1}_y) of length 1 and weight $\nu^{\otimes R}(x, y)$. Note that $\nu^{\otimes R}(x, y) > 0$ for any $x, y \in \Omega^R$. Call it a *short edge*. The sum of finite weights is b.

Completeness. We first prove that edge cuts that correspond to *dictators* behave the same as the fractional solution that gives $\frac{1}{r}$ to every short edge. Fix $q \in [R]$ and let E_q be the set of short edges defined by

$$E_q := \{(v^i_x, v^{i+1}_y) : 0 \le i < b, y_q \ne x_q + 1 \mod R \text{ or } (x_q, y_q) = (0, 1)\}.$$

When $(x, y) \in \Omega_1 \times \Omega_2$ is sampled according to ν, the probability that $y_q \ne x_q + 1 \mod R$ or $(x_q, y_q) = (0, 1)$ is at most $\frac{2}{r}$. The total weight of E_q is $\frac{2b}{r}$.

Lemma 17.6.1. *After removing edges in E_q, the length of the shortest path is at least $a(b-r+1)$.*

Proof. Let $p = (s, v_{x^1}^{i_1}, \ldots, v_{x^z}^{i_z}, t)$ be a path from s to t where $i_j \in \{0, \ldots, b\}$ and $x^j \in \Omega^R$ for each $1 \leq j \leq z$. Let $y_j := (x^j)_q \in \{0, \ldots, r-1\}$ for each $1 \leq j \leq z$.

For each $1 \leq j < z$, the edge (p_j, p_{j+1}) is either a long edge or a short edge, and either taken forward (i.e., $i_j < i_{j+1}$) or backward (i.e., $i_j > i_{j+1}$). Let $z_{\mathsf{LF}}, z_{\mathsf{SF}}, z_{\mathsf{LB}}, z_{\mathsf{SB}}$ be the number of long edges taken forward, short edges taken forward, long edges taken backward, and shot edges taken backward, respectively ($z_{\mathsf{LF}} + z_{\mathsf{SF}} + z_{\mathsf{LB}} + z_{\mathsf{SB}} = z - 1$). By considering how i_j changes,

$$z_{\mathsf{LF}} + z_{\mathsf{SF}} - z_{\mathsf{LB}} - z_{\mathsf{SB}} = b. \tag{17.1}$$

Consider how y_j changes. Taking a long edge does not change y_j. Taking a short edge forward increases y_j by 1 mod r, taking a short edge backward decreases y_j by 1 mod r. Since E_q is cut, y_j can never change from 0 to 1. This implies

$$z_{\mathsf{SF}} - z_{\mathsf{SB}} \leq r - 1. \tag{17.2}$$

(17.1) − (17.2) yields $z_{\mathsf{LF}} - z_{\mathsf{LB}} \geq b - r + 1$. The total length of p is at least $a \cdot z_{\mathsf{LF}} \geq a(b-r+1)$. $\square$

Soundness. We first bound the correlation $\rho(\Omega_1, \Omega_2; \nu)$. The following lemma of Wenner [Wen13] gives a convenient way to bound the correlation.

Lemma 17.6.2 (Corollary 2.18 of [Wen13])**.** *Let $(\Omega_1 \times \Omega_2, \delta\mu + (1-\delta)\mu')$ be two correlated spaces such that the marginal distribution of at least one of Ω_1 and Ω_2 is identical on μ and μ'. Then,*

$$\rho(\Omega_1, \Omega_2; \delta\mu + (1-\delta)\mu') \leq \sqrt{\delta \cdot \rho(\Omega_1, \Omega_2; \mu)^2 + (1-\delta) \cdot \rho(\Omega_1, \Omega_2; \mu')^2}.$$

When (x, y) is sampled from ν, they are completely independent with probability $\frac{1}{r}$. Therefore, we have $\rho := \rho(\Omega_1, \Omega_2; \nu) \leq \sqrt{1 - \frac{1}{r}}$. By Sheppard's Formula,

$$\begin{aligned}\underline{\Gamma}_\rho(\frac{1}{2}, \frac{1}{2}) &= \frac{1}{4} + \frac{1}{2\pi} \arcsin(-\rho) \geq \frac{1}{4} - \frac{1}{2\pi} \arccos(\frac{1}{\sqrt{r}}) \\ &= \sum_{n=0}^{\infty} \frac{(2n)!}{4^n (n!)^2 (2n+1)} (\frac{1}{\sqrt{r}})^{2n+1} \geq \frac{1}{\sqrt{r}}.\end{aligned}$$

Apply Theorem 3.3.10 ($\rho \leftarrow \rho, \alpha \leftarrow \frac{1}{r^3}, \epsilon \leftarrow \frac{\underline{\Gamma}_\rho(\frac{1}{2},\frac{1}{2})}{3}$) to get τ and d. We will later apply this theorem with the parameters obtained here.

Fix an arbitrary subset $C \subseteq E$ of short edges. For $0 \le i < b$, let $C_i = C \cap (v^i \times v^{i+1})$. Call a pair $(i, i+1)$ as the ith layer, and say it is blocked when $\nu^{\otimes R}(C_i) \ge \frac{\underline{\Gamma}_\rho(\frac{1}{2},\frac{1}{2})}{2}$. Let b' be the number of blocked layers. For $0 \le i \le b$, let $S_i \subseteq v^i$ be such that $x \in S_i$ if there exists a path $(s, p_0, \ldots, p_i = v^i_x)$ such that

- For $0 \le i' \le i$, $p_{i'} \in v^{i'}$.
- For $0 \le i' < i$, $(p_{i'}, p_{i'+1})$ is short if and only if the i'th layer is unblocked.

Let $f_i : \Omega^R \mapsto [0,1]$ be the indicator function of S_i. We prove that if none of f_i reveals any *influential coordinate*, S_b is nonempty, implying that there exists a path using b' long edges and $b - b'$ short edges. . Therefore, even after removing edges in C, the length of the shortest path is at most $2 + ab' + (b - b')$.

Lemma 17.6.3. *Suppose that for any* $0 \le i \le b$ *and* $1 \le j \le R$*,* $\mathsf{Inf}_j^{\le d}[f_i] \le \tau$*. Then* $S_b \ne \emptyset$*.*

Proof. Assume towards contradiction that $S_b = \emptyset$. Since $S_0 = \Omega^R$ and $S_i = S_{i+1}$ if the ith layer is blocked (and we use long edges), there must exist i such that the ith layer is unblocked and $\mu^{\otimes R}(S_i) \ge \frac{1}{2}, \mu^{\otimes R}(S_{i+1}) < \frac{1}{2}$. All short edges between S_i and $v^{i+1} \setminus S_{i+1}$ are in C_i. Theorem 3.3.10 implies that $\nu^{\otimes R}(C_i) > \frac{2}{3}\underline{\Gamma}_\rho(\frac{1}{2}, \frac{1}{2})$. This contradicts the fact that the ith layer is unblocked. □

In summary, in the completeness case, if we cut edges of total weight $k := k(a, b, r) = \frac{2b}{r}$, the length of the shortest path is at least $l := l(a, b, r) = a(b - r + 1)$. In the soundness case, even after cutting edges of total weight k', at most $\frac{2k'}{\underline{\Gamma}_\rho(\frac{1}{2},\frac{1}{2})} \le 2k'\sqrt{r}$ layers are blocked, the length of the shortest path is at most $l' = 2 + (b - 2k'\sqrt{r}) + 2ak'\sqrt{r}$.

- Let $a = 4, b = 2r - 1$ so that $k \le 4, l = 4r$. Requiring $l' \ge l$ results in $k' = \Omega(\sqrt{r})$, giving a gap of $\Omega(\sqrt{r}) = \Omega(\sqrt{l})$ between the completeness case and the soundness case for LENGTH-BOUNDED EDGE CUT.
- Let $a = \sqrt{r}, b = 2r - 1$ so that $k \le 4, l = r^{1.5}$. Requiring $k' \le 4$ results in $l' = O(r)$, giving a gap of $\Omega(\sqrt{r}) = \Omega(l^{1/3})$ for SHORTEST PATH EDGE INTERDICTION. Generally, $k' \le O(r^\epsilon)$ results in $l' \le O(r^{1+\epsilon})$, giving an $(O(r^\epsilon), O(r^{1/2-\epsilon}))$-bicriteria gap for any $\epsilon \in (0, \frac{1}{2})$.

17.7 SHORT PATH VERTEX CUT

We propose our dictatorship test for SHORT PATH VERTEX CUT that will be used for proving Unique Games hardness. It is parameterized by positive integers a, b, r, R and small $\epsilon > 0$. It is inspired by the integrality gap instances by Baier et al. [BEH+10] Mahjoub and and McCormick [MM10], and made such that the vertex cuts that correspond to *dictators* behave the same as the fractional solution that cuts $\frac{1}{r}$ fraction of every vertex. All graphs in this section are undirected.

For positive integers a, b, r, R, and $\epsilon > 0$, define $\mathcal{D}^{\mathsf{V}}_{a,b,r,R,\epsilon} = (V, E)$ be the graph defined as follows. Consider the probability space (Ω, μ) where $\Omega := \{0, \dots, r-1, *\}$, and $\mu : \Omega \mapsto [0, 1]$ with $\mu(*) = \epsilon$ and $\mu(x) = \frac{1-\epsilon}{r}$ for $x \neq *$.

- $V = \{s, t\} \cup \{v^i_x\}_{0 \le i \le b, x \in \Omega^R}$. Let v^i denote the set of vertices $\{v^i_x\}_x$.
- For $0 \le i \le b$ and $x \in \Omega^R$, $\mathsf{wt}(v^i_x) = \mu^{\otimes R}(x)$. Note that the sum of weights is $b + 1$.
- For any $0 \le i \le b$, there are edges from s to each vertex in v_i with length $ai + 1$ and edges from each vertex in v_i to t with length $(b - i)a + 1$.
- For $x, y \in \Omega^R$, we call that x and y are *compatible* if
  - For any $1 \le j \le R$: $[y_j = (x_j + 1) \mod r]$ or $[y_j = *]$ or $[x_j = *]$.
- For any $0 \le i < b$ and compatible $x, y \in \Omega^R$, we have an edge (v^i_x, v^{i+1}_y) of length 1 (called a *short edge*).
- For any i, j such that $0 \le i < j - 1 < b$ and compatible $x, y \in \Omega^R$, we have an edge (v^i_x, v^j_y) of length $(j - i)a$ (called a *long edge*).

Completeness. We first prove that vertex cuts that correspond to *dictators* behave the same as the fractional solution that gives $\frac{1}{r}$ to every vertex. For any $q \in [R]$, let $V_q := \{v^i_x : 0 \le i \le b, x_q = * \text{ or } 0\}$. Note that the total weight of V_q is $(b + 1)(\epsilon + \frac{1-\epsilon}{r})$.

Lemma 17.7.1. *After removing vertices in V_q, the length of the shortest path is at least $a(b - r + 2)$.*

Proof. Let $p = (s, v^{i_1}_{x^1}, \dots, v^{i_z}_{x^z}, t)$ be a path from s to t where $i_j \in \{0, \dots, b\}$ and $x^j \in \Omega^R$ for each $1 \le j \le z$. Let $y_j := (x^j)_q \in \{0, \dots, r-1\}$ for each $1 \le j \le z$.

For each $1 \le j < z$, the edge $(v^{i_j}_{x^j}, v^{i_{j+1}}_{x^{j+1}})$ is either a long edge or a short edge, and either taken forward (i.e., $i_j < i_{j+1}$) or backward (i.e., $i_j > i_{j+1}$). Let $z_{\mathsf{LF}}, z_{\mathsf{SF}}, z_{\mathsf{LB}}, z_{\mathsf{SB}}$

be the number of long edges taken forward, short edges taken forward, long edges taken backward, and shot edges taken backward, respectively ($z_{\mathsf{LF}} + z_{\mathsf{SF}} + z_{\mathsf{LB}} + z_{\mathsf{SB}} = z - 1$). For $1 \leq j \leq z_{\mathsf{LF}}$ (resp. z_{LB}), consider the jth long edge taken forward (resp. backward) — it is $(v_{x^{j'}}^{i_{j'}}, v_{x^{j'+1}}^{i_{j'+1}})$ for some j'. Let s_j^{F} (resp. s_j^{B}) be $|i_{j'} - i_{j'+1}|$. The following equality holds by observing how i_j changes.

$$i_1 + \sum_{j=1}^{z_{\mathsf{LF}}} s_j^{\mathsf{F}} + z_{\mathsf{SF}} - \sum_{j=1}^{z_{\mathsf{LB}}} s_j^{\mathsf{B}} - z_{\mathsf{SB}} = i_z \quad \Rightarrow \quad i_1 + \sum_{j=1}^{z_{\mathsf{LF}}} s_j^{\mathsf{F}} + z_{\mathsf{SF}} - z_{\mathsf{LB}} - z_{\mathsf{SB}} - i_z \geq 0. \tag{17.3}$$

Consider how y_j changes. Taking any edge forward increases y_j, and taking any edge backward decreases y_j. Since y_j can never be 0 or $*$, we can conclude that

$$z_{\mathsf{LF}} + z_{\mathsf{SF}} - z_{\mathsf{LB}} - z_{\mathsf{SB}} \leq r - 2. \tag{17.4}$$

(17.3) – (17.4) yields

$$i_1 - i_z + \sum_{j=1}^{z_{\mathsf{LF}}} (s_j^{\mathsf{F}} - 1) \geq 2 - r \quad \Rightarrow \quad i_1 - i_z + \sum_{j=1}^{z_{\mathsf{LF}}} s_j^{\mathsf{F}} \geq 2 - r. \tag{17.5}$$

The total length of p is

$$\begin{aligned} & 2 + a\Big(i_1 + b - i_z + \sum_{j=1}^{z_{\mathsf{LF}}} s_j^{\mathsf{F}} + \sum_{j=1}^{z_{\mathsf{LB}}} s_j^{\mathsf{B}}\Big) + z_{\mathsf{SF}} + z_{\mathsf{SB}} \\ \geq \quad & a\Big(i_1 + b - i_z + \sum_{j=1}^{z_{\mathsf{LF}}} s_j^{\mathsf{F}}\Big) \\ \geq \quad & a(b - r + 2). \end{aligned}$$

□

Soundness. To analyze soundness, we define a correlated probability space $(\Omega_1 \times \Omega_2, \nu)$ where both Ω_1, Ω_2 are copies of $\Omega = \{0, \dots, r-1, *\}$. It is defined by the following process to sample $(x, y) \in \Omega^2$.

- Sample $x \in \{0, \dots, r-1\}$. Let $y = (x + 1) \mod r$.
- Change x to $*$ with probability ϵ. Do the same for y independently.

Note that the marginal distribution of both x and y is equal to μ. Assuming $\epsilon < \frac{1}{2r}$, the minimum probability of any atom in $\Omega_1 \times \Omega_2$ is ϵ^2. For any $(x, y) \in \Omega^2$ with nonzero probability, $(*, y)$ and $(*, *)$ also have nonzero probabilities, so we can apply Lemma 3.3.3 to have $\rho(\Omega_1, \Omega_2; \nu) \le \rho := 1 - \epsilon^4/2$.

Apply Theorem 3.3.10 ($\rho \leftarrow \rho, \alpha \leftarrow \epsilon^2, \epsilon \leftarrow \frac{\Gamma_\rho(\frac{\epsilon}{3}, \frac{\epsilon}{3})}{2}$) to get τ and d. We will later apply this theorem with the parameters obtained here. Fix an arbitrary subset $C \subseteq V$, and $C_i := C \cap v^i$. For $0 \le i \le b$, call v^i *blocked* if $\mu^{\otimes R}[C_i(x)] \ge 1 - \epsilon$. At most $\lfloor \frac{\mathsf{wt}(C)}{1-\epsilon} \rfloor$ v^i's can be blocked. Let k' be the number of blocked v^i's, and $z = b + 1 - k'$ be the number of unblocked v^i's. Let $\{v^{i_1}, \ldots, v^{i_z}\}$ be the set of unblocked v^i's with $i_1 < i_2 < \cdots < i_z$.

For $1 \le j \le z$, let $S_j \subseteq v^{i_j}$ be such that $x \in S_j$ if there exists a path $(p_0 = s, p_1, \ldots, p_{j-1}, v_x^{i_j})$ such that each $p_{j'} \in v^{i_{j'}} \setminus C$ ($1 \le j' < j$). For $1 \le j \le z$, let $f_j : \Omega^R \mapsto [0, 1]$ be the indicator function of S_j.

We prove that if none of f_j reveals any *influential coordinate*, $\mu^{\otimes R}(S_z) > 0$. Since any path passing $v^{i_1}, \ldots, v^{i_z}$ (bypassing only blocked v^i's) uses short edges at least $b - 2k'$ times, so the length of the shortest path after removing C is at most $2 + (b - 2k') + 2ak'$.

Lemma 17.7.2. *Suppose that for any* $1 \le j \le z$ *and* $1 \le i \le R$, $\mathsf{Inf}_i^{\le d}[f_j] \le \tau$. *Then* $\mu^{\otimes R}(S_z) > 0$.

Proof. We prove by induction that $\mu^{\otimes R}(S_j) \ge \frac{\epsilon}{3}$. It holds when $j = 1$ since v^{i_1} is unblocked. Assuming $\mu^{\otimes R}(S_j) \ge \frac{\epsilon}{3}$, since S_j does not reveal any influential coordinate, Theorem 3.3.10 shows that for any subset $T_{j+1} \subseteq v^{i_{j+1}}$ with $\mu^{\otimes R}(T_{j+1}) \ge \frac{\epsilon}{3}$, there exists an edge between S_j and T_{j+1}. If $S'_{j+1} \subseteq v^{i_{j+1}}$ is the set of neighbors of S_j, we have $\mu^{\otimes R}(S'_{j+1}) \ge 1 - \frac{\epsilon}{3}$. Since $v^{i_{j+1}}$ is unblocked, $\mu^{\otimes R}(S'_{j+1} \setminus C) \ge \frac{2\epsilon}{3}$, completing the induction. $\square$

In summary, in the completeness case, if we cut vertices of total weight $k := k(a, b, r, \epsilon) = (b+1)(\epsilon + \frac{1-\epsilon}{r})$, the length of the shortest path is at least $l := l(a, b, r, \epsilon) = a(b - r + 2)$. In the soundness case, even after cutting vertices of total weight k', the length of the shortest path is at most $2 + (b - \frac{k'}{1-\epsilon}) + 2a(\frac{k'}{1-\epsilon})$.

- Let $a = 4, b = 2r - 2$ and ϵ small enough so that $k \le 2, l = 4r$. Requiring $l' \ge l$ results in $k' = \Omega(r)$, giving a gap of $\Omega(r) = \Omega(l)$ for LENGTH-BOUNDED VERTEX CUT.

- Let $a = r, b = 2r - 2$ and ϵ small enough so that $k \le 2, l = r^2$. Requiring $k' \le 2$ results in $l' = O(r)$, giving a gap of $\Omega(r) = \Omega(\sqrt{l})$ for SHORTEST PATH

VERTEX INTERDICTION. Generally, $k' \leq O(r^\epsilon)$ results in $l' \leq O(r^{1+\epsilon})$, giving an $(O(r^\epsilon), O(r^{1-\epsilon}))$-bicriteria gap for any $\epsilon \in (0,1)$.

17.8 RMFC

We present our dictatorship test for the RMFC problem. Our test is inspired by the integrality gap example in Chalermsook and Chuzhoy [CC10], which is suggested by Khanna and Olver. This test will be used in Chapter 18 to prove the hardness result based on Conjecture 3.2.4. All graphs in this section are undirected. We will prove hardness of RMFC where $T = \{t\}$ for a single vertex t.

Given positive integers b and R, let $B = (b!)\cdot(\sum_{i=1}^{b} \frac{b!}{i})$, $\Omega = \{*, 1, \ldots, B\}^R$. Consider the probability space (Ω, μ) where $\mu : \Omega \mapsto [0,1]$ with $\mu(*) = \epsilon$ and $\mu(x) = \frac{1-\epsilon}{B}$ for $x \neq *$. We define $\mathcal{D}^{\mathsf{F}}_{b,R,\epsilon} = (V, E)$ as follows.

- $V = \{s, t\} \cup (\{v^i_x\}_{1 \leq i \leq b, x \in \Omega^R})$. Let $v^i := \{v^i_x\}_{x \in \Omega^R}$. The weight a vertex v^i_x is $i \cdot \mu^{\otimes R}(x)$.
- There is an edge from s to each vertex in v^i, from each vertex in v^b to t.
- For $x, y \in \Omega^R$, we call that x and y are *compatible* if
  - For any $1 \leq j \leq R$: $[y_j = x_j]$ or $[y_j = *]$ or $[x_j = *]$.
- For any $0 \leq i < b$ and compatible $x, y \in \Omega^R$, we have an edge (v^i_x, v^{i+1}_y).

Completeness. We first prove that vertex cuts that correspond to *dictators* are efficient. Let $H_i = 1 + \frac{1}{2} + \cdots + \frac{1}{i} = \frac{\sum_{j=1}^{i} \frac{i!}{j}}{i!}$ be the ith harmonic number. For $1 \leq i \leq b$, let $B_i = \frac{H_i}{H_b} B$ and $B_0 = 0$. Each B_i is an integer since $B = (b!) \cdot (\sum_{i=1}^{b} \frac{b!}{i})$, and $\sum_{i=1}^{b} (B_i - B_{i-1}) = B$.

For any $q \in [R]$, we consider the solution where on Day i ($1 \leq i \leq b$), we save

$$V^i_q := \{v^i_x : x_q = * \text{ or } B_{i-1} + 1 \leq x_q \leq B_i\}.$$

Note each day the total weight that the total weight of V_q is $i(\epsilon + \frac{1}{i \cdot H_b}) \leq b\epsilon + \frac{1}{H_b}$.

Lemma 17.8.1. *In above solution, t is never burnt.*

Proof. Fix an arbitrary $p = (s, v^{i_1}_{x^1}, \ldots, v^{i_z}_{x^z}, t)$ from s to t, and let $y_j = (x^j)_q$ $(1 \le j \le z)$. Since $i_j \le j$ for any j, V^i_q is saved before we arrive $v^{i_j}_{x^j}$. Therefore $y_1 = y_2 = \cdots = y_z$. There exists $r \in \{1, \ldots, b\}$ such that $y_1 \in \{B_{r-1}+1, \ldots, B_r\}$. Then p intersects V^r_q. $\square$

Soundness. To analyze soundness, we define a correlated probability space $(\Omega_1 \times \Omega_2, \nu)$ where both Ω_1, Ω_2 are copies of $\Omega = \{*, 1, \ldots, B\}$. It is defined by the following process to sample $(x, y) \in \Omega^2$.

- Sample $x \in \{1, \ldots, B\}$. Let $y = x$.
- Change x to $*$ with probability ϵ. Do the same for y independently.

Note that the marginal distribution of both x and y is equal to μ. Assuming $\epsilon < \frac{1}{2B}$, the minimum probability of any atom in $\Omega_1 \times \Omega_2$ is ϵ^2. Furthermore, in our correlated space, $\nu(x, *) > 0$ for all $x \in \Omega_1$ and $\nu(*, x) > 0$ for all $x \in \Omega_2$. Therefore, we can apply Lemma 3.3.3 to conclude that $\rho(\Omega_1, \Omega_2; \nu) \le \rho := 1 - \frac{\epsilon^4}{2}$. Apply Theorem 3.3.10 ($\rho \leftarrow \rho, \alpha \leftarrow \epsilon^2, \epsilon \leftarrow \frac{\Gamma_\rho(\frac{1}{3}, \frac{1}{3})}{2}$) to get τ and d. We will later apply this theorem with the parameters obtained here.

Fix an arbitrary solution where we save $C_i \subseteq V$ on Day i with $\mathsf{wt}(C_i) \le k'$. Let $S_i \subseteq v^i$ be the set of vertices of v^i burnt at the end of Day i. Let $f_i : \Omega^R \mapsto [0, 1]$ be the indicator function of S_i $(1 \le i \le b)$. We prove that if none of f_i reveals any *influential coordinate*, unless k' is large, $\mu^{\otimes R}(S_i)$ is large for all i, so t will be burnt on Day $b + 1$.

Lemma 17.8.2. *Suppose that for any* $1 \le i \le b$ *and* $1 \le j \le R$, $\mathsf{Inf}^{\le d}_j[f_i] \le \tau$. *If* $k' \le \frac{1}{3}$, $\mu^{\otimes R}(S_i) \ge \frac{1}{3}$ *for all* $1 \le i \le b$.

Proof. We prove by induction on i. It is easy to see $\mu^{\otimes R}(S_1) \ge \frac{1}{3}$ since the $\mathsf{wt}(v^1) = 1$ but $k' \le \frac{1}{3}$. Suppose that the claim holds for i. For any $T \subseteq v^{i+1}$ with $\mu^{\otimes R}(T) \le \frac{1}{3}$, since S_i does not reveal any influential coordinate, Theorem 3.3.10 shows that there exists an edge between S_i and T. It implies that $\mu^{\otimes R}(N(S_i)) \ge \frac{2}{3}$, where $N(S_i) \subseteq v^{i+1}$ denotes the set of neighbors of S_i in v^{i+1}. The total weight of saved vertices up to Day i is at most $ik' \le \frac{i}{3}$. Since $\mathsf{wt}(v^i) = i$, even if all saved vertices are in v^i, $\mu^{\otimes R}(v^i \cap (C_1 \cup \cdots \cup C_i)) \le \frac{1}{3}$. Since $S_{i+1} = N(S_i) \setminus (C_1 \cup \cdots \cup C_i)$, $\mu^{\otimes R}(S_{i+1}) \ge \frac{1}{3}$, the induction is complete. $\square$

In summary, in the completeness case, we save vertices of total weight at most $b\epsilon + \frac{1}{H_b}$ and save t. In the soundness case, we fail to save t unless we spend total weight at least $\frac{1}{3}$ each day. By taking ϵ small enough, the gap becomes $\Omega(\log b)$.

Chapter 18

Reduction from UNIQUE GAMES to Cut Problems

18.1 General Reduction

We now introduce our reduction from UNIQUE GAMES to our problems. Recall that we constructed the following dictatorship tests.

- $\mathcal{D}^{\mathsf{M}}_{r,k,R,\epsilon}$ for DIRECTED MULTICUT, $\mathcal{D}^{\mathsf{M}}_{G',H,h,\ell,R,\epsilon}$ for DIRECTED MULTICUT(H) (let H', H, h, ℓ be a LP gap instance).
- $\mathcal{D}^{\mathsf{global}}_{R,\epsilon}$ for NODE DOUBLE CUT, $\mathcal{D}^{\mathsf{st}}_{a,b,R,\epsilon}$ for s-t NODE DOUBLE CUT.
- $\mathcal{D}^{\mathsf{vc}}_{k,R,\epsilon}$ for VERTEX COVER ON k-PARTITE GRAPHS.
- $\mathcal{D}^{\mathsf{E}}_{a,b,r,R}$ for SHORTEST PATH EDGE CUT, $\mathcal{D}^{\mathsf{V}}_{a,b,r,R,\epsilon}$ for SHORT PATH VERTEX CUT.
- $\mathcal{D}^{\mathsf{F}}_{b,R,\epsilon}$ for RMFC.

Fix a problem, and let $\mathcal{D} = (V_{\mathcal{D}}, E_{\mathcal{D}})$ be the dictatorship test for the problem with the chosen parameters. $\mathcal{D}^{\mathsf{E}}$ is edge-weighted and all others are vertex-weighted, and our reduction will take care of this difference whenever relevant.

Given an instance $\mathcal{L}(B(U_B \cup W_B, E_B), [R], \{\pi(u,w)\}_{(u,w)\in E_B})$ of UNIQUE GAMES, we describe how to reduce it to a graph $G = (V_G, E_G)$. We assign to each vertex $w \in W_B$ a copy of $V_{\mathcal{D}}$ and for each *terminal* of $V_{\mathcal{D}}$, merge all $|W_B|$ copies into one. The merged terminals are

- $\{s,t\}$: NODE DOUBLE CUT, s-t NODE DOUBLE CUT, SHORT PATH EDGE CUT, SHORT PATH VERTEX CUT, and RMFC.

- $\{s_i,t_i\}_{i\in[k]}$: DIRECTED MULTICUT.

- $h(V_H)$: DIRECTED MULTICUT(H).

- VERTEX COVER has no terminal.

For any $w \in W_B, v \in V_{\mathcal{D}}$, the vertex weight of (w,v) is $\frac{\mathsf{wt}(v)}{|W_B|}$, so that the sum of vertex weights (except terminals) is $b+1$ for SHORT PATH VERTEX CUT and $\frac{b(b+1)}{2}$ for RMFC, and r^k for DIRECTED MULTICUT, 4 for NODE DOUBLE CUT, ab for s-t NODE DOUBLE CUT, k for VERTEX COVER ON k-PARTITE GRAPHS.

For a permutation $\sigma : [R] \to [R]$, let $x \circ \sigma := (x_{\sigma(1)}, \ldots, x_{\sigma(R)})$. To describe the set of edges, consider the random process where $u \in U_B$ is sampled uniformly at random, and its two neighbors w^1, w^2 are independently sampled. For each edge $(v^{i_1}_{x^1}, v^{i_2}_{x^2}) \in E_{\mathcal{D}}$, we create an edge $((w_1, v^{i_1}_{x^1\circ\pi(u,w^1)}), (w_2, v^{i_2}_{x^2\circ\pi(u,w^2)}))$. Call this edge is *created by* u. For SHORT PATH EDGE CUT, the weight of each edge is the weight in $\mathcal{D}^{\mathsf{E}}$ times the probability that (u, w_1, w_2) are sampled. The sum of weights is b. For each edge incident on a terminal (i.e., (X, v^i_x) or (v^i_x, X) where $X \in \{s,t\} \cup \{s_i,t_i\}_i$), we add the corresponding edge $(X, (w, v^i_x))$ or $((w, v^i_x), X)$ for each $w \in W_B$. For SHORT PATH EDGE CUT, their wegiths are ∞ as in $\mathcal{D}^{\mathsf{E}}$.

18.2 Completeness

Suppose there exists a labeling l and a subset $W' \subseteq W_B$ with $|W'| \geq (1-\eta)|W_B|$ such that l satisfy every edge incident on W'.

DIRECTED MULTICUT. For every $w \in W'$, we cut the following vertices.

$$\{(w, v^{\alpha}_x) : \alpha \in [r]^k, x_{l(w)} = * \text{ or } 0\}.$$

For $w \notin W'$, we cut every vertex in $\{w\} \times \mathcal{D}$. The total cost is at most $(\epsilon + \frac{1-\epsilon}{r})r^k + \eta r^k \leq r^{k-1}(1 + r\epsilon + r\eta)$. The completeness analysis for the dictatorship test, Lemma 17.2.1, ensures that there is no path from s_i to t_i for any i.

DIRECTED MULTICUT(H). For every $w \in W'$, we cut the following vertices.

$$\{(w, v_x^\alpha) : \alpha \in [r]^k, x_{l(w)} = * \text{ or } x_{l(w)} < \ell(\alpha)\}.$$

For $w \notin W'$, we cut every vertex in $\{w\} \times \mathcal{D}$. The total cost is at most $\mathsf{FRAC} + (1/r + \epsilon + \eta)|V_{G'}^N|$. The completeness analysis for the dictatorship test, Lemma 17.2.7, ensures that there is no path from $h(s)$ to $h(t)$ for any $(s, t) \in E_H$.

s-t NODE DOUBLE CUT. Let $D = (V_D, A_D)$ be the graph constructed in Section 17.3 and I_D be $V_D \setminus \{s, t\}$. For every $w \in W'$, we remove the following vertices.

$$\{(w, v_x^\alpha) : \alpha \in I_D, x_{l(w)} = * \text{ or } 0\}.$$

For $w \notin W'$, we remove every vertex in $\{w\} \times I_{\mathcal{D}}$. The total weight is at most $ab/(b - 2a) + ab\epsilon + ab\eta$. The completeness analysis for the dictatorship test ensures that no vertex in V_G can reach both s and t. The proof of Lemma 17.3.3 works verbatim — for each vertex $(w_j, v_{x_j}^{\alpha_j})$ with $x_j \in \Omega^R$, consider $(x_j)_{l(w_j)}$ in place of $(x_j)_q$.

NODE DOUBLE CUT. Let $D = (V_D, A_D)$ be the graph constructed in Section 17.4 and I_D be $V_D \setminus \{s, t\}$. For every $w \in W'$, we remove the following vertices.

$$\{(w, v_x^\alpha) : \alpha \in I_D, x_{l(w)} = * \text{ or } 0\}.$$

For $w \notin W'$, we remove every vertex in $\{w\} \times I_{\mathcal{D}}$. The total weight is at most $4(1 + \epsilon)/3 + 4\eta$. The completeness analysis for the dictatorship test ensures that no vertex in V_G can reach both s and t. The proof of Lemma 17.4.2 works verbatim — for each vertex $(w_j, v_{x_j}^{\alpha_j})$ with $x_j \in \Omega^R$, consider $(x_j)_{l(w_j)}$ in place of $(x_j)_q$.

VERTEX COVER ON k-PARTITE GRAPHS. For every $w \in W'$, we remove the following vertices.

$$\{(w, v_x^\alpha) : \alpha \in [k], x_{l(w)} = * \text{ or } 0\}.$$

For $w \notin W'$, we remove every vertex in $\{w\} \times V_{\mathcal{D}}$. The total weight is at most $k(1 + \epsilon)/2 + k\eta$. The completeness analysis for the dictatorship test, Lemma 17.5.1, ensures that every edge of G is covered — for each edge $\{(w, v_x^i), (w', v_y^j)\}$, consider $x_{l(w)}$ and $y_{l(w')}$ in place of x_q and y_q.

SHORT PATH EDGE CUT. For every triple (u, w_1, w_2) such that $u \in U_B$ and (u, w_1), $(u, w_2) \in E_B$, we cut the following edges.

$$\{((w_1, v_x^i), (w_2, v_y^{i+1}) : 0 \le i < b, y_{l(w_2)} \ne x_{l(w_1)}{+}1 \mod R \text{ or } (x_{l(w_1)}, y_{l(w_2)}) = (0,1)\}.$$

For $w \notin W'$, we additionally cut every edge incident on $\{w\} \times \mathcal{D}$. The total cost is at most $\frac{2b}{r} + 2\eta b$. The completeness analysis for the dictatorship test ensures that the length of the shortest path is at least $a(b - r + 1)$. The proof of Lemma 17.6.1 works if we have $y_j = x^j_{l(w_j)}$.

SHORT PATH VERTEX CUT. For every $w \in W'$, we cut the following vertices.

$$\{(w, v_x^i) : 0 \le i \le b, x_{l(w)} = * \text{ or } 0\}.$$

For $w \notin W'$, we cut every vertex in $\{w\} \times \mathcal{D}$. The total cost is $(b+1)(\epsilon + \frac{1-\epsilon}{r}) + \eta(b+1)$. The completeness analysis for the dictatorship test ensures that the length of the shortest path is at least $a(b - r + 2)$. The proof of Lemma 17.7.1 works if we have $y_j = x^j_{l(w_j)}$.

RMFC. For $w \in W'$, on Day $i (1 \le i \le b)$, we save every vertex in

$$\{(w, v_x^i) : x_{l(w)} = * \text{ or } B_{i-1} \le x_{l(w)} \le B_i\},$$

where $B_i = \frac{H_i}{H_b} B$. For $w \notin W'$, on Day $i\,(1 \le i \le b)$, we save every vertex in (w, v^i). This ensures that fire never spreads to vertices associated with $w \notin W'$. Each day, the total cost of saved vertices is at most $b\epsilon + \frac{1}{H_b} + b\eta$. The completeness analysis for the dictatorship test ensures that t is saved in this case. The proof of Lemma 17.8.1 works if we have $y_j = x^j_{l(w_j)}$.

18.3 Soundness for Cut / Interdiction Problems

We present the soundness analysis for DIRECTED MULTICUT, s-t NODE DOUBLE CUT, NODE DOUBLE CUT, k-HYPERGRAPH VERTEX COVER, SHORT PATH EDGE CUT, and SHORT PATH VERTEX CUT. The soundness analysis of RMFC is in Section 18.3.1. We first discuss how to extract an influential coordinate for each $u \in U_B$.

DIRECTED MULTICUT. Fix an arbitrary $C \subseteq V_G$ with the total cost k', and consider the graph after cutting vertices in C. Let $\beta > 0$ be another small parameter to be determined later. If $k' \leq k(1-\epsilon)(1-\beta)(r-1)^{k-1}$, we prove that we can decode influential coordinates for many vertices of the UNIQUE GAMES instance.

For each $w \in W_B$, $i \in [k]$, $1 \leq j \leq r^k$, and a sequence $\overline{\alpha} = (\alpha_1, \ldots, \alpha_j) \in ([r]^k)^j$, let $g_{w,i,j,\overline{\alpha}} : \Omega^R \mapsto \{0,1\}$ such that $g_{w,i,j,\overline{\alpha}}(x) = 1$ if and only if there exists a path $p = (s_i, (w_1, v_{x^1}^{\alpha_1}), \ldots, (w_{j-1}, v_{x^{j-1}}^{\alpha_{j-1}}), (w, v_x^{\alpha_j}))$ for some $w_1, \ldots, w_{j-1} \in W_B$ and $x^1, \ldots, x^{j-1} \in \Omega^R$.

For $u \in U_B, 0 \leq j \leq b$, and $\overline{\alpha} \in ([r]^k)^j$, let $f_{u,i,j,\overline{\alpha}} : \Omega_R \mapsto [0,1]$ be such that

$$f_{u,i,j,\overline{\alpha}}(x) = \mathbb{E}_{w \in N(u)}[g_{w,i,j,\overline{\alpha}}(x \circ \pi^{-1}(u,w))],$$

where $N(u)$ is the set of neighbors of u in the UNIQUE GAMES instance.

Let $\underline{\Gamma}(u)$ be the expected weight of $C \cap (\{w\} \times \mathcal{D})$, where w is a random neighbor of u. $\mathbb{E}_u[\underline{\Gamma}(u)] = k' \leq k(1-\epsilon)(1-\beta)(r-1)^{k-1}$, so at least β fraction of u's have $\mathbb{E}_u[\underline{\Gamma}(u)] \leq k(1-\epsilon)(r-1)^{k-1}$. For such u, since any s_i-t_i pair is disconnected, the soundness analysis for the dictatorship test shows that there exists $q \in [R], 1 \leq j \leq r^k, \overline{\alpha}$ such that $\mathsf{Inf}_q^{\leq d}[f_{u,i,j,\overline{\alpha}}] \geq \tau$ (d and τ do not depend on u).

DIRECTED MULTICUT(H). Fix an arbitrary $C \subseteq V_G$ with the total cost k', and consider the graph after cutting vertices in C. Let $\beta > 0$ be another small parameter to be determined later. If $k' \leq \mathsf{Opt}(1-\epsilon)(1-\beta)$, we prove that we can decode influential coordinates for many vertices of the UNIQUE GAMES instance.

For each $w \in W_B$, $s \in V_H^S$, $i \in [k]$, $1 \leq j \leq |V_{G'}^N|$, and a sequence $\overline{\alpha} = (\alpha_1, \ldots, \alpha_j) \in (V_{G'}^N)^j$, let $g_{w,s,j,\overline{\alpha}} : \Omega^R \mapsto \{0,1\}$ such that $g_{w,s,j,\overline{\alpha}}(x) = 1$ if and only if there exists a path $p = (h(s), (w_1, v_{x^1}^{\alpha_1}), \ldots, (w_{j-1}, v_{x^{j-1}}^{\alpha_{j-1}}), (w, v_x^{\alpha_j}))$ for some $w_1, \ldots, w_{j-1} \in W_B$ and $x^1, \ldots, x^{j-1} \in \Omega^R$.

For $u \in U_B, 0 \leq j \leq b$, and $\overline{\alpha} \in (V_{G'}^N)^j$, let $f_{u,s,j,\overline{\alpha}} : \Omega_R \mapsto [0,1]$ be such that

$$f_{u,s,j,\overline{\alpha}}(x) = \mathbb{E}_{w \in N(u)}[g_{w,s,j,\overline{\alpha}}(x \circ \pi^{-1}(u,w))],$$

where $N(u)$ is the set of neighbors of u in the UNIQUE GAMES instance.

Let $\underline{\Gamma}(u)$ be the expected weight of $C \cap (\{w\} \times \mathcal{D})$, where w is a random neighbor of u. $\mathbb{E}_u[\underline{\Gamma}(u)] = k' \leq \mathsf{Opt}(1-\epsilon)(1-\beta)$, so at least β fraction of u's have $\mathbb{E}_u[\underline{\Gamma}(u)] \leq \mathsf{Opt}(1-\epsilon)$. For such u, since any $(h(s), h(t))$ pair is disconnected for each $s, t \in E_H$, the soundness analysis for the dictatorship test shows that there exists $q \in [R], 1 \leq j \leq |V_{G'}^N|, \overline{\alpha}$ such that $\mathsf{Inf}_q^{\leq d}[f_{u,s,j,\overline{\alpha}}] \geq \tau$ (d and τ do not depend on u).

NODE DOUBLE CUT and s-t NODE DOUBLE CUT. Fix an arbitrary $C \subseteq V_G \setminus \{s, t\}$, and consider the graph after removing vertices in C. We will show that if $\mathsf{wt}(C)$ is small and no vertex can reach both s and t, we can decode influential coordinates for many vertices of the UNIQUE GAMES instance. For NODE DOUBLE CUT, since every vertex in V_G has an incoming arc from either s or t, it implies that any solution to NODE DOUBLE CUT must reveal influential coordinates or $\mathsf{wt}(C)$ must be large. Recall the graph $D = (V_D, A_D)$ constructed in Section 17.4 (for NODE DOUBLE CUT) or Section 17.3 (for s-t NODE DOUBLE CUT), and $I_D = V_D \setminus \{s, t\}$.

For each $w \in W_B, r \in \{s, t\}, 1 \le j \le |I_D|$, and a sequence $\overline{\alpha} = (\alpha_1, \ldots, \alpha_j) \in (I_D)^j$, let $g_{w,r,j,\overline{\alpha}} : \Omega^R \to \{0, 1\}$ such that $g_{w,i,j,\overline{\alpha}}(x) = 1$ if and only if there exists a path $p = ((w, v_x^{\alpha_1}), (w_2, v_{x^2}^{\alpha_2}), \ldots, (w_j, v_{x^j}^{\alpha_j}), r)$ for some $w_2, \ldots, w_j \in W_B$ and $x^2, \ldots, x^j \in \Omega^R$.

For $u \in U_B, 1 \le j \le |I_D|$, and $\overline{\alpha} \in (I_D)^j$, let $f_{u,r,j,\overline{\alpha}} : \Omega_R \to [0, 1]$ be such that

$$f_{u,r,j,\overline{\alpha}}(x) = \mathbb{E}_{w \in N(u)}[g_{w,r,j,\overline{\alpha}}(x \circ \pi^{-1}(u, w))],$$

where $N(u)$ is the set of neighbors of u in the UNIQUE GAMES instance.

Let $S := 2(1 - \epsilon)$ (for NODE DOUBLE CUT) or $S := (2a - 1)(1 - \epsilon)$ (for s-t NODE DOUBLE CUT) be the lower bound on the weight in the soundness analysis of the respective dictatorship tests. Let $S' := (1 - \beta)S$ for some $\beta > 0$ that will be determined later, and assume that the total weight of removed vertices is at most S'. Let $\gamma(u)$ be the expected weight of $C \cap (\{w\} \times I_{\mathcal{D}})$, where w is a random neighbor of u. Since the instance of UNIQUE GAMES is biregular,

$$\begin{aligned} \mathbb{E}_{u \in U_B}[\gamma(u)] &= \mathbb{E}_{u \in U_B}[\mathbb{E}_{w \in N(u)}[\mathsf{wt}(C \cap (\{w\} \times I_{\mathcal{D}}))]] \\ &= \mathbb{E}_{w \in W_B}[\mathsf{wt}(C \cap (\{w\} \times I_{\mathcal{D}}))] \le S' = (1 - \beta)S. \end{aligned}$$

Therefore, at least β fraction of u's have $\gamma(u) \le S$. For such u, since no vertex can reach both s and t, the soundness analysis for the dictatorship test shows that there exists $q \in [R], r \in \{s, t\}, 1 \le j \le |I_D|, \overline{\alpha}$ such that $\mathsf{Inf}_q^{\le d}[f_{u,r,j,\overline{\alpha}}] \ge \tau$ (d and τ do not depend on u).

VERTEX COVER ON k-PARTITE GRAPHS. Fix an arbitrary $C \subseteq V_G$, and consider the graph after removing vertices in C. We will show that if $\mathsf{wt}(C)$ is small and every edge is removed, we can decode influential coordinates for many vertices of the UNIQUE GAMES instance.

For each $w \in W_B$ and $j \in [k]$, let $g_{w,j} : \Omega^R \to \{0, 1\}$ such that $g_{w,j}(x) = 1$ if and only if $(w, v_x^j) \notin C$. For $u \in U_B$ and $1 \le j \le [k]$, let $f_{u,j} : \Omega_R \to [0, 1]$ be such that

$$f_{u,j}(x) = \mathbb{E}_{w \in N(u)}[g_{w,j}(x \circ \pi^{-1}(u, w))],$$

where $N(u)$ is the set of neighbors of u in the UNIQUE GAMES instance.

Let $S := (1-\epsilon)(k-1)$ be the lower bound on the weight in the soundness analysis of the dictatorship test. Let $S' := (1-\beta)S$ for some $\beta > 0$ that will be determined later, and assume that the total weight of removed vertices is at most S'. Let $\gamma(u)$ be the expected weight of $C \cap (\{w\} \times V_{\mathcal{D}})$, where w is a random neighbor of u. Since the instance of UNIQUE GAMES is biregular,

$$\begin{aligned} &\mathbb{E}_{u \in U_B}[\gamma(u)] = \mathbb{E}_{u \in U_B}[\mathbb{E}_{w \in N(u)}[\mathsf{wt}(C \cap (\{w\} \times I_{\mathcal{D}}))]] \\ = \ &\mathbb{E}_{w \in W_B}[\mathsf{wt}(C \cap (\{w\} \times I_{\mathcal{D}}))] \leq S' = (1-\beta)S. \end{aligned}$$

Therefore, at least β fraction of u's have $\gamma(u) \leq S$. For such u, since every edge is removed, the soundness analysis for the dictatorship test shows that there exists $q \in [R], 1 \leq j \leq [k]$ such that $\mathsf{Inf}_q^{\leq d}[f_{u,j}] \geq \tau$ (d and τ do not depend on u).

SHORT PATH EDGE CUT. Fix an arbitrary $C \subseteq E_G$ with the total cost k', and consider the graph after cutting edges in C. We will show that if the length of the shortest path is greater than $l' = 2 + b - 4k'\sqrt{r} + 4ak'\sqrt{r}$, we can decode influential coordinates for many vertices of the UNIQUE GAMES instance.

For each $w \in W_B$, $0 \leq j \leq b$, and a sequence $\overline{c} = (c_1, \ldots, c_j) \in \{\mathsf{L}, \mathsf{S}\}^j$, let $g_{w,j,\overline{c}} : \Omega^R \mapsto \{0,1\}$ such that $g_{w,j,\overline{c}}(x) = 1$ if and only if there exists a path $p = (s, p_0 = (w_0, v^0_{x^0}), \ldots, p_{j-1} = (w_{j-1}, v^{j-1}_{x^{j-1}}), p_j = (w, v^j_x))$ for some $w_0, \ldots, w_{j-1} \in W_B$ and $x^0, \ldots, x^{j-1} \in \Omega^R$ such that $(p_{j'-1}, p_{j'})$ is long if and only if $c_{j'} = \mathsf{L}$ for $1 \leq j' \leq j$.

For $u \in U_B, 0 \leq j \leq b$, and $\overline{c} \in \{\mathsf{L}, \mathsf{S}\}^j$, let $f_{u,j,\overline{c}} : \Omega_R \mapsto [0,1]$ be such that

$$f_{u,j,\overline{c}}(x) = \mathbb{E}_{w \in N(u)}[g_{w,j,\overline{c}}(x \circ \pi^{-1}(u,w))],$$

where $N(u)$ is the set of neighbors of u in the UNIQUE GAMES instance.

Let $\underline{\Gamma}(u)$ be the sum of weights of the edges created by u in C. $\mathbb{E}_u[\underline{\Gamma}(u)] = k'$, so at least $\frac{1}{2}$ fraction of u's have $\mathbb{E}_u[\underline{\Gamma}(u)] \leq 2k'$. For such u, since the length of the shortest path is greater than $l' = 2+b-4k'\sqrt{r}+4ak'\sqrt{r}$, the soundness analysis for the dictatorship test shows that there exist $j \in \{0, \ldots, b\}, q \in [R], \overline{c}$ such that $\mathsf{Inf}_q^{\leq d}[f_{u,j,\overline{c}}] \geq \tau$ (d and τ do not depend on u).

SHORT PATH VERTEX CUT. Fix an arbitrary $C \subseteq V_G$ with the total cost k', and consider the graph after cutting vertices in C. We will show that if the length of the shortest path is greater than $l' = 2 + (b - 4k') + 8ak'$, we can decode influential coordinates for many vertices of the UNIQUE GAMES instance.

For each $w \in W_B$, $1 \leq j \leq b$, and a sequence $\bar{i} = (i_1 < \cdots < i_j) \in \{0, \ldots, b\}^j$, let $g_{w,j,\bar{i}} : \Omega^R \mapsto \{0,1\}$ such that $g_{w,j,\bar{i}}(x) = 1$ if and only if there exists a path $p = (s, (w_1, v_{x^1}^{i_1}), \ldots, (w_{j-1}, v_{x^{j-1}}^{i_{j-1}}), (w, v_x^{i_j}))$ for some $w_1, \ldots, w_{j-1} \in W_B$ and $x^1, \ldots, x^{j-1} \in \Omega^R$.

For $u \in U_B, 0 \leq j \leq b$, and $\bar{i} \in \{0, \ldots, b\}^i$, let $f_{u,j,\bar{i}} : \Omega_R \mapsto [0,1]$ be such that

$$f_{u,j,\bar{i}}(x) = \mathbb{E}_{w \in N(u)}[g_{w,j,\bar{i}}(x \circ \pi^{-1}(u,w))],$$

where $N(u)$ is the set of neighbors of u in the UNIQUE GAMES instance.

Let $\underline{\Gamma}(u)$ be the expected weight of $C \cap (\{w\} \times \mathcal{D})$, where w is a random neighbor of u. $\mathbb{E}_u[\underline{\Gamma}(u)] = k'$, so at least $\frac{1}{2}$ fraction of u's have $\mathbb{E}_u[\underline{\Gamma}(u)] \leq 2k'$. For such u, Since the length of the shortest path is greater than $l' = 2 + (b - 4k') + 8ak'$, the soundness analysis for the dictatorship test shows that there exists $q \in [R], 1 \leq j \leq b, \bar{i}$ such that $\mathsf{Inf}_q^{\leq d}[f_{u,j,\bar{i}}] \geq \tau$ (d and τ do not depend on u).

Finishing Up. The above analyses for DIRECTED MULTICUT, DIRECTED MULTICUT(H), s-t NODE DOUBLE CUT, NODE DOUBLE CUT, k-HYPERGRAPH VERTEX COVER, SHORT PATH EDGE CUT, and SHORT PATH VERTEX CUTcan be abstracted as follows. Each vertex $u \in U_B$ is associated with $\{f_{u,h} : \Omega^R \mapsto [0,1]\}_{h \in I}$ for some index set I ($|I|$ is upper bounded by some function of b for SHORT PATH EDGE CUT and SHORT PATH VERTEX CUT, some function of r and k for DIRECTED MULTICUT, some absolute constant for NODE DOUBLE CUT, some function of a and b for s-t NODE DOUBLE CUT, some function on k on VERTEX COVER ON k-PARTITE GRAPHS). For at least β fraction of $u \in U_B$ ($\beta = \frac{1}{2}$ for SHORT PATH EDGE CUT and SHORT PATH VERTEX CUT), there exist $i \in I$ and $q \in [R]$ such that $\mathsf{Inf}_q^{\leq d}[f_{u,i}] \geq \tau$. Set $l(u) = q$ for those vertices. Since

$$\begin{aligned}\mathsf{Inf}_q^{\leq d}(f_{u,i}) &= \sum_{\alpha_q \neq 0, |\alpha| \leq d} \widehat{f_{u,i}}(\alpha)^2 = \sum_{\alpha_q \neq 0, |\alpha| \leq d} (\mathbb{E}_w[\widehat{f_{w,i}}(\pi(u,w)^{-1}(\alpha))]^2) \\ &\leq \sum_{\alpha_q \neq 0, |\alpha| \leq d} \mathbb{E}_w[\widehat{f_{w,i}}(\pi(u,w)^{-1}(\alpha))^2] = \mathbb{E}_w[\mathsf{Inf}_{\pi(u,w)^{-1}(q)}^{\leq d}(f_{w,i})],\end{aligned}$$

at least $\tau/2$ fraction of u's neighbors satisfy $\mathsf{Inf}_{\pi(u,w)^{-1}(q)}^{\leq d}(f_{w,i}) \geq \tau/2$. There are at most $2d/\tau$ coordinates with degree-d influence at least $\tau/2$ for a fixed h, so their union over $i \in I$ yields at most $\frac{2d \cdot |I|}{\tau}$ coordinates. Choose $l(w)$ uniformly at random among those coordinates (if there is none, set it arbitrarily). The above probabilistic strategy satisfies at least $\beta(\frac{\tau}{2})(\frac{\tau}{2d \cdot |I|})$ fraction of all edges. Taking η smaller than this quantity proves the soundness of the reductions.

18.3.1 Soundness for RMFC

Fix an arbitrary solution $C_1, \ldots, C_b \subseteq V$ such that C_i is saved on Day i and the weight of each C_i is at most $k' = \frac{1}{10}$. Suppose that t is saved. We will prove that the UNIQUE GAMES instance admits a good labeling.

For each $w \in W_B$, $1 \le i \le b$, let $g_{w,i} : \Omega^R \mapsto \{0,1\}$ such that $g_{w,i}(x) = 1$ if and only if (w, v_x^i) is burning on Day i. Let Day i^* be the first day where $\mathbb{E}_{w,x}[g_{w,i^*}(x)] \ge \frac{1}{2}$ and $\mathbb{E}_{w,x}[g_{w,i^*+1}(x)] \le \frac{1}{2}$. Such i^* must exist since $\mathbb{E}_{w,x}[g_{w,1}] \ge 1 - k' \ge \frac{1}{2}$ but $\mathbb{E}_{w,x}[g_{w,b}] = 0$. For each $w \in W_B$, let $g_w := g_{w,i^*}$ and let $f_w : \Omega^R \mapsto \{0,1\}$ be such that $f_w(x) = 1$ if and only if there exists (w', x') such that the vertex $(w', v_{x'}^{i^*})$ is burning on Day i and there exists an edge $((w', v_{x'}^{i^*}), (w, v_x^{i^*+1}))$. We must have $\mathbb{E}_{w,x}[f_w(x)] \le \frac{1}{2} + \frac{1}{10} = \frac{3}{5}$, since we can save at most $k' = \frac{1}{10}$ fraction of $\{g_{w,i^*+1}\}_w$ before Day $i^* + 1$.

By an averaging argument, at least $\frac{1}{4}$ fraction of $w \in W_B$ satisfies $\mathbb{E}_x[g_{w,i^*}] \ge \frac{1}{4}$. Call them *heavy* vertices. By the expansion of the UNIQUE GAMES instance, at least $\frac{9}{10}$ fraction of $u \in U_B$ has a heavy neighbor, and at least $\frac{9}{10}$ fraction of $w \in W_B$ has a heavy $w' \in W_B$ such that $(u, w), (u, w') \in E_B$ for some $u \in U_B$ (say w is *reachable* from w').

By Theorem 3.3.10, there exist τ and d such that for each heavy w', if $\mathsf{Inf}_j^{\le d}[g_{w'}] \le \tau$ for all $j \in [R]$, all w reachable from w' should satisfy $\mathbb{E}_x[f_w(x)] \ge \frac{9}{10}$ (say w' *reveals an influential coordinate* if such j exists). At least $\frac{1}{4} - \frac{1}{10} = 0.15$ fraction of w' are heavy and reveal an influential coordinate, since otherwise by the expansion $\mathbb{E}_{w,x}[f_w(x)] \ge (\frac{9}{10})^2 > \frac{3}{5}$.

Another expansion argument ensures that at least $\frac{9}{10}$ fraction of $u \in U_B$ is a neighbor of heavy w with an influential coordinate. Call such u *good* and let $h_u : \Omega^R \mapsto \{0,1\}$ such that $h_u(x) = g_w(x \circ \pi^{-1}(u, w))$. Finally, call $w \in W_B$ *good* if $\mathbb{E}_x[f_w(x)] \le \frac{9}{10}$. Since $\mathbb{E}_{w,x}[f_w(x)] \le \frac{3}{5}$, the fraction of good w is at least $\frac{1}{3}$. Theorem 3.3.10 ensures that if there is $(u, w) \in E_B$ where both u and w are good, there exists $j \in [R]$ such that $\min(\mathsf{Inf}_j^{\le d}[h_u], \mathsf{Inf}_{\pi(u,w)^{-1}(j)}^{\le d}[f_w]) \ge \tau$.

Our labeling strategy for UNIQUE GAMES is as follows. Each good u will get a random label from $\{j : \mathsf{Inf}_j^{\le d}[h_u] \ge \tau\}$, and each good w will get a random label from $\{j : \mathsf{Inf}_j^{\le d}[f_w] \ge \tau\}$. Other vertices get an arbitrary label. Since at least $\frac{9}{10}$ fraction of $u \in U_B$ are good, $\frac{1}{3}$ fraction of $w \in W_B$ are good, and the UNIQUE GAMES instance is biregular, at least $\frac{9}{10} - \frac{2}{3} \ge \frac{1}{5}$ fraction of edges are between good vertices. For each f_w or h_u, the number of coordinates j with degree-d influence at least τ is at most $\frac{d}{\tau}$. Therefore, this strategy satisfies at least $\frac{1}{5} \cdot (\frac{d}{\tau})^2$ fraction of edges in expectation. Taking η smaller than this quantity proves the soundness of the reduction.

18.4 Final Results

Combining our completeness and soundness analyses and taking ϵ and η small enough, we prove our main results.

DIRECTED MULTICUT. It is hard to distinguish the following cases.

- Completeness: There is a cut of weight at most $r^{k-1}(1 + r\epsilon + r\eta)$ that separates every s_i and t_i.
- Soundness: Every multicut must have weight at least $k(1-\epsilon)(1-\beta)(r-1)^{k-1}$.

This immediately implies Theorem 16.2.1 by taking large r and small ϵ, β, η.

DIRECTED MULTICUT(H). It is hard to distinguish the following cases.

- Completeness: There is a cut of weight at most $\mathsf{FRAC} + (1/r + \epsilon + \eta)|V_{G'}^N|$ that separates $h(s)$ and $h(t)$ for every $(s,t) \in E_H$.
- Soundness: Every multicut must have weight at least $\mathsf{Opt}(1-\epsilon)(1-\beta)$.

This immediately implies Theorem 17.2.4 by taking large r and small ϵ, β, η.

s-t NODE DOUBLE CUT. It is hard to distinguish the following cases.

1. Completeness: There is a $\{s,t\}$-double cut of weight at most $ab/(b-2a) + ab\epsilon + ab\eta$.
2. Soundness: There is no $\{s,t\}$-double cut of weight less than $(2a-1)(1-\epsilon)(1-\beta)$.

The gap is

$$\frac{(2a-1)(1-\epsilon)(1-\beta)}{\frac{ab}{b-2a} + ab\epsilon + ab\eta},$$

which approaches to 2 by taking a large, b larger, and ϵ, η, β small. This proves Theorem 16.2.3.

NODE DOUBLE CUT. It is hard to distinguish the following cases.

1. Completeness: There is a $\{s,t\}$-double cut of weight at most $4(1+\epsilon)/3+4\eta$.

2. Soundness: There is no global double cut of weight less than $2(1-\epsilon)(1-\beta)$.

The gap is

$$\frac{2(1-\epsilon)(1-\beta)}{\frac{4(1+\epsilon)}{3}+4\eta},$$

which approaches to 1.5 by taking ϵ, η, β small. This proves Theorem 16.2.4.

VERTEX COVER ON k-PARTITE GRAPHS. It is hard to distinguish the following cases.

1. Completeness: There is a vertex cover of weight at most $k(1+\epsilon)/2+k\eta$.

2. Soundness: There is no vertex cover of weight less than $(k-1)(1-\epsilon)(1-\beta)$.

The gap is

$$\frac{(k-1)(1-\epsilon)(1-\beta)}{\frac{k(1+\epsilon)}{2}+k\eta},$$

which approaches to $2(k-1)/k$ by taking ϵ, η, β small. In particular, it approahces to $4/3$ for $k=3$ and $3/2$ for $k=4$. Take large r and small ϵ, β, η.

SHORT PATH EDGE CUT. It is hard to distinguish the following cases.

- Completeness: There is a cut of weight at most $k := \frac{2b}{r} + 2\eta b$ such that the length of the shortest path after the cut is at least $l := a(b-r+1)$.

- Soundness: For every cut of weight k', the length of the shortest path is at most $l' := 2 + b - 4k'\sqrt{r} + 4ak'\sqrt{r}$.

Setting $a = 4$, $b = 2r-1$ yields $k \le 4$ and $l = 4r$. Since $l' \ge 4r$ implies $k' = \Omega(\sqrt{r})$, we prove the first case of Theorem 16.2.5. Setting $a = \sqrt{r}$ and $b = 2r-1$ yields $k \le 4$ and $l = r^{1.5}$. Since $l' = O(k'r)$, we prove the last two cases of Theorem 16.2.5.

SHORT PATH VERTEX CUT. It is hard to distinguish the following cases.

- Completeness: There is a cut of weight at most $k := (b+1)(\epsilon + \frac{1-\epsilon}{r}) + \eta(b+1)$ such that the length of the shortest path after the cut is at least $l := a(b-r+2)$.
- Soundness: For every cut of weight k', the length of the shortest path is at most $l' := 2 + (b-4k') + 8ak'$.

Setting $a=4$, $b=2r-2$ yields $k \leq 2$ and $l=4r$. Since $l' \geq 4r$ implies $k' = \Omega(r)$, we prove the first case of Theorem 16.2.4. Setting $a=r$ and $b=2r-2$ yields $k \leq 2$ and $l=r^2$. Since $l' = O(k'r)$, we prove the last two cases of Theorem 16.2.4.

RMFC. It is hard to distinguish the following cases.

- Completeness: There is a solution where we save vertices of cost $b\epsilon + \frac{1}{H_b} + b\eta = O(\frac{1}{\log b})$ each day to eventually save t.
- Soundness: Saving vertices of $\frac{1}{10}$ each day cannot save t.

This immediately implies Theorem 16.2.6 by taking small ϵ and η.

Chapter 19

Concluding Remarks: What Now?

19.1 Yet Another Summary

At the risk of being redundant, we briefly summarize the results of this thesis here, with the personal opinions on their contribution in the approximation algorithms literature and future directions.

Part I. In Part I, we studied Constraint Satisfaction Problems. One of the nicest features of CSPs is that each problem in the class is formally defined by simply specifying a predicate $P \subseteq D^k$ for some domain D, so each problem is denoted by CSP(P) for some predicate P.[1] Given a property T for optimization problems (e.g., NP-hardness to compute the exact optimal solution), this formal and simple description of each problem in the class often allows us to give a characterization on P in order for CSP(P) to have the property T. Most notably, the famous CSP dichotomy conjecture [FV98] states that for every CSP(P), deciding whether every constraint can be simultaneously satisfied or not is either in $\mathbf{P}$ or NP-hard. Our results in this part can be interpreted as characterizing CSPs with respect to other natural properties related to approximation algorithms.

- HARD CSP and BALANCE CSP (Chapter 4): We studied two variants of MAX CSP that allow only some assignments to be feasible. Among Boolean CSPs, it was known that only MAX 2-SAT and MAX HORN-SAT admit a *robust algorithm*. For

[1] We note that in the exact CSP literature, it is more general and typical to define a CSP by specifying a *set of predicates* instead of one predicate. Also, there is another way to define problems by specifying *constraint hypergraphs*.

each of HARD CSP and BALANCE CSP on the Boolean domain, our results gave a characterization on the predicates that admit a robust algorithm.

- SYMMETRIC CSP (Chapter 5): We revisited the notion of *approximation resistance* introduced by Austrin and Håstad [AH13]. For general MAX CSP, it was known that the simple necessary conditions to be approximation resistant were not sufficient, and the complete characterization is currently unknown and likely to be technically complicated [KTW14]. We show that modulo a simple analytic conjecture, there is a very simple characterization of approximation resistance if we consider a natural subclass of MAX CSP called SYMMETRIC CSP.

Historically, many previous results on such characterizations studied the exact solvability and used algebraic techniques, but recently there are many approximability results that combine techniques from algebra and convex relaxations [BK12, DKM14, DKK$^+$17]. This synergy between algebra and convex relaxations may broaden our understanding of CSPs.

Part II. In Part II, we studied variants of CSPs, mainly motivated by the intersection of computer science with other fields. These problems can be captured as a CSP, but additional conceptual insights were required in order to apply the traditional tools developed for MAX CSP.

- UNIQUE COVERAGE (Chapter 6): Besides resolving the approximability of a fundamental and practical optimization problem, one of the main messages of this chapter is to *bypass* conjectures to prove optimal hardness results. This work is one of few examples where nontrivial hardness results were proved first assuming the Unique Games Conjecture and the Feige's Random 3SAT Hypothesis, and these assumptions were removed later. While there are not many technical evidences for or against these conjectures, removing these assumptions from their important known consequences will be valuable contribution.

- GRAPH PRICING (Chapter 7): It is one of the problems in this thesis that *could have gone either way*. The simplicity of the current best approximation algorithm [BB07, LBA$^+$07] motivated many researchers to try to improve the current best approximation algorithm, but we finally proved that it is optimal under the Unique Games Conjecture. While we are allowed to give a real value to each vertex, we introduced an intermediate problem that bridges this somewhat continuous problem and the tools developed for the discrete problems.

- LDPC decoding (Chapter 8): While decoding a given message is equivalent to finding the closest codeword and can be captured as an optimization problem, error correcting codes have been studied somewhat separately from other combinatorial optimization problems partially due to their differences in techniques; the design of many error correcting codes involves algebraic or information theoretic tools, while algorithms for other combinatorial optimization problems typically rely on their convex relaxations. There is a decoding algorithm using an LP relaxation [Fel03] known in the literature, and this LP relaxation can be systematically strengthened to the Sherali-Adams or Sum-of-Squares hierarchies. Unfortunately our results prove that these hierarchies will not improve the decoding performance greatly, but another interesting question is to study the relationship between the convex relaxation based algorithms and other classes of algorithms. For example, can the Sum-of-Square hierarchies capture iterative message-passing algorithms used to decode LDPC codes? Could our results on the limitation of the convex hierarchies give insights for proving general hardness of the problem?

Part III. Part III studied the complexity of coloring a hypergraph under the promise that the input hypergraph admits one of the three strong notions of coloring, namely low-discrepancy, rainbow, and strong coloring. While we present nontrivial approximation algorithms that exploit the structure of such colorings, our main contribution in this part is hardness results, showing that under these strong promises, it is still NP-hard to weakly color the hypergraph.

As we emphasized, we unified almost all previous coloring hardness techniques in our *recipe* to prove our results. However, there are many open problems remaining in this direction of *coloring a hypergraph with very strong structures*. This may suggest that we need a new set of tools to close this gap. For example, for K-uniform hypergraphs, we show that coloring a $(K/2)$-rainbow colorable graph will be hard, but it is still consistent with our knowledge that $(K-1)$-rainbow colorable graph is still hard to color. For strong coloring, Brakensiek and Guruswami [BG16] showed that for $t = \lceil 3k/2 \rceil$, it is NP-hard to find a 2-weak coloring of a hypergraph that admits t-strong coloring, and conjectured that the same conclusion holds even when $t = K + 1$. Can we close this gap?

More ambitiously, given a computational task that is NP-hard without the restriction on its instances (e.g., weak coloring in this thesis), can we characterize which promises on instances make the task tractable in polynomial time or still NP-hard? See another work of Brakensiek and Guruswami [BG17a] that formalizes this question in the CSP perspective.

Part IV. Part IV mainly studied H-TRANSVERSAL, and gave some characterization on its approximability depending on H. Even though we showed that it is hard to approximate all 2-connected H, and gave efficient approximation algorithms for k-Star and k-Path, there are some obvious open problems in this direction.

- Can we get $O(\log k)$-approximation whenever H is a tree with k vertices?
- The notion of vertex connectivity is local and does not exactly capture the approximability of H-TRANSVERSAL. For example, when H is one large cycle and a single edge glued at one vertex, H is 1-vertex connected by definition, but the approximability of H-TRANSVERSAL should be closer to that of a large cycle. Can we find or define a natural property of H that captures the approximability of H-TRANSVERSAL more accurately?

Besides the vertex deletion version of H-TRANSVERSAL, there is the edge deletion version where the best approximation ratio $O(k^2)$ and the best hardness ratio $\Omega(k)$ do not match even when H is a clique. Also the packing version of H-TRANSVERSAL, H-PACKING, is less understood that the covering counterpart. It would be interesting to study these equally natural variants in the future.

Part V. Part V proved the improved hardness results for numerous cut problems including DIRECTED MULTICUT, LENGTH-BOUNDED CUT, SHORTEST PATH INTERDICTION, RMFC, bicuts, and double cuts. One notable feature of this part is that all results were achieved by the common framework called length-control dictatorship tests that convert LP gap instances to computational hardness results. This leaves many conceptual open questions.

- Are there any other well-studied cut problems whose hardness can be proved via this framework?
- Is there a way to characterize the subclass of cut problems where LP gap instances imply hardness results (perhaps assuming the UGC)? The work of Chekuri and Madan [CM17] shows that the are some cut problems where the best known approximation ratio is even better than the integrality gap of the standard LP relaxation, so not all known cut problems can belong to this subclass. Unlike CSPs, it seems hard to formally capture all known cut problems in one definition, but giving a sufficient condition that captures most known hardness results would advance our understanding on the approximability of cut problems.

- Computational hardness results based on integrality gaps were proved by Raghavendra [Rag08], Kumar et al. [KMTV11], and Ene et al. [EVW13] for variants of CSPs, but the results in this part is the first of this kind that concerns more structured problems on graphs. Could we hope a similar result for other natural classes of structured graph problems such as connectivity (network design) problems?

19.2 Future Directions Beyond This Thesis

This thesis, like every other thesis in science, only contains the successful results, but there are numerous failed attempts to prove optimal approximabilities during the author's Ph.D. study, and even more problems and techniques that have not been explored. Among them, we collect a few future directions that may be interesting to pursue.

19.2.1 Bypassing the Unique Games Conjecture

As there is not much evidence for the Unique Games Conjecture, it is natural to consider the results based on the Unique Games Conjecture and try to prove them without it. The following is the list of problems where the goal is to prove NP-hardness of approximation, since the corresponding UG-hardness is already known or the UGC is not applicable due to the technical nature of problems.

- MAX k-CSP with perfect completeness: By the result of Chan [Cha13], it is known that MAX k-CSP is NP-hard to approximate within a factor better than $\Theta(\frac{k}{2^k})$, matching the current best algorithm of Charikar et al. [CMM07], and the previous UG-hardness result of Samorodnitsky and Treivsan [ST09, AM09]. However, when the instance is promised to admit an assignment that satisfies every constraint (also known as *perfect completeness*), the best algorithm still achieves $\Omega(\frac{k}{2^k})$-approximation while the best hardness remains at $\tilde{O}(2^{k^{1/3}}/2^k)$ [Hua13]. Closing this gap remains an outstanding open question in the approximability of MAX CSP. Very recently Brakensiek and Guruswami [BG17b] proved a hardness ratio of $O(k^2/2^k)$ assuming the *V-Label Cover Conjecture*.

- MAX HORN-SAT: Guruswami and Zhou [GZ12] proved that given an $(1-\epsilon)$-satisfiable instance of MAX HORN-3-SAT, it is UG-hard to find an assignment satisfying more than $(1-\frac{1}{O(\log(1/\epsilon))})$ of the constraints. Proving the same hardness without relying on the UGC will be interesting. Guruswami and Zhou's result is achieved by constructing a SDP gap instance and applying Raghavendra's result [Rag08] to convert it to UG-hardness. We hope that a more direct and combinatorial reduction from LABEL COVER may bypass the dependence on the UGC. This approach was successful on UNIQUE COVERAGE , which was another result of Guruswami and Zhou that we converted to NP-hardness.

- FEEDBACK VERTEX SET: Under the UGC, there are two different proofs showing that FEEDBACK VERTEX SET does not admit a constant factor approximation algorithm. The first one is given by Guruswami et al. [GMR08] based on the tools for MAX CSP, and the other is given by Svensson [Sve13]. This thesis contains the further simplification of Svensson's proof that inspires length-control dictatorship tests for other problems. We believe that proof of the same statement without the UGC will reveal many applications beyond FEEDBACK VERTEX SET.

19.2.2 Packing and Assignment Problems

Other than CSPs, coloring, covering, and cut problems mainly studied in this thesis, packing problems and assignment (scheduling) problems form other major classes of combinatorial optimization problems that have been actively studied. The following problems are outstanding open problems in these classes where the best approximation ratio and the hardness ratio are far apart. For these problems, tight hardness even assuming the UGC is not known.

- Min-Max and Max-Min Allocation: MINMAX ALLOCATION and MAXMIN ALLOCATION are also fundamental optimization problems that have resisted attempts to understand their approximability. MAXMIN ALLOCATION admit an $O(n^\epsilon)$-approximation algorithm while the best hardness remains at 2 [BCG09, CCK09]. MINMAX ALLOCATION is also known as SCHEDULING UNRELATED PARALLEL MACHINES, and the optimal approximation ratio is between 1.5 and 2 [LST90]. The best hardness results are achieved via a simple reduction from 3-SAT. It would be interesting to see the modern theory of hardness of approximation is applicable to these problems.

- k-SET PACKING and disjoint path problems: Our results for cut and interdiction problems are based on the solid understanding of k-HYPERGRAPH VERTEX COVER, which is UG-hard to approximate within a factor $k - \epsilon$ and NP-hard to approximate within a factor $k - 1 - \epsilon$. k-SET PACKING has a larger gap between algorithms and hardness, where the best algorithm achieves $\frac{k+1}{3}$ [Cyg13] and the best hardness remains at $\Omega(\frac{k}{\log k})$ [HSS06]. MAXIMUM INDEPENDENT SET ON k-REGULAR

GRAPHS is an important special case that is also hard to approximate within a factor $\Omega(\frac{k}{\text{polylog}(k)})$ [AKS09, Cha13]. These hardness results use different techniques. We believe that more unified understanding of k-SET PACKING and its special cases will lead to tighter results for k-SET PACKING itself as well as its important variants.

One special case is the following natural disjoint paths problems: Given an undirected graph G and two pairs of terminals $(s_1, t_1), (s_2, t_2)$, how can we find one path from s_1 to t_1 and as many paths from s_2 to t_2, where all paths must be internally vertex disjoint? What if we want to find the same number of s_1-t_1 paths and s_2-t_2 paths?

19.2.3 FPT Approximation

Like the notion of approximation algorithms, fixed parameter tractable (FPT) algorithms are another natural notion that is designed to cope with NP-hardness of numerous combinatorial optimization problems. While the size of the input n is the only parameter in the traditional combinatorial optimization, a parameterized optimization problem comes with another natural parameter k, and an algorithm is called FPT if it runs in time $f(k) \cdot n^{O(1)}$ for some computable function f. While the FPT literature has mainly focused on the exact optimization, there are many interesting parameterized optimization problems that can be studied using the lens of approximation algorithms. Our results for k-PATH TRANSVERSAL and k-VERTEX SEPARATOR indeed give FPT approximation algorithms parameterized by k. Another interesting problem in this direction is the following fundamental problem discussed in Vazirani's textbook [Vaz01].

EDGE k-CUT

Input: A graph $G = (V, E)$.

Output: $F \subseteq E$ such that the subgraph $(V, E \setminus F)$ has at least k connected components.

Goal: Minimize $|F|$.

When k is a constant, this problem admits $n^{\Theta(k)}$-time algorithm that computes an optimal F [GH94, KS96]. It also admits an exact FPT algorithm whose running time is $f(\mathsf{Opt}) \cdot n^{O(1)}$ [KT11], where Opt denotes the size of the optimal k-cut. In the approximation algorithms literature, there is a 2-approximation algorithm [GBH00] whose running time is $n^{O(1)}$ even for large values of k. Assuming the Small Set Expansion Hypothesis, it is NP-hard to have a $(2-\epsilon)$-approximation algorithm for any $\epsilon > 0$ [Man17]. It would be interesting to study whether $(2-\epsilon)$-approximation for some $\epsilon > 0$ is fixed parameter tractable when parameterized by only k; a simple reduction from CLIQUE shows that it is W[1]-hard to find the exact optimum when parameterized by k, but computing an $(1+\epsilon)$-approximate solution for any $\epsilon > 0$ may be in FPT.

19.2.4 Continuous Problems

Optimization problems are studied in a wide range of academic fields outside theoretical computer science such as machine learning and economics. While the notion of approximation algorithms in the theory community have been mainly associated with combinatorial optimization problems, these fields other motivate numerous *continuous optimization problems*, where the domain of feasible solutions is an inherently continuous set. Some problems, including clustering problems in the Euclidean space, have both continuous and combinatorial flavors. GRAPH PRICING is another example.

Often the approximabilities of these problems are not well understood and the techniques for the combinatorial problems are not easily applicable to these continuous problems. We believe that these continuous optimization problems enrich the field of approximation algorithms not only by giving new, practically relevant problems but also by revealing fundamental algorithmic and complexity-theoretic aspects of optimization that are not easily observable in combinatorial problems. We elaborate on three of my results on these problems that were not included in the thesis.

- Polynomial Optimization: In a recent work with Bhattiprolu, Guruswami, Ghosh and Tulsiani [BGG$^+$16], we studied the problem of maximizing an n-variate degree-d homogeneous polynomial f over the unit sphere. Besides being a natural and fundamental problem in its own right, it has connections to widely studied questions in many other areas including quantum information theory (via Quantum Merlin-Arthur games [BH13, BKS14]), Small Set Expansion Hypothesis and the Unique Games Conjecture (via $2 \to 4$ norm [BBH$^+$12, BKS14]), and tensor decomposition and PCA [BKS15, GM15, MR14, HSS15, MSS16].

  Our results include $\tilde{O}((\frac{n}{q})^{d/2-1})$-approximation algorithms for general polynomials, and $\tilde{O}((\frac{n}{q})^{d/4-1/2})$-approximation algorithms for polynomials with nonnegative coefficients and random polynomials. Our lower bounds do not currently match the upper bounds, and it is an interesting open problem to close this gap.

- Clustering: k-MEANS is one of the most fundamental clustering problems, but its optimal approximation ratio is still not well understood. While the best approximation algorithm achieves 6.357-approximation [ANFSW16], we proved the current best NP-hardness ratio 1.0013, improving over Awasthi et al. [ACKS15]. The bottleneck of proving stronger hardness result is that the problem is defined in the Euclidean space $\mathbb{R}^n$, which makes it difficult to *embed* hard combinatorial problems without much loss. Indeed, our hard instances are based on VERTEX COVER in 3-regular graphs and extremely sparse — each point is in the hypercube $\{0,1\}^d$ and has 1 in exactly two coordinates, and each coordinate has exactly three points with 1 there. This sparsity prevents loss incurred by embedding VERTEX COVER to k-MEANS, even though VERTEX COVER in 3-regular graphs has a good approximation algorithm. We believe that a carefully designed special case of hypergraph vertex cover problem, with lossless embedding technique would yield a better inapproximability result for k-MEANS.

- Sparse Birkhoff-Von Neuman Decomposition: With Janardhan Kulkarni and Mohit Singh [KLS17], we studied the following problem. Given a bipartite graph $G = (V, E)$, and a point $x \in [0,1]^E$ in its matching polytope, find the convex decomposition $x = \sum_{i=1}^{k} \lambda_i M_i$, with minimum k such that $\lambda_i \geq 0$ for each i, $\sum_i \lambda_i = 1$, and each M_i is an integral matching. We proved that an $O(\log k)$-approximation is possible when the optimal sparsity k is constant, and also showed that the problem remains NP-hard in that case.

  This problem has applications in routing and switch scheduling. Indeed, the question of writing the given point as a sparse convex combination of vertices can be asked in any polytope. When the polytope is $[0,1]^n$, we can recover the optimal decomposition when k is constant, but approximabilities for large k is open. We believe that this question is worth further research effort.

Bibliography

[AAC07] Amit Agarwal, Noga Alon, and Moses S. Charikar. Improved approximation for directed cut problems. In *Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing*, STOC '07, pages 671–680, New York, NY, USA, 2007. ACM. 2.5.1, 16.2

[ABG$^+$07] Noga Alon, Béla Bollobás, András Gyárfás, Jenő Lehel, and Alex Scott. Maximum directed cuts in acyclic digraphs. *Journal of Graph Theory*, 55(1):1–13, 2007. 7.1.2, 7.1.2, 7.4.1, 7.5.1, 7.5.1

[ABG13] Per Austrin, Siavosh Benabbas, and Konstantinos Georgiou. Better balance by being biased: a 0.8776-approximation for Max Bisection. In *Proceedings of the 24th annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '13, pages 373–387, 2013. 4.1.1, 4.3.2

[ABLT06] Sanjeev Arora, Béla Bollobás, László Lovász, and Iannis Tourlakis. Proving integrality gaps without knowing the linear program. *Theory of Computing*, 2(1):19–51, 2006. 7.1.2, 7.5.1, 7.5.1, 7.5.5

[ABM12] Per Austrin, Siavosh Benabbas, and Avner Magen. On quadratic threshold CSPs. *Discrete Mathematics & Theoretical Computer Science*, 14(2):205–228, 2012. 5.1.1, 5.1.3, 5.3, 5.3.2

[ABZ06] Adi Avidor, Ido Berkovitch, and Uri Zwick. Improved approximation algorithms for Max NAE-SAT and Max SAT. In Thomas Erlebach and Giuseppe Persinao, editors, *Approximation and Online Algorithms*, volume 3879 of *Lecture Notes in Computer Science*, pages 27–40. 2006. 4.3.2

[ABZ16] David Adjiashvili, Andrea Baggio, and Rico Zenklusen. Firefighting on trees beyond integrality gaps. *arXiv preprint arXiv:1601.00271*, 2016. 2.5.6, 16.2

[AC06] Sanjeev Arora and Eden Chlamtac. New approximation guarantee for chromatic number. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 215–224. ACM, 2006. 1.2.1

[ACG+99] Giorgio Ausiello, Pierluigi Crescenzi, Giorgio Gambosi, Viggo Kann, Alberto Marchetti-Spaccamela, and Marco Protasi. Complexity and approximation: Combinatorial optimization problems and their approximability properties. 1999. 3.1

[ACG+10] Matthew Andrews, Julia Chuzhoy, Venkatesan Guruswami, Sanjeev Khanna, Kunal Talwar, and Lisa Zhang. Inapproximability of edge-disjoint paths and low congestion routing on undirected graphs. *Combinatorica*, 30(5):485–520, 2010. 8.5, 13.2

[ACHS12] Elliot Anshelevich, Deeparnab Chakrabarty, Ameya Hate, and Chaitanya Swamy. Approximability of the firefighter problem. *Algorithmica*, 62(1-2):520–536, 2012. 2.5.6, 2.5.6, 16.2, 16.2

[ACKS15] Pranjal Awasthi, Moses Charikar, Ravishankar Krishnaswamy, and Ali Kemal Sinop. The hardness of approximation of euclidean k-means. In *31st International Symposium on Computational Geometry (SoCG 2015)*, volume 34, pages 754–767. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015. 19.2.4

[ACMM05] Amit Agarwal, Moses Charikar, Konstantin Makarychev, and Yury Makarychev. $O(\sqrt{\log n})$-approximation algorithms for Min UnCut, Min 2CNF deletion, and directed cut problems. In *Proceedings of the 37th annual ACM Symposium on Theory of Computing*, STOC '05, pages 573–581, 2005. 4.5

[ADS12] Sanjeev Arora, Constantinos Daskalakis, and David Steurer. Message-passing algorithms and improved LP decoding. *IEEE Transactions on Information Theory*, 58(12):7260–7271, 2012. 2.2.2, 8.1, 8.6

[AGH14] Per Austrin, Venkatesan Guruswami, and Johan Håstad. $(2+\epsilon)$-SAT is NP-hard. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 1–10. IEEE, 2014. 2.3.1, 9.2.1, 9.2.1, 3, 9.3

[AH05] Geir Agnarsson and Magnús M Halldórsson. Strong colorings of hypergraphs. In *Approximation and Online Algorithms*, pages 253–266. 2005. 9.1

[AH13] Per Austrin and Johan Håstad. On the usefulness of predicates. *ACM Transactions on Computation Theory*, 5(1):1:1–1:24, May 2013. 1.3.1, 2.1.2, 5.1, 5.1.1, 5.1.2, 5.1.2, 5.1.3, 5.2.1, 5.2.1, 5.3.1, 5.4, 9.1, 19.1

[AK13] Per Austrin and Subhash Khot. A characterization of approximation resistance for even k-partite CSPs. In *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*, ITCS '13, pages 187–196, 2013. 1.3.1, 5.1.1

[AKMH96] Noga Alon, Pierre Kelsen, Sanjeev Mahajan, and Ramesh Hariharan. Approximate hypergraph coloring. *Nordic Journal of Computing*, 3(4):425–439, 1996. 2.3, 9.1, 11.3.2, 11.3.3

[AKS09] P. Austrin, S. Khot, and M. Safra. Inapproximability of vertex cover and independent set in bounded degree graphs. In *Proceedings of the 24th annual IEEE Conference on Computational Complexity*, CCC '09, pages 74–80, 2009. 3.3.1, 4.2.1, 4.2.1, 4.2.1, 4.2.1, 12.2.1, 19.2.2

[ALM$^+$98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998. 1.2, 1.2.1, 1.3.1, 3.1

[Alo14] Noga Alon. Personal communication. 2014. 2.3, 9.1

[AM09] Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. *Computational Complexity*, 18(2):249–271, 2009. 2.1.2, 2.1.2, 5.1, 5.1.1, 10.7.3, 19.2.1

[ANFSW16] Sara Ahmadian, Ashkan Norouzi-Fard, Ola Svensson, and Justin Ward. Better guarantees for k-means and euclidean k-median by primal-dual algorithms. *arXiv preprint arXiv:1612.07925*, 2016. 19.2.4

[Aom77] Kazuhiko Aomoto. Analytic structure of Schläfli function. *Nagoya Mathematical Journal*, 68:1–16, 1977. 11.1

[AR98] Yonatan Aumann and Yuval Rabani. An o (log k) approximate min-cut max-flow theorem and approximation algorithm. *SIAM Journal on Computing*, 27(1):291–301, 1998. 1.2.1

[AR06] Konstantin Andreev and Harald Racke. Balanced graph partitioning. *Theory of Computing Systems*, 39(6):929–939, 2006. 12.3

[ARV09] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of ACM*, 56(2), April 2009. 12.3

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of np. *Journal of the ACM*, 45(1):70–122, 1998. 1.2, 1.2.1, 1.3.1, 3.1

[Aus10] P. Austrin. Towards sharp inapproximability for any 2-CSP. *SIAM Journal on Computing*, 39(6):2430–2463, 2010. 2

[AZ06] Matthew Andrews and Lisa Zhang. Logarithmic hardness of the undirected edge-disjoint paths problem. *Journal of the ACM*, 53(5):745–761, 2006. 13.2

[Ban10] Nikhil Bansal. Constructive algorithms for discrepancy minimization. In *Proceedings of the 51st annual IEEE symposium on Foundations of Computer Science*, FOCS '10, pages 3–10. IEEE, 2010. 9.1

[BB07] Maria-Florina Balcan and Avrim Blum. Approximation algorithms and online mechanisms for item pricing. *Theory of Computing*, 3:179–195, 2007. 1.3.1, 2.2.3, 7.1, 19.1

[BB08] Halil Bayrak and Matthew D Bailey. Shortest path network interdiction with asymmetric information. *Networks*, 52(3):133–140, 2008. 2.5.5, 16.2

[BBF95] V. Bafna, P. Berman, and T. Fujito. Constant ratio approximations of the weighted feedback vertex set problem for undirected graphs. In *Proceedings of the 6th International Symposium on Algorithms and Computation*, ISAAC '95, pages 142–151, 1995. 12.3

[BBH$^+$12] Boaz Barak, Fernando G.S.L. Brandao, Aram W. Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 307–326, 2012. 19.2.4

[BCC$^+$10] Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. Detecting high log-densities: an $o(n^{1/4})$ approximation for densest k-subgraph. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 201–210. ACM, 2010. 2.4.2, 12.2.2, 13.5

[BCD97] A. Brandstdt, V. Chepoi, and F. Dragan. Clique r-domination and clique r-packing problems on dually chordal graphs. *SIAM Journal on Discrete Mathematics*, 10(1):109–127, 1997. 12.3

[BCG09] MohammadHossein Bateni, Moses Charikar, and Venkatesan Guruswami. Maxmin allocation via degree lower-bounded arborescences. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, STOC '09, pages 543–552, 2009. 19.2.2

[BCK15] Boaz Barak, Siu On Chan, and Pravesh Kothari. Sum of squares lower bounds from pairwise independence. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, STOC '15, pages 97–106, 2015. 1.2, 5.1.1

[BCK+17] K. Berczi, K. Chandrasekaran, T. Kiraly, E. Lee, and C. Xu. Global and fixed-terminal cuts in digraphs. *Manuscript*, 2017. 1.4, 2.5.2, 2.5.3, 2.5.4, 16.2, 16.2, 16.2, 17.3.1, 17.3.1

[BCV+12] Aditya Bhaskara, Moses Charikar, Aravindan Vijayaraghavan, Venkatesan Guruswami, and Yuan Zhou. Polynomial integrality gaps for strong sdp relaxations of densest k-subgraph. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 388–405. SIAM, 2012. 2.4.2, 12.2.2, 13.5

[BEH+10] Georg Baier, Thomas Erlebach, Alexander Hall, Ekkehard Köhler, Petr Kolman, Ondřej Pangrác, Heiko Schilling, and Martin Skutella. Length-bounded cuts and flows. *ACM Transactions on Algorithms*, 7(1):4:1–4:27, December 2010. Preliminary version in ICALP '06. 2.5.5, 2.5.5, 16.2, 16.2, 16.3, 17.6, 17.7

[BG96] A. Becker and D. Geiger. Optimization of pearl's method of conditioning and greedy-like approximation algorithms for the vertex feedback set problem. *Artificial Intelligence*, 83(1):167–188, 1996. 12.3

[BG16] Joshua Brakensiek and Venkatesan Guruswami. New hardness results for graph and hypergraph colorings. In *Proceedings of the 31st Conference on Computational Complexity*, CCC '16, pages 14:1–14:27, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. 19.1

[BG17a] Joshua Brakensiek and Venkatesan Guruswami. Promise constraint satisfaction: Algebraic structure and a symmetric boolean dichotomy. *arXiv preprint arXiv:1704.01937*, 2017. 19.1

[BG17b] Joshua Brakensiek and Venkatesan Guruswami. The quest for strong inapproximability results with perfect completeness. *Electronic Colloquium on Computational Complexity (ECCC)*, 2017. TR17-080. 19.2.1

[BGG+16] Vijay Bhattiprolu, Mrinalkanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. Weak decoupling, polynomial folds, and approximate optimization over the sphere. *arXiv preprint arXiv:1611.05998*, 2016. 19.2.4

[BGL15] Vijay V. S. P. Bhattiprolu, Venkatesan Guruswami, and Euiwoong Lee. Approximate Hypergraph Coloring under Low-discrepancy and Related Promises. In Naveen Garg, Klaus Jansen, Anup Rao, and José D. P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*, volume 40 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 152–174, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. 1.4

[BGMT12] Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP gaps from pairwise independence. *Theory of Computing*, 8(1):269–289, 2012. 5.1.1, 8.3, 8.3

[BGS98] Mihir Bellare, Oded Goldreich, and Madhu Sudan. Free bits, pcps, and nonapproximability—towards tight results. *SIAM Journal on Computing*, 27(3):804–915, 1998. 1.2, 3.3.1

[BGU14] Louay Bazzi, Badih Ghazi, and Ruediger L Urbanke. Linear programming decoding of spatially coupled codes. *IEEE Transactions on Information Theory*, 60(8):4677–4698, 2014. 8.6

[BGV89] Michael O Ball, Bruce L Golden, and Rakesh V Vohra. Finding the most vital arcs in a network. *Operations Research Letters*, 8(2):73–76, 1989. 2.5.5, 16.2

[BH13] Fernando GSL Brandao and Aram W Harrow. Quantum de finetti theorems under local measurements with applications. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 861–870. ACM, 2013. 19.2.4

[BJH99] Károly Böröczky Jr and Martin Henk. Random projections of regular polytopes. *Archiv der Mathematik*, 73(6):465–473, 1999. 11.1, 11.2.3

[BJK⁺13] Botjan Brear, Marko Jakovac, Jn Katreni, Gabriel Semaniin, and Andrej Taranenko. On the vertex k-path cover. *Discrete Applied Mathematics*, 161(1314):1943 – 1949, 2013. 12.2.1

[BK09] N. Bansal and S. Khot. Optimal long code test with one free bit. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '09, pages 453–462, Oct 2009. 3.2, 3.3.1, 16.2

[BK10] Nikhil Bansal and Subhash Khot. Inapproximability of hypergraph vertex cover and applications to scheduling problems. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming*, ICALP '10, pages 250–261, 2010. 2, 3.3.1, 4.2.2, 9.2.1, 9.3, 12.2.1

[BK12] Libor Barto and Marcin Kozik. Robust satisfiability of constraint satisfaction problems. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, STOC '12, pages 931–940, 2012. 19.1

[BKKS11] Boštjan Brešar, František Kardoš, Ján Katrenič, and Gabriel Semanišin. Minimum k-path vertex cover. *Discrete Applied Mathematics*, 159(12):1189–1195, 2011. 12.2.1

[BKS14] Boaz Barak, Jonathan A Kelner, and David Steurer. Rounding sum-of-squares relaxations. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 31–40. ACM, 2014. 19.2.4

[BKS15] Boaz Barak, Jonathan A Kelner, and David Steurer. Dictionary learning and tensor decomposition via the sum-of-squares method. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 143–151. ACM, 2015. 19.2.4

[Blu94] Avrim Blum. New approximation algorithms for graph coloring. *Journal of the ACM (JACM)*, 41(3):470–516, 1994. 1.2.1

[BM02] David Burshtein and Gadi Miller. Bounds on the performance of belief propagation decoding. *IEEE Transactions on Information Theory*, 48(1):112–122, 2002. 8.1

[BM04] David Burshtein and Gadi Miller. Asymptotic enumeration methods for analyzing ldpc codes. *IEEE Transactions on Information Theory*, 50(6):1115–1131, 2004. 2, 8.1

[BM10] Andrei A. Bulatov and Dániel Marx. Constraint satisfaction problems and global cardinality constraints. *Communications of the ACM*, 53(9):99–106, 2010. 4.1.1

[BNN15] Cristina Bazgan, Andr Nichterlein, and Rolf Niedermeier. A refined complexity analysis of finding the most vital edges for undirected shortest paths. In Vangelis Th. Paschos and Peter Widmayer, editors, *Algorithms and Complexity*, volume 9079 of *Lecture Notes in Computer Science*, pages 47–60. Springer International Publishing, 2015. 2.5.5, 16.2

[Bod94] H. L. Bodlaender. On disjoint cycles. *International Journal of Foundations of Computer Science*, 5(01):59–68, 1994. 12.3

[BP13] A. Bernáth and G. Pap. Blocking optimal arborescences. In *Proceedings of the 16th International Conference on Integer Programming and Combinatorial Optimization (IPCO)*, pages 74–85, 2013. 2.5.3, 2, 16.2

[BPRS13] Béla Bollobás, David Pritchard, Thomas Rothvoß, and Alex Scott. Cover-decomposition and polychromatic numbers. *SIAM Journal on Discrete Mathematics*, 27(1):240–256, 2013. 9.1

[BR94] M. Bellare and J. Rompel. Randomness-efficient oblivious sampling. In *Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '94, pages 276–287, 1994. 6.2.2, 6.2.1

[BS14] Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. *Proceedings of International Congress of Mathematicians (ICM)*, 2014. 3.4.1

[BYGNR98] R. Bar-Yehuda, D. Geiger, J. Naor, and R. Roth. Approximation algorithms for the feedback vertex set problem with applications to constraint satisfaction and bayesian inference. *SIAM Journal on Computing*, 27(4):942–959, 1998. 12.3

[Cam15] Eglantine Camby. Connecting hitting sets and hitting paths in graphs. *Doctoral Thesis*, 2015. 12.2.1

[CC08] Miroslav Chlebík and Janka Chlebíková. Approximation hardness of dominating set problems in bounded degree graphs. *Information and Computation*, 206(11):1264–1275, 2008. 12.2.1, 14.1

[CC10] Parinya Chalermsook and Julia Chuzhoy. Resource minimization for fire containment. In *Proceedings of the Twenty-first Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '10, pages 1334–1349, Philadelphia, PA, USA, 2010. Society for Industrial and Applied Mathematics. 2.5.6, 2.5.6, 16.2, 16.2, 16.3, 17.8

[CCHM12] R. Chitnis, M. Cygan, M. Hajiaghayi, and D. Marx. Directed subset feedback vertex set is fixed-parameter tractable. In *Proceedings of the 39th International Colloquium on Automata, Languages, and Programming*, ICALP '12, pages 230–241, 2012. 12.3

[CCK09] Deeparnab Chakrabarty, Julia Chuzhoy, and Sanjeev Khanna. On allocating goods to maximize fairness. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 107–116. IEEE, 2009. 19.2.2

[CCKK12] Parinya Chalermsook, Julia Chuzhoy, Sampath Kannan, and Sanjeev Khanna. Improved hardness results for profit maximization pricing problems with unlimited supply. In *Proceedings of the 15th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, APPROX '12, pages 73–84. 2012. 6.1

[CD82] HW Corley and Y Sha David. Most vital links and nodes in weighted networks. *Operations Research Letters*, 1(4):157–160, 1982. 2.5.5, 16.2

[CF96] Hui Chen and Alan M Frieze. Coloring bipartite hypergraphs. In *Proceedings of the 5th international conference on Integer Programming and Combinatorial Optimization*, IPCO '96, pages 345–358, 1996. 2.3, 9.1

[CGHW98] F. A. Chudak, M. X. Goemans, D. S. Hochbaum, and D. P. Williamson. A primal-dual interpretation of two 2-approximation algorithms for the feedback vertex set problem in undirected graphs. *Operations Research Letters*, 22(45):111–118, 1998. 12.3

[Cha13] Siu On Chan. Approximation resistance from pairwise independent subgroups. In *Proceedings of the 45th annual ACM Symposium on Theory of Computing*, STOC '13, pages 447–456, 2013. 2.4.1, 3.1, 5.1.1, 7.4, 8.1.1, 8.3, 8.3, 8.3, 8.3, 8.3, 8.3.4, 9.3, 10.1.1, 10.4.3, 12.2.1, 13.1, 13.1.2, 19.2.1, 19.2.2

[Che01] Wing-Sum Cheung. Generalizations of Hölder's inequality. *International Journal of Mathematics and Mathematical Sciences*, 26(1):7–10, 2001. 5.2.2

[CHIS12] Mahdi Cheraghchi, Johan Håstad, Marcus Isaksson, and Ola Svensson. Approximating linear threshold predicates. *ACM Trans. Comput. Theory*, 4(1):2:1–2:31, March 2012. 5.1.1

[CHK13] Rajesh Chitnis, MohammadTaghi Hajiaghayi, and Guy Kortsarz. Fixed-parameter and approximation algorithms: A new look. In *Parameterized and Exact Computation*, pages 110–122. Springer, 2013. 12.3

[Chl07] Eden Chlamtac. Approximation algorithms using hierarchies of semidefinite programming relaxations. In *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*, pages 691–701. IEEE, 2007. 1.2.1

[CK09] Julia Chuzhoy and Sanjeev Khanna. Polynomial flow-cut gaps and hardness of directed cut problems. *Journal of the ACM*, 56(2), 2009. 2.5.1, 16.2

[CKK+06] Shuchi Chawla, Robert Krauthgamer, Ravi Kumar, Yuval Rabani, and D Sivakumar. On the hardness of approximating multicut and sparsest-cut. *Computational Complexity*, 15(2):94–114, 2006. 1.2.1

[CKL01] Maw-Shang Chang, Ton Kloks, and Chuan-Min Lee. In *Graph-Theoretic Concepts in Computer Science*, volume 2204 of *Lecture Notes in Computer Science*, pages 32–43. 2001. 12.3

[CKLN13] Parinya Chalermsook, Shiva Kintali, Richard J. Lipton, and Danupon Nanongkai. Graph pricing problem on bounded treewidth, bounded genus and k-partite graphs. *Chicago Journal of Theoretical Computer Science*, 2013(13), November 2013. 1.3.1, 2.2.3, 7.1

[CKR05] Gruia Calinescu, Howard Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. *SIAM Journal on Computing*, 34(2):358–372, 2005. 15.1, 15.2.2

[CLL+08] J. Chen, Y. Liu, S. Lu, B. O'sullivan, and I. Razgon. A fixed-parameter algorithm for the directed feedback vertex set problem. *Journal of the ACM*, 55(5):21:1–21:19, 2008. 12.3

[CLN13] Parinya Chalermsook, Bundit Laekhanukit, and Danupon Nanongkai. Independent set, induced matching, and pricing: Connections and tight (subexponential time) approximation hardnesses. In *Proceedings of the 54th annual IEEE symposium on Foundations of Computer Science*, FOCS '13, pages 370–379, 2013. 7.1, 7.2.2

[CLN14] Parinya Chalermsook, Bundit Laekhanukit, and Danupon Nanongkai. Pre-reduction graph products: Hardnesses of properly learning DFAs and approximating EDP on DAGs. 2014. 12.3

[CM16] Chandra Chekuri and Vivek Madan. Simple and fast rounding algorithms for directed and node-weighted multiway cut. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 797–807, 2016. 2.5.1, 16.2

[CM17] Chandra Chekuri and Vivek Madan. Approximating multicut and the demand graph. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 855–874. SIAM, 2017. 16.2.3, 16.3, 17.2.1, 17.2.4, 19.1

[CMM06] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for unique games. In *Proceedings of the 38th annual ACM Symposium on Theory of Computing*, STOC '06, pages 205–214, 2006. 7.1.1, 1, 7.7.1

[CMM07] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for maximum constraint satisfaction problems. In *Proceedings of the 18th annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, pages 62–68, 2007. 4.1, 19.2.1

[CMM09] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality gaps for Sherali-Adams relaxations. In *Proceedings of the 41st annual ACM Symposium on Theory of Computing*, STOC '09, pages 283–292, 2009. 7.1.1, 7.1.2, 7.5.1, 7.5.3

[CMM10] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Local global tradeoffs in metric embeddings. *SIAM Journal on Computing*, 39(6):2487–2512, 2010. 7.5.7, 7.5.12

[CMWY09] F. Chataigner, G. Mani, Y. Wakabayashi, and R. Yuster. Approximation algorithms and hardness results for the clique packing problem. *Discrete Applied Mathematics*, 157(7):1396 – 1406, 2009. 12.3

[Coo71] Stephen A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, STOC '71, pages 151–158, New York, NY, USA, 1971. ACM. 1.1

[CPPW11] M. Cygan, M. Pilipczuk, M. Pilipczuk, and J. O. Wojtaszczyk. Subset feedback vertex set is fixed-parameter tractable. In *Proceedings of the 38th International Colloquim on Automata, Languages and Programming*, ICALP '11, pages 449–461, 2011. 12.3

[CSS10] Nadia Creignou, Henning Schnoor, and Ilka Schnoor. Nonuniform boolean constraint satisfaction problems with cardinality constraint. *ACM Transactions on Computational Logic*, 11(4), 2010. 4.1.1

[CT12] Eden Chlamtac and Madhur Tulsiani. Convex relaxations and integrality gaps. In *Handbook on Semidefinite, Conic and Polynomial Optimization*, volume 166 of *International Series in Operations Research and Management Science*, pages 139–169. 2012. 2.2.3, 3.4.1, 7.1

[CV16] Parinya Chalermsook and Daniel Vaz. New integrality gap results for the firefighters problem on trees. In *International Workshop on Approximation and Online Algorithms*, pages 65–77. Springer, 2016. 2.5.6, 16.2

[Cyg13] Marek Cygan. Improved approximation for 3-dimensional matching via bounded pathwidth local search. In *Proceedings of the 54th annual IEEE symposium on Foundations of Computer Science*, FOCS '13, pages 509–518, 2013. 12.3, 19.2.2

[DDKW08] Constantinos Daskalakis, Alexandros G Dimakis, Richard M Karp, and Martin J Wainwright. Probabilistic analysis of linear programming decoding. *IEEE Transactions on Information Theory*, 54(8):3565–3578, 2008. 2.2.2, 8.1

[DDvH14] Pål Grønås Drange, Markus Sortland Dregi, and Pim vant Hof. On the computational complexity of vertex integrity and component order connectivity. In *Algorithms and Computation*, pages 285–297. Springer, 2014. 2.4.2, 12.2.2, 12.3, 13.5

[DFHS08] Erik D Demaine, Uriel Feige, MohammadTaghi Hajiaghayi, and Mohammad R Salavatipour. Combination can be hard: Approximability of the unique coverage problem. *SIAM Journal on Computing*, 38(4):1464–1483, 2008. Preliminary version in SODA '06. 2.2.1, 6.1

[DG13] Irit Dinur and Venkatesan Guruswami. PCPs via low-degree long code and hardness for constrained hypergraph coloring. In *Proceedings of the 54th annual symposium on Foundations of Computer Science*, FOCS 13, pages 340–349, 2013. 1.3.1, 9.1

[DGKR05] Irit Dinur, Venkatesan Guruswami, Subhash Khot, and Oded Regev. A new multilayered PCP and the hardness of hypergraph vertex cover. *SIAM Journal on Computing*, 34(5):1129–1146, 2005. 2.4, 3.1, 4.2.2, 8.1, 9.2.1, 10.2.7, 12.1, 12.2.1, 13.1, 13.1.1

[DJP+94] E. Dahlhaus, D. Johnson, C. Papadimitriou, P. Seymour, and M. Yannakakis. The complexity of multiterminal cuts. *SIAM Journal on Computing*, 23(4):864–894, 1994. 1.2.1, 2.5.2, 16.2

[DK13] I. Dinur and G. Kol. Covering CSPs. In *Proceedings of the 28th annual IEEE Conference on Computational Complexity*, CCC '13, pages 207–218, 2013. 4.1

[DK15] Pavel Dvořák and Dušan Knop. Parametrized complexity of length-bounded cuts and multi-cuts. In Rahul Jain, Sanjay Jain, and Frank Stephan, editors, *Theory and Applications of Models of Computation*, volume 9076 of *Lecture Notes in Computer Science*, pages 441–452. Springer International Publishing, 2015. 2.5.5, 16.2

[DKK+17] Víctor Dalmau, Marcin Kozik, Andrei Krokhin, Konstantin Makarychev, Yury Makarychev, and Jakub Opršal. Robust algorithms with polynomial loss for near-unanimity csps. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 340–357. SIAM, 2017. 19.1

[DKM14] Víctor Dalmau, Andrei Krokhin, and Rajsekar Manokaran. Towards a characterization of constant-factor approximable min csps. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 847–857. SIAM, 2014. 19.1

[DLMS08] Guillermo Durán, Min Chih Lin, Sergio Mera, and Jayme L Szwarcfiter. Algorithms for finding clique-transversals of graphs. *Annals of Operations Research*, 157(1):37–45, 2008. 12.3

[dlVKM07] Wenceslas Fernandez de la Vega and Claire Kenyon-Mathieu. Linear programming relaxations of maxcut. In *Proceedings of the 18th annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, pages 53–61. Society for Industrial and Applied Mathematics, 2007. 7.1.2, 7.5.1

[DMR09] Irit Dinur, Elchanan Mossel, and Oded Regev. Conditional hardness for approximate coloring. *SIAM Journal on Computing*, 39(3):843–873, 2009. 1.2.1, 9.1

[DPT+02] Changyan Di, David Proietti, I. Emre Telatar, Thomas J Richardson, and Rüdiger L Urbanke. Finite-length analysis of low-density parity-check codes on the binary erasure channel. *IEEE Transactions on Information Theory*, 48(6):1570–1579, 2002. 2.2.2, 8.1

[DRS05] Irit Dinur, Oded Regev, and Clifford D. Smyth. The hardness of 3-Uniform hypergraph coloring. *Combinatorica*, 25(1):519–535, 2005. 1.3.1, 9.1

[DS05] I. Dinur and S. Safra. On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, pages 439–485, 2005. 4.2.1, 12.3

[DT97] D. Dor and M. Tarsi. Graph decomposition is NP-complete: A complete proof of Holyer's conjecture. *SIAM Journal on Computing*, 26(4):1166–1187, 1997. 12.1, 12.3

[Edm65] Jack Edmonds. Paths, trees, and flowers. *Canadian Journal of mathematics*, 17(3):449–467, 1965. 1.1

[EGT92] Paul Erdős, Tibor Gallai, and Zsolt Tuza. Covering the cliques of a graph with vertices. *Discrete Mathematics*, 108(1):279–289, 1992. 12.3

[EL75] Paul Erdos and László Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. *Infinite and finite sets*, 10(2):609–627, 1975. 2.3, 9.1

[ENRS99] Guy Even, Joseph Naor, Satish Rao, and Baruch Schieber. Fast approximate graph partitioning algorithms. *SIAM Journal on Computing*, 28(6):2187–2214, 1999. 2.4.2, 12.2.2, 12.3, 15.1, 1, 15.1, 15.2.1

[ENRS00] G. Even, J. Naor, S. Rao, and B. Schieber. Divide-and-conquer approximation algorithms via spreading metrics. *Journal of the ACM*, 47(4):585–616, 2000. 12.3, 15.1, 1

[ENSS98] G. Even, J. Naor, B. Schieber, and M. Sudan. Approximating minimum feedback sets and multicuts in directed graphs. *Algorithmica*, 20:151–174, 1998. 4.4, 12.3, 12.3

[EVW13] Alina Ene, Jan Vondrák, and Yi Wu. Local distribution and the symmetry gap: Approximability of multiway partitioning problems. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 306–325. SIAM, 2013. 2.5.1, 16.1, 16.2, 16.3, 19.1

[Fei98] Uriel Feige. A threshold of ln n for approximating set cover. *Journal of the ACM*, 45(4):634–652, July 1998. 1.2.1, 1, 3.1, 6.1.1

[Fei02] Uriel Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, STOC '02, pages 534–543, 2002. 6.1

[Fel03] Jon Feldman. *Decoding error-correcting codes via linear programming*. PhD thesis, Massachusetts Institute of Technology, 2003. 1.3.1, 2.2.2, 8.1, 8.1.1, 8.6, 19.1

[FHL08] Uriel Feige, MohammadTaghi Hajiaghayi, and James R Lee. Improved approximation algorithms for minimum weight vertex separators. *SIAM Journal on Computing*, 38(2):629–657, 2008. 12.3, 15.1

[FHNN15] Till Fluschnik, Danny Hermelin, André Nichterlein, and Rolf Niedermeier. Fractals for kernelization lower bounds, with an application to length-bounded cut problems. *arXiv preprint arXiv:1512.00333*, 2015. 2.5.5, 16.2

[FJ95] Alan Frieze and Mark Jerrum. Improved approximation algorithms for Max k-Cut and Max Bisection. In *Proceedings of the 4th conference on Integer Programming and Combinatorial Optimization*, IPCO '95, pages 1–13, 1995. 4.3.2

[FK98] Uriel Feige and Joe Kilian. Zero knowledge and the chromatic number. *Journal of Computer and System Sciences*, 57(2):187 – 199, 1998. 3.1

[FL01] Uriel Feige and Michael Langberg. The RPR2 rounding technique for semidefinite programs. In *Proceedings of the 28th International Colloquium on Automata, Languages and Programming*, ICALP '01, pages 213–224, 2001. 4.3.2

[FMS+07] Jon Feldman, Tal Malkin, Rocco A Servedio, Clifford Stein, and Martin J Wainwright. LP decoding corrects a constant fraction of errors. *IEEE Transactions on Information Theory*, 53(1):82–89, 2007. 1.3.1, 8.1, 6

[FS07] Z. Friggstad and M. R. Salavatipour. Approximability of packing disjoint cycles. In *Proceedings of the 18th International conference on Algorithms and Computation*, ISAAC'07, pages 304–315, 2007. 12.3

[FS12] Tomas Feder and Carlos Subi. Packing edge-disjoint triangles in given graphs. *Electronic Colloquium on Computational Complexity (ECCC)*, 2012. TR12-013. 12.3

[FV98] Tomás Feder and Moshe Y Vardi. The computational structure of monotone monadic snp and constraint satisfaction: A study through datalog and group theory. *SIAM Journal on Computing*, 28(1):57–104, 1998. 19.1

[FWK05] Jon Feldman, Martin J Wainwright, and David R Karger. Using linear programming to decode binary linear codes. *IEEE Transactions on Information Theory*, 51(3):954–972, 2005. 1.3.1, 2.2.2, 8.1, 8.1.1

[Gal62] Robert G Gallager. Low-density parity-check codes. *IRE Transactions on Information Theory*, 8(1):21–28, 1962. 2.2.2, 8.1, 2

[GBH00] Nili Guttmann-Beck and Refael Hassin. Approximation algorithms for minimum k-cut. *Algorithmica*, 27(2):198–207, 2000. 19.2.3

[GH94] O. Goldschmidt and D. Hochbaum. A polynomial algorithm for the k-cut problem for fixed k. *Math. Oper. Res.*, 19(1):24–37, Feb 1994. 2.5.2, 2.5.4, 16.2, 16.2, 19.2.3

[GHH+14] Venkatesan Guruswami, Johan Håstad, Prahladh Harsha, Srikanth Srinivasan, and Girish Varma. Super-polylogarithmic hypergraph coloring hardness via low-degree long codes. In *Proceedings of the 46th annual ACM Symposium on Theory of Computing*, STOC '14, 2014. 1.3.1, 9.1

[GHI+13] Daniel Gross, Monika Heinig, Lakshmi Iswara, L William Kazmierczak, Kristi Luttrell, John T Saccoman, and Charles Suffel. A survey of component order connectivity models of graph theoretic networks. *SWEAS Trans. Math*, 12(9):895–910, 2013. 12.3

[GHK+05] Venkatesan Guruswami, Jason D. Hartline, Anna R. Karlin, David Kempe, Claire Kenyon, and Frank McSherry. On profit-maximizing envy-free pricing. In *Proceedings of the 16th annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '05, pages 1164–1173, 2005. 2.2.3, 7.1

[GHM+11] V. Guruswami, J. Håstad, R. Manokaran, P. Raghavendra, and M. Charikar. Beating the random ordering is hard: Every ordering CSP is approximation resistant. *SIAM Journal on Computing*, 40(3):878–914, 2011. 2, 7.1.2

[GHS02] Venkatesan Guruswami, Johan Håstad, and Madhu Sudan. Hardness of approximate hypergraph coloring. *SIAM Journal on Computing*, 31(6):1663–1686, 2002. 1.3.1, 9.1, 9.2.1, 9.3

[GJS76] Michael R Garey, David S. Johnson, and Larry Stockmeyer. Some simplified np-complete graph problems. *Theoretical computer science*, 1(3):237–267, 1976. 1.2.1

[GK04] Venkatesan Guruswami and Sanjeev Khanna. On the hardness of 4-coloring a 3-colorable graph. *SIAM Journal on Discrete Mathematics*, 18(1):30–40, 2004. 1.2.1

[GKS10] Parikshit Gopalan, Subhash Khot, and Rishi Saket. Hardness of reconstructing multivariate polynomials over finite fields. *SIAM Journal on Computing*, 39(6):2598–2621, 2010. 3.1, 10.2.5

[GL14] Venkatesan Guruswami and Euiwoong Lee. Complexity of approximating CSP with balance / hard constraints. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*, ITCS '14, pages 439–448, 2014. 1.4

[GL15a] Badih Ghazi and Euiwoong Lee. LP/SDP hierarchy lower bounds for decoding random LDPC codes. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '15, pages 1326–1342, 2015. 1.4

[GL15b] Venkatesan Guruswami and Euiwoong Lee. Inapproximability of h-transversal/packing. *arXiv preprint arXiv:1506.06302*, 2015. 1, 12.3

[GL15c] Venkatesan Guruswami and Euiwoong Lee. Inapproximability of H-Transversal/Packing. In *Approximation, Randomization, and Combinatorial*

Optimization. Algorithms and Techniques (APPROX/RANDOM 2015), volume 40 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 284–304, Dagstuhl, Germany, 2015. 1.4

[GL15d] Venkatesan Guruswami and Euiwoong Lee. Strong inapproximability results on balanced rainbow-colorable hypergraphs. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '15, pages 822–836, 2015. 1.4, 10.7.2, 10.7.2

[GL16a] Venkatesan Guruswami and Euiwoong Lee. Complexity of approximating CSP with balance/hard constraints. *Theory of Computing Systems*, 59(1):76–98, 2016. 1.4

[GL16b] Venkatesan Guruswami and Euiwoong Lee. Nearly optimal NP-hardness of unique coverage. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '16, pages 1724–1730, 2016. 1.4

[GL16c] Venkatesan Guruswami and Euiwoong Lee. Simple proof of hardness of feedback vertex set. *Theory of Computing*, 12(6):1–11, 2016. 12.3, 16.3

[GL17] B. Ghazi and E. Lee. LP/SDP hierarchy lower bounds for decoding random LDPC codes. *IEEE Transactions on Information Theory*, PP(99):1–1, 2017. 1.4

[GM15] Rong Ge and Tengyu Ma. Decomposing overcomplete 3rd order tensors using sum-of-squares algorithms. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, page 829, 2015. 19.2.4

[GMR08] V. Guruswami, R. Manokaran, and P. Raghavendra. Beating the random ordering is hard: Inapproximability of maximum acyclic subgraph. In *Proceedings of the 49th annual IEEE symposium on Foundations of Computer Science*, FOCS '08, pages 573–582, 2008. 4.4, 12.3, 19.2.1

[GMR$^+$11] Venkatesan Guruswami, Yury Makarychev, Prasad Raghavendra, David Steurer, and Yuan Zhou. Finding almost-perfect graph bisections. In *Proceedings of the 2nd symposium on Innovations in Computer Science*, ICS '11, pages 321–337, 2011. 4.1.1, 4.3.2

[GMT09] Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. Optimal Sherali-Adams gaps from pairwise independence. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 5687 of *Lecture Notes in Computer Science*, pages 125–139. 2009. 8.1.1, 8.3, 8.3, 8.3, 8.3, 8.3, 8.3.2, 8

[GPR00] Venkatesan Guruswami and C. Pandu Rangan. Algorithmic aspects of clique-transversal and clique-independent sets. *Discrete Applied Mathematics*, 100(3):183 – 202, 2000. 12.3

[GPRC+01] Venkatesan Guruswami, C. Pandu Rangan, M. S. Chang, G. J. Chang, and C. K. Wong. The K_r-packing problem. *Computing*, 66(1):79–89, 2001. 12.3

[GS10] Venkatesan Guruswami and Rishi Saket. On the inapproximability of vertex cover on k-partite k-uniform hypergraphs. In *Proceedings of the 37th International Colloquium Conference on Automata, Languages and Programming*, ICALP '10, pages 360–371, 2010. 7.1.2

[GSS15] Venkatesan Guruswami, Sushant Sachdeva, and Rishi Saket. Inapproximability of minimum vertex cover on k-uniform k-partite hypergraphs. *SIAM Journal on Discrete Mathematics*, 29(1):36–58, 2015. 1, 16.3

[GT05] Venkatesan Guruswami and Luca Trevisan. The complexity of making unique choices: Approximating 1-in-k Sat. In *Proceedings of the 8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, APPROX '05, pages 99–110. 2005. 6.1

[GT11] Petr A. Golovach and Dimitrios M. Thilikos. Paths of bounded length and their cuts: Parameterized complexity and algorithms. *Discrete Optimization*, 8(1):72 – 86, 2011. Parameterized Complexity of Discrete Optimization. 2.5.5, 16.2

[GTW13] Anupam Gupta, Kunal Talwar, and David Witmer. Sparsest cut on bounded treewidth graphs: Algorithms and hardness results. In *Proceedings of the 45th annual ACM Symposium on Theory of Computing*, STOC '13, pages 281–290, 2013. 2.2.3, 7.1

[GVY94] Naveen Garg, Vijay V Vazirani, and Mihalis Yannakakis. Multiway cuts in directed and node weighted graphs. In *Automata, Languages and Programming*, pages 487–498. Springer, 1994. 2.5.1, 16.2

[GVY96] Naveen Garg, Vijay V Vazirani, and Mihalis Yannakakis. Approximate max-flow min-(multi) cut theorems and their applications. *SIAM Journal on Computing*, 25(2):235–251, 1996. 1.2.1, 15.1

[GVY04] N. Garg, V. Vazirani, and M. Yannakakis. Multiway cuts in node weighted graphs. *Journal of Algorithms*, 50(1):49–61, 2004. 2.5.4, 16.2

[GW94] Michel Goemans and David Williamson. New $\frac{3}{4}$-approximation algorithms for the maximum satisfiability problem. *SIAM Journal on Discrete Mathematics*, 7(4):656–666, 1994. 4.1, 4.1.1, 4.3.2

[GW95] Michel Goemans and David Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, November 1995. 1.2, 1.2.1, 1.3.1, 1.3.1, 4.1.1, 4.3.2, 7.1.1

[GZ12] Venkatesan Guruswami and Yuan Zhou. Tight bounds on the approximability of almost-satisfiable horn SAT and exact hitting set. *Theory of Computing*, 8(1):239–267, 2012. Preliminary version in SODA '11. 2.2.1, 4.1, 4.2.2, 4.2.2, 6.1, 19.2.1

[Har95] Bert Hartnell. Firefighter! an application of domination. presentation. In *25th Manitoba Conference on Combinatorial Mathematics and Computing, University of Manitoba in Winnipeg, Canada*, 1995. 2.5.6, 16.2

[Hås96] J. Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. In *Proceedings of the 37th Annual Symposium on Foundations of Computer Science*, FOCS '96, pages 627–636, Oct 1996. 3.1

[Hås01] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, July 2001. 1.2, 1.3.1, 2.3, 4.1, 5.1.1, 9.1

[Has05] G Hast. *Beating a random assignment. KTH, Stockholm*. PhD thesis, Ph. D Thesis, 2005. 5.1.1, 5.3

[Hås08] Johan Håstad. Every 2-CSP allows nontrivial approximation. *Computational Complexity*, 17(4):549–566, 2008. 7.1.1

[Hås14] Johan Håstad. On the NP-hardness of Max-Not-2. *SIAM Journal on Computing*, 49:179–193, 2014. 9.3

[HJ06] Mohammad Taghi Hajiaghayi and Kamal Jain. The prize-collecting generalized steiner tree problem via a new approach of primal-dual schema. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 631–640. Society for Industrial and Applied Mathematics, 2006. 13.5, 13.5.1

[HK04] Jonas Holmerin and Subhash Khot. A new PCP outer verifier with applications to homogeneous linear equations and Max-bisection. In *Proceedings of the Thirty-sixth Annual ACM Symposium on Theory of Computing*, STOC '04, pages 11–20, 2004. 4.1

[HKNP05] P. Hell, S. Klein, L.T. Nogueira, and F. Protti. Packing r-cliques in weighted chordal graphs. *Annals of Operations Research*, 138(1):179–187, 2005. 12.3

[HLP52] Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. *Inequalities*. Cambridge university press, 1952. 10.3.5

[Hoc96] Dorit S Hochbaum. *Approximation algorithms for NP-hard problems*. PWS Publishing Co., 1996. 3.1

[Hol02] Jonas Holmerin. Vertex cover on 4-regular hyper-graphs is hard to approximate within $2 - \epsilon$. In *Proceedings of the 34th annual ACM Symposium on Theory of Computing*, STOC '02, pages 544–552, 2002. 9.3

[HSS06] Elad Hazan, Shmuel Safra, and Oded Schwartz. On the complexity of approximating k-set packing. *Computational Complexity*, 15(1):20–39, 2006. 2.4, 12.1, 19.2.2

[HSS15] Samuel B Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-square proofs. In *Proceedings of The 28th Conference on Learning Theory*, pages 956–1006, 2015. 19.2.4

[Hua13] Sangxia Huang. Approximation resistance on satisfiable instances for predicates with few accepting inputs. In *Proceedings of the 45th annual ACM symposium on Symposium on Theory of Computing*, STOC '13, pages 457–466, 2013. 9.3, 19.2.1

[Hua15] Sangxia Huang. *Hardness of Constraint Satisfaction and Hypergraph Coloring: Constructions of Probabilistically Checkable Proofs with Perfect Completeness*. PhD thesis, KTH Royal Institute of Technology, 2015. 1.3.1, 2.3, 9.1

[HZ01] Eran Halperin and Uri Zwick. A unified framework for obtaining improved approximation algorithms for maximum graph bisection problems. In *Proceedings of the 8th conference on Integer Programming and Combinatorial Optimization*, IPCO '01, pages 210–225, 2001. 4.3.2

[IW02] Eitan Israeli and R Kevin Wood. Shortest-path network interdiction. *Networks*, 40(2):97–111, 2002. 2.5.5, 16.2

[JM15] Bart M. P. Jansen and Dániel Marx. Characterizing the easy-to-find subgraphs from the viewpoint of polynomial-time algorithms, kernels, and turing kernels. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '15, pages 616–629, 2015. 12.3

[Kar72] Richard M. Karp. *Reducibility among Combinatorial Problems*, pages 85–103. Springer US, Boston, MA, 1972. 1.1, 1.2.1

[Kat16] Ján Katrenič. A faster FPT algorithm for 3-path vertex cover. *Information Processing Letters*, 116(4):273–278, 2016. 12.2.1

[KBB+07] Leonid Khachiyan, Endre Boros, Konrad Borys, Khaled Elbassioni, Vladimir Gurvich, Gabor Rudolf, and Jihui Zhao. On short paths interdiction problems: Total and node-wise limited interdiction. *Theory of Computing Systems*, 43(2):204–233, 2007. 2.5.5, 16.2

[KH83] D. Kirkpatrick and P. Hell. On the complexity of general graph factor problems. *SIAM Journal on Computing*, 12(3):601–609, 1983. 12.1, 12.3

[Kho02a] Subhash Khot. Hardness results for coloring 3-colorable 3-uniform hypergraphs. In *Proceedings of the 43rd annual IEEE symposium on Foundations of Computer Science*, FOCS '02, pages 23–32. IEEE, 2002. 3.1, 9.3, 10.1.1, 10.2.5, 10.2.4, 10.2.5, 10.2.5, 10.2.6

[Kho02b] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th annual ACM Symposium on Theory of Computing*, STOC '02, pages 767–775, 2002. 1.2, 1.3.1, 3.2, 3.2.2, 5.1.1, 6.1, 12.3, 16.1

[Kho06] Subhash Khot. Ruling out ptas for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM Journal on Computing*, 36(4):1025–1071, 2006. 6.1, 13.5

[Kho10] S. Khot. On the unique games conjecture (invited survey). In *Proceedings of the 25th annual IEEE Conference on Computational Complexity*, CCC '10, pages 99–121. IEEE, 2010. 3.2, 3.3

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for Max-Cut and other 2-variable CSPs?

SIAM Journal on Computing, 37(1):319–357, 2007. 1.2.1, 3.2, 4.1, 7.1.2, 7.4, 7.4.3, 10.7.1, 16.1

[KKMS09] Rohit Khandekar, Tracy Kimbrel, Konstantin Makarychev, and Maxim Sviridenko. On hardness of pricing items for single-minded bidders. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 5687 of *Lecture Notes in Computer Science*, pages 202–216. 2009. 1.3.1, 2.2.3, 7.1, 7.1.2, 7.2.1, 7.2, 7.3, 7.3.2

[KLN10] Guy Kortsarz, Michael Langberg, and Zeev Nutov. Approximating maximum subgraphs without short cycles. *SIAM J. Discrete Math.*, 24(1):255–269, 2010. 12.3

[Klo12] Ton Kloks. Packing interval graphs with vertex-disjoint triangles. *arXiv preprint arXiv:1202.1041*, 2012. 12.3

[KLS00] Sanjeev Khanna, Nathan Linial, and Shmuel Safra. On the hardness of approximating the chromatic number. *Combinatorica*, 20(3):393–415, 2000. 1.2.1

[KLS17] Janardhan Kulkarni, Euiwoong Lee, and Mohit Singh. Minimum birkhoff-von neumann decomposition. In *Proceedings of the 21th International Conference on Integer Programming and Combinatorial Optimization*, IPCO '17, 2017. 19.2.4

[KM10] Andrew King and Gary MacGillivray. The firefighter problem for cubic graphs. *Discrete Mathematics*, 310(3):614–621, 2010. 2.5.6, 16.2

[KMR11] Robert Krauthgamer, Aranyak Mehta, and Atri Rudra. Pricing commodities. *Theoretical Computer Science*, 412(7):602 – 613, 2011. Selected papers from WAOA 2007: 5th Workshop on Approximation and Online Algorithms. 1.3.1, 2.2.3, 7.1

[KMS98a] David Karger, Rajeev Motwani, and Madhu Sudan. Approximate graph coloring by semidefinite programming. *Journal of the ACM*, 45(2):246–265, March 1998. 1.2.1

[KMS98b] David Karger, Rajeev Motwani, and Madhu Sudan. Approximate graph coloring by semidefinite programming. *Journal of the ACM (JACM)*, 45(2):246–265, 1998. 11.3

[KMTV11] Amit Kumar, Rajsekar Manokaran, Madhur Tulsiani, and Nisheeth K. Vishnoi. On LP-based approximability for strict CSPs. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '11, pages 1560–1573. SIAM, 2011. 1.3.2, 7.1.2, 16.1, 1, 16.3, 19.1

[Kne36] Hellmuth Kneser. Der simplexinhalt in der nichteuklidischen geometrie. *Deutsche Math*, 1:337–340, 1936. 11.1

[KNS$^+$07] M. Krivelevich, Z. Nutov, M. R. Salavatipour, J. V. Yuster, and R. Yuster. Approximation algorithms and hardness results for cycle packing problems. *ACM Trans. Algorithms*, 3(4), November 2007. 12.3

[KNS09] Robert Krauthgamer, Joseph Seffi Naor, and Roy Schwartz. Partitioning graphs into balanced components. In *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 942–949. Society for Industrial and Applied Mathematics, 2009. 2.4.2, 12.2.2, 12.3, 15.1

[Knu94] Donald E Knuth. The sandwich theorem. *The Electronic Journal of Combinatorics*, 1(1):A1, 1994. 3.4.3

[KR08] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within $2 - \epsilon$. *Journal of Computer and System Sciences*, 74(3):335–349, 2008. 2.5.1, 3.2, 3.2, 3.2.3, 3.3.1, 4.2.1, 4.2.2, 12.2.1, 12.3, 16.2

[KRU12] Shrinivas Kudekar, Tom Richardson, and Rüdiger Urbanke. Spatially coupled ensembles universally achieve capacity under belief propagation. In *Proceedings of 2012 IEEE International Symposium on Information Theory*, ISIT 2012, pages 453–457, 2012. 2.2.2, 8.1, 3, 8.6

[KS96] David R Karger and Clifford Stein. A new approach to the minimum cut problem. *Journal of the ACM (JACM)*, 43(4):601–640, 1996. 19.2.3

[KS14a] S. Khot and R. Saket. Hardness of finding independent sets in 2-colorable and almost 2-colorable hypergraphs. In *Proceedings of the 25th annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '14, pages 1607–1625, 2014. 1.3.1, 9.3

[KS14b] Subhash Khot and Rishi Saket. Hardness of coloring 2-colorable 12-uniform hypergraphs with $exp(\log^{\Omega(1)} n)$ colors. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 206–215. IEEE, 2014. 2.3, 9.1

[KSTW01] Sanjeev Khanna, Madhu Sudan, Luca Trevisan, and David Williamson. The approximability of constraint satisfaction problems. *SIAM Journal on Computing*, 30(6):1863–1920, 2001. 4.2.2, 8.1.1

[KT11] Ken-ichi Kawarabayashi and Mikkel Thorup. The minimum k-way cut of bounded size is fixed-parameter tractable. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 160–169. IEEE, 2011. 19.2.3

[KT12] Ken-ichi Kawarabayashi and Mikkel Thorup. Combinatorial coloring of 3-colorable graphs. In *Proceedings of the 53rd IEEE Annual Symposium on Foundations of Computer Science*, FOCS '12, pages 68–75. IEEE, 2012. 1.2.1

[KT17] Ken-Ichi Kawarabayashi and Mikkel Thorup. Coloring 3-colorable graphs with less than $n^{1/5}$ colors. *Journal of the ACM*, 64(1):4, 2017. 1.2.1

[KTW14] Subhash Khot, Madhur Tulsiani, and Pratik Worah. A characterization of strong approximation resistance. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 634–643, 2014. 1.3.1, 5.1.1, 19.1

[KV05] Subhash A Khot and Nisheeth K Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into l_1. In *Proceedings of the 46th annual IEEE symposium on Foundations of Computer Science*, FOCS '05, pages 53–62, 2005. 1.2, 2.2.3, 7.1.1

[Las01] Jean B. Lasserre. An explicit exact SDP relaxation for nonlinear 0-1 programs. In *Proceedings of the 8th conference on Integer Programming and Combinatorial Optimization*, IPCO '01, pages 293–303, 2001. 1.2, 3.4.1, 8.1

[Lau03] Monique Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming. *Mathematics of Operations Research*, 28(3):470–496, 2003. 8.1

[LBA$^+$07] Eui-woong Lee, David Buchfuhrer, Lachlan Andrew, Ao Tang, and Steven Low. Pricing in the presence of peering. In *Proceedings of the 45th annual Allerton Conference on Computing, Communications and Control*, Allerton '07, pages 286–291, 2007. 1.3.1, 2.2.3, 7.1, 19.1

[LBA$^+$08] Eui-woong Lee, David Buchfuhrer, Lachlan Andrew, Ao Tang, and Steven Low. Progress on pricing with peering. In *Proceedings of the 42nd annual Coneference on Information Sciences and Systems*, CISS '08, pages 286–291, 2008. 7.1

[LCS02] CM Lee, MS Chang, and SC Sheu. The clique transversal and clique independence of distance hereditary graphs. In *Proceedings of the 19th Workshop on Combinatorial Mathematics and Computation Theory, Taiwan*, pages 64–69, 2002. 12.3

[Lee12] Chuan-Min Lee. Weighted maximum-clique transversal sets of graphs. *ISRN Discrete Mathematics*, 2011, 2012. 12.3

[Lee15] Euiwoong Lee. Hardness of graph pricing through generalized max-dicut. In *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing*, STOC '15, pages 391–399, 2015. 1.4

[Lee17a] Euiwoong Lee. Improved hardness for cut, interdiction, and firefighter problems. In *Proceedings of the 44th International Conference on Automata, Languages and Programming*, ICALP '17, 2017. 1.4

[Lee17b] Euiwoong Lee. Partitioning a graph into small pieces with applications to path transversal. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '17, pages 1546–1558, 2017. 1.4

[Lev73] Leonid A Levin. Universal sequential search problems. *Problemy Peredachi Informatsii*, 9(3):115–116, 1973. 1.1

[LLR95] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. 1.2.1

[LLZ02] Michael Lewin, Dror Livnat, and Uri Zwick. Improved rounding techniques for the Max 2-Sat and Max Di-Cut problems. In *Proceedings of the 9th international conference on Integer Programming and Combinatorial Optimization*, IPCO '02, pages 67–82, 2002. 1.3.1, 4.1.1, 4.3.2, 7.4

[LM12] Shachar Lovett and Raghu Meka. Constructive discrepancy minimization by walking on the edges. In *Proceedings of the 53rd annual IEEE symposium on Foundations of Computer Science*, FOCS 12, pages 61–67, 2012. 9.1

[LNLP78] László Lovász, V Neumann-Lara, and M Plummer. Mengerian theorems for paths of bounded length. *Periodica Mathematica Hungarica*, 9(4):269–276, 1978. 2.5.5, 16.2

[Lov79] László Lovász. On the shannon capacity of a graph. *Information Theory, IEEE Transactions on*, 25(1):1–7, 1979. 2.3, 3.4.3, 9.1

[LR99] Tom Leighton and Satish Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM*, 46(6):787–832, November 1999. 1.2.1

[LRV13] Anne Louis, Prasad Raghavendra, and Santosh Vempala. The complexity of approximating vertex expansion. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 360–369. IEEE, 2013. 12.3

[LS91] László Lovász and Alexander Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1(2):166–190, 1991. 1.2

[LS02] Simon Litsyn and Vladimir Shevelev. On ensembles of low-density parity-check codes: asymptotic distance distributions. *IEEE Transactions on Information Theory*, 48(4):887–908, 2002. 2.2.2, 8.1

[LST90] Jan Karel Lenstra, David B Shmoys, and Éva Tardos. Approximation algorithms for scheduling unrelated parallel machines. *Mathematical programming*, 46(1-3):259–271, 1990. 19.2.2

[LY93] Carsten Lund and Mihalis Yannakakis. The approximation of maximum subgraph problems. In *Automata, Languages and Programming*, volume 700 of *Lecture Notes in Computer Science*, pages 40–51. 1993. 12.3

[LY94] Carsten Lund and Mihalis Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, 41(5):960–981, 1994. 6.1.1

[Mac99] David JC MacKay. Good error-correcting codes based on very sparse matrices. *IEEE Transactions on Information Theory*, 45(2):399–431, 1999. 2.2.2, 8.1

[Man17] Pasin Manurangsi. Inapproximability of maximum biclique problems, minimum k-cut and densest at-least-k-subgraph from the small set expansion hypothesis. *arXiv preprint arXiv:1705.03581*, 2017. 19.2.3

[Mar08] Dániel Marx. Parameterized complexity and approximation algorithms. *The Computer Journal*, 51(1):60–78, 2008. 12.3

[MB01] G. Miller and D. Burshtein. Bounds on the maximum-likelihood decoding error probability of low-density parity-check codes. *IEEE Transactions on Information Theory*, 47(7):2696–2710, Nov 2001. 2.2.2, 8.1

[McD93] Colin McDiarmid. A random recolouring method for graphs and hypergraphs. *Combinatorics, Probability and Computing*, 2(03):363–365, 1993. 2.3, 9.1

[MM10] A Ridha Mahjoub and S Thomas McCormick. Max flow and min cut with bounded-length paths: complexity, algorithms, and approximation. *Mathematical programming*, 124(1-2):271–284, 2010. 2.5.5, 16.2, 16.3, 17.6, 17.7

[MMG89] K. Malik, A. K. Mittal, and S. K. Gupta. The k most vital arcs in the shortest path problem. *Oper. Res. Lett.*, 8(4):223–227, August 1989. 2.5.5, 16.2

[MNRS08] Rajsekar Manokaran, Joseph Seffi Naor, Prasad Raghavendra, and Roy Schwartz. Sdp gaps and ugc hardness for multiway cut, 0-extension, and metric labeling. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 11–20. ACM, 2008. 16.1, 16.3

[MOO10] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Annals of Mathematics*, 171:295–341, 2010. 4.2.1

[MOR+06] Elchanan Mossel, Ryan O'Donnell, Oded Regev, Jeffrey E Steif, and Benny Sudakov. Non-interactive correlation distillation, inhomogeneous markov chains, and the reverse bonami-beckner inequality. *Israel Journal of Mathematics*, 154(1):299–336, 2006. 10.1.1, 10.3

[Mor11] David P Morton. Stochastic network interdiction. *Wiley Encyclopedia of Operations Research and Management Science*, 2011. 2.5.5, 16.2

[Mos10] Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010. 3.3, 3.3.3, 3.3.10, 7.1.2, 7.4.1, 9.3, 10.1.1, 10.1.1, 10.2, 10.4.3, 10.4.4, 10.5.3, 10.5.5, 16.3

[MOS13] Elchanan Mossel, Krzysztof Oleszkiewicz, and Arnab Sen. On reverse hypercontractivity. *Geometric and Functional Analysis*, 23(3):1062–1097, 2013. 10.1.1, 10.3.1, 10.3

[MR14] Dániel Marx and Igor Razgon. Fixed-parameter tractability of multicut parameterized by the size of the cutset. *SIAM Journal on Computing*, 43(2):355–388, 2014. Preliminary version in STOC '11. 19.2.4

[MSS16] Tengyu Ma, Jonathan Shi, and David Steurer. Polynomial-time tensor decompositions with sum-of-squares. In *Proceedings of the 57th annual IEEE symposium on Foundations of Computer Science*, FOCS '16, pages 438–446, 2016. 19.2.4

[MY05] Jun Murakami and Masakazu Yano. On the volume of a hyperbolic and spherical tetrahedron. *Communications in analysis and geometry*, 13(2):379, 2005. 11.1

[NSS95] Moni Naor, Leonard J Schulman, and Aravind Srinivasan. Splitters and near-optimal derandomization. In *Foundations of Computer Science, 1995. Proceedings., 36th Annual Symposium on*, pages 182–191. IEEE, 1995. 15.3

[NZ01] Joseph Naor and Leonid Zosin. A 2-approximation algorithm for the directed multiway cut problem. *SIAM Journal on Computing*, 31(2):477–482, 2001. 2.5.1, 16.2

[O'D14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014. 3.3, 3.3.8, 10.3

[OFN12] K. Okumoto, T. Fukunaga, and H. Nagamochi. Divide-and-conquer algorithms for partitioning hypergraphs and submodular systems. *Algorithmica*, 62(3):787–806, 2012. 16.1

[Ott48] Richard Otter. The number of trees. *Annals of Mathematics*, pages 583–599, 1948. 7.5.1

[OW09] Ryan O'Donnell and Yi Wu. Conditional hardness for satisfiable 3-CSPs. In *Proceedings of the 41st annual ACM Symposium on Theory of Computing*, STOC '09, pages 493–502, 2009. 9.3

[Par00] Pablo A Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000. 1.2, 3.4.1

[PS13] Feng Pan and Aaron Schild. Interdiction problems on planar graphs. In Prasad Raghavendra, Sofya Raskhodnikova, Klaus Jansen, and JosD.P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 8096 of *Lecture Notes in Computer Science*, pages 317–331. Springer Berlin Heidelberg, 2013. 2.5.5, 16.2

[PY91] Christos Papadimitriou and Mihalis Yannakakis. Optimization, approximation, and complexity classes. *Journal of Computer and System Sciences*, 43(3):425 – 440, 1991. 4.3.1

[Rac08] Harald Racke. Optimal hierarchical decompositions for congestion minimization in networks. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, pages 255–264, New York, NY, USA, 2008. ACM. 12.3, 15.5.2

[Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the 40th annual ACM symposium on Theory of computing*, STOC '08, pages 245–254, 2008. 1.2, 1.3.1, 1.3.1, 1.3.2, 3.2, 6.1, 7.1.2, 16.1, 16.3, 19.1, 19.2.1

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998. 1.2, 1.3.1, 3.1

[Rog61] CA Rogers. An asymptotic expansion for certain schläfli functions. *Journal of the London Mathematical Society*, 1(1):78–80, 1961. 11.1

[Rog64] Claude Ambrose Rogers. *Packing and covering*. Number 54. Cambridge University Press, 1964. 11.1, 11.2.3

[Rot14] Thomas Rothvoß. Constructive discrepancy minimization for convex sets. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 140–145. IEEE, 2014. 9.1

[RR09] Dieter Rautenbach and Friedrich Regen. On packing shortest cycles in graphs. *Information Processing Letters*, 109(14):816 – 821, 2009. 12.3

[RS09] Prasad Raghavendra and David Steurer. Integrality gaps for strong SDP relaxations of unique games. In *Proceedings of the 50th annual IEEE symposium on Foundations of Computer Science*, FOCS '09, pages 575–585, 2009. 2.2.3, 7.1.1

[RSU01] Thomas J Richardson, Mohammad Amin Shokrollahi, and Rüdiger L Urbanke. Design of capacity-approaching irregular low-density parity-check codes. *IEEE Transactions on Information Theory*, 47(2):619–637, 2001. 3, 8.6

[RT12] Prasad Raghavendra and Ning Tan. Approximating CSPs with global cardinality constraints using SDP hierarchies. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '12, pages 373–387, 2012. 4.1.1, 4.3.2

[RU01] Thomas J Richardson and Rüdiger L Urbanke. The capacity of low-density parity-check codes under message-passing decoding. *IEEE Transactions on Information Theory*, 47(2):599–618, 2001. 2.2.2, 8.1

[RU08] Tom Richardson and Ruediger Urbanke. *Modern coding theory*. Cambridge University Press, 2008. 8.1

[RV98] Sridhar Rajagopalan and Vijay V Vazirani. Primal-dual rnc approximation algorithms for set cover and covering integer programs. *SIAM Journal on Computing*, 28(2):525–540, 1998. 14.1

[SA90] H. Sherali and W. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3(3):411–430, 1990. 1.2, 2.2.3, 3.4.1, 7.1, 8.1

[Sak14] Rishi Saket. Hardness of finding independent sets in 2-colorable hypergraphs and of satisfiable csps. In *Proceedings of the 29th annual IEEE Conference on Computational Complexity*, CCC '14, pages 78–89, 2014. 9.3

[Sch58] Ludwig Schläfli. On the multiple integral $\int \cdots \int$ dx dy... dz, whose limits are $p1 = a1x + b1y + \cdots + h1z > 0, p2 > 0, ..., pn > 0$, and $x2 + y2 + \cdots + z2 < 1$. *Quart. J. Math*, 2(1858):269–300, 1858. 11.1

[Sch78] Thomas Schaefer. The complexity of satisfiability problems. In *Proceedings of the 10th annual ACM Symposium on Theory of Computing*, STOC '78, pages 216–226, 1978. 2.1.1, 4.1, 5.1.1

[Sch03] Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer, 2003. 1.1, 2.5.3, 16.2

[Sch05] Alexander Schrijver. On the history of combinatorial optimization (till 1960). In G.L. Nemhauser K. Aardal and R. Weismantel, editors, *Discrete Optimization*, volume 12 of *Handbooks in Operations Research and Management Science*, pages 1 – 68. Elsevier, 2005.

[Sch08] G. Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *Proceedings of the 49th annual IEEE symposium on Foundations of Computer Science*, FOCS '08, pages 593–602, Oct 2008. Newer version available at the author's homepage. 1.2, 8.1, 8.1.1, 8.3, 8.3, 8.3

[Sey95] P. D. Seymour. Packing directed circuits fractionally. *Combinatorica*, 15(2):281–288, 1995. 12.3

[SLK14] Erfang Shan, Zuosong Liang, and Liying Kang. Clique-transversal sets and clique-coloring in planar graphs. *European Journal of Combinatorics*, 36:367–376, 2014. 12.3

[Spe85] Joel Spencer. Six standard deviations suffice. *Transactions of the American Mathematical Society*, 289(2):679–706, 1985. 9.1

[SPG13] J Cole Smith, Mike Prince, and Joseph Geunes. Modern network interdiction problems and algorithms. In *Handbook of Combinatorial Optimization*, pages 1949–1987. Springer, 2013. 16.2

[SS94] Michael Sipser and Daniel A Spielman. Expander codes. In *Proceedings of the 35th annual IEEE symposium on Foundations of Computer Science*, FOCS 1994, pages 566–576, 1994. 2.2.2, 8.1

[SS13] S. Sachdeva and R. Saket. Optimal inapproximability for scheduling problems via structural hardness for hypergraph vertex cover. In *Proceedings of the 28th annual IEEE Conference on Computational Complexity*, CCC '13, pages 219–229, 2013. 9.2.1, 9.3

[SSZ04] Michael Saks, Alex Samorodnitsky, and Leonid Zosin. A lower bound on the integrality gap for minimum multicut in directed networks. *Combinatorica*, 24(3):525–530, 2004. 16.3, 17.2, 17.2

[ST09] Alex Samorodnitsky and Luca Trevisan. Gowers uniformity, influence of variables, and PCPs. *SIAM Journal on Computing*, 39(1):323–360, 2009. 10.2.2, 10.2.3, 19.2.1

[Sti08] Henning Stichtenoth. *Algebraic function fields and codes*. GMT 254, Springer-Varlg, Berlin, 2008. 3

[STT07] Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. Tight integrality gaps for Lovasz-Schrijver LP relaxations of vertex cover and max cut. In *Proceedings of the 39th annual ACM Symposium on Theory of Computing*, STOC '07, pages 302–310, 2007. 7.5.1

[Sve13] Ola Svensson. Hardness of vertex deletion and project scheduling. *Theory of Computing*, 9(24):759–781, 2013. Preliminary version in APPROX '12. 2, 4.4, 4.4.1, 7.1.2, 12.3, 16.3, 19.2.1

[Tod12] Michael J. Todd. Cornell ORIE6327: Semidefinite programming. *Lecture 8-10*, 2012. 3.4.3

[Tre05] Luca Trevisan. Inapproximability of combinatorial optimization problems. *Paradigms of Combinatorial Optimization, 2nd Edition*, pages 381–434, 2005. 3.1

[Tre12] Luca Trevisan. On khots unique games conjecture. *Bull. Amer. Math. Soc.(NS)*, 49(1):91–111, 2012. 3.2

[TSSW00] Luca Trevisan, Gregory B Sorkin, Madhu Sudan, and David P Williamson. Gadgets, approximation, and linear programming. *SIAM Journal on Computing*, 29(6):2074–2097, 2000. 1.2.1

[Tul09] Madhur Tulsiani. CSP gaps and reductions in the Lasserre hierarchy. In *Proceedings of the 41st annual ACM Symposium on Theory of Computing*, STOC '09, pages 303–312, 2009. 1.2, 5.1.1, 8.1.1, 8.3, 8.3, 8.3, 8.3, 8.5

[Tur41] Paul Turán. On an extremal problem in graph theory. *Mat. Fiz. Lapok*, 48(436-452):137, 1941. 13.3.2

[Tuz91] Zsolt Tuza. Covering all cliques of a graph. In S.T. Hedetniemi, editor, *Topics on Domination*, volume 48 of *Annals of Discrete Mathematics*, pages 117 – 126. Elsevier, 1991. 12.3

[TV15] L. Tseng and N. Vaidya. Fault-Tolerant Consensus in Directed Graphs. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing (PODC 2015)*, pages 451–460, 2015. 2.5.3, 16.2

[TZ11] Jianhua Tu and Wenli Zhou. A factor 2 approximation algorithm for the vertex cover p3 problem. *Information Processing Letters*, 111(14):683–686, 2011. 12.2.1

[Vaz01] Vijay V. Vazirani. Approximation algorithms. Springer, 2001. 1.2.1, 3.1, 19.2.3

[Vid13] Michael Viderman. LP decoding of codes with expansion parameter above 2/3. *Information Processing Letters*, 113(7):225–228, 2013. 6

[Wen13] Cenny Wenner. Circumventing d-to-1 for approximation resistance of satisfiable predicates strictly containing parity of width four. *Theory of Computing*, 9(23):703–757, 2013. 3.1, 1, 9.3, 10.1.1, 10.1.1, 10.2.1, 10.2.1, 1, 10.4.3, 10.4.3, 10.4.3, 10.4.3, 17.6, 17.6.2

[Wen14] Cenny Wenner. Parity is positively useless. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques: The 17th. International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, pages 433–448. Schloss Dagstuhl, 2014. 9.1

[Wig83] Avi Wigderson. Improving the performance guarantee for approximate graph coloring. *Journal of the ACM (JACM)*, 30(4):729–735, 1983. 1.2.1, 11.3.2

[Wil13] Ryan Williams. Stanford CS266: Parameterized algorithms and complexity. *Lecture 4*, 2013. 15.3

[WS11] David P Williamson and David B Shmoys. *The design of approximation algorithms*. Cambridge university press, 2011. 1.2.1, 3.1

[Ye01] Yinyu Ye. A .699-approximation algorithm for Max-Bisection. *Mathematical Programming*, 90(1):101–111, 2001. 4.3.2

[Yus07] Raphael Yuster. Combinatorial and computational aspects of graph packing and graph decomposition. *Computer Science Review*, 1(1):12 – 26, 2007. 12.3

[Yus14] R. Yuster. Edge-disjoint cliques in graphs with high minimum degree. *SIAM Journal on Discrete Mathematics*, 28(2):893–910, 2014. 12.3

[YZ14] Yuichi Yoshida and Yuan Zhou. Approximation schemes via sherali-adams hierarchy for dense constraint satisfaction problems and assignment problems. In *Proceedings of the 5th conference on Innovations in Theoretical Computer Science*, pages 423–438. ACM, 2014. 2.2.3, 7.1

[Zwi98a] Uri Zwick. Approximation algorithms for constraint satisfaction problems involving at most three variables per constraint. In *Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '98, pages 201–210, 1998. 5.1.1, 11.1

[Zwi98b] Uri Zwick. Finding almost-satisfying assignments. In *Proceedings of the 30th annual ACM Symposium on Theory of Computing*, STOC '98, pages 551–560, 1998. 4.1, 4.2.2