Convex Relaxations for Graph and Inverse Eigenvalue Problems

Thesis by
Utkan Onur Candogan

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2020
Defended August 15th, 2019

Utkan Onur Candogan
ORCID: 0000-0002-1416-4909

ACKNOWLEDGEMENTS

I have been extremely lucky to have Venkat as an advisor during my PhD. Thanks to his support, I have been able to explore my interests both within and outside research, and navigate my PhD and my career in a way that best fits my desires. His endless confidence in me has been an invaluable source of motivation. Thank you Venkat, for your guidance in materializing this thesis. Evidently, it will stay as one of my proudest accomplishments throughout my life.

I am grateful to my committee members, Babak Hassibi, Steven Low, P. P. Vaidyanathan and Adam Wierman particularly for being inspiring teachers and researchers. Special thanks to Professors Mathieu Desbrun and Adam Wierman for being accessible and nurturing a friendly environment in Annenberg.

A big thank you to my close friends Andrea Coladangelo, Umut Dincer, Dimitar Ho, Brian Hong, Sinan Kefeli, Sahin Lale, Hoang Le, Jenish Mehta, Oguzhan Teke, Nina Tompkin, Zeynep Turan, and Hikmet Yildiz. Your friendship has definitely made the journey fun. I will certainly miss the backyard bonfires.

Very special thanks to my girlfriend, Corina Panda, for being the kindest and sweetest person I have ever known. Your endless support through this adventure has been unparalleled. Thank you so much for putting up with my terrible jokes :)

I am indebted to my loving family. A big thank you to my parents Siddika and Zafer, for placing my and my brother's education in front of everything else. It is not a coincidence that soon both of your sons will have PhDs. Your unconditional love and sacrifices have made everything possible. A huge thank you to my brother Ozan, who has been my primary inspiration for starting graduate school. Your support and guidance throughout the journey have been extremely helpful, even relieving at times.

Reflecting back on my previous several years, I feel extremely lucky and grateful to have gone through this unique experience. Without a doubt, I got much more out of this journey than I could have anticipated when I started in 2013.

Having successfully completed this ambitious piece of work fills my heart with confidence and optimism for the bright days ahead.

ABSTRACT

This thesis is concerned with presenting convex optimization based tractable solutions for three fundamental problems:

1. *Planted subgraph problem*: Given two graphs, identifying the subset of vertices of the larger graph corresponding to the smaller one.

2. *Graph edit distance problem*: Given two graphs, calculating the number of edge/vertex additions and deletions required to transform one graph into the other.

3. *Affine inverse eigenvalue problem*: Given a subspace $\mathcal{E} \subset \mathbb{S}^n$ and a vector of eigenvalues $\lambda \in \mathbb{R}^n$, finding a symmetric matrix with spectrum λ contained in $\mathcal{E}$.

These combinatorial and algebraic problems frequently arise in various application domains such as social networks, computational biology, chemoinformatics, and control theory. Nevertheless, exactly solving them in practice is only possible for very small instances due to their complexity. For each of these problems, we introduce convex relaxations which succeed in providing exact or approximate solutions in a computationally tractable manner.

Our relaxations for the two graph problems are based on convex graph invariants, which are functions of graphs that do not depend on a particular labeling. One of these convex relaxations, coined the Schur-Horn orbitope, corresponds to the convex hull of all matrices with a given spectrum, and plays a prominent role in this thesis. Specifically, we utilize relaxations based on the Schur-Horn orbitope in the context of the planted subgraph problem and the graph edit distance problem. For both of these problems, we identify conditions under which the Schur-Horn orbitope based relaxations exactly solve the corresponding problem with overwhelming probability. Specifically, we demonstrate that these relaxations turn out to be particularly effective when the underlying graph has a spectrum comprised of few distinct eigenvalues with high multiplicities. In addition to relaxations based on the Schur-Horn orbitope, we also consider outer-approximations based on other convex graph invariants such as the stability number and the maximum-cut value for the graph edit distance

problem. On the other hand, for the inverse eigenvalue problem, we investigate two relaxations arising from a sum of squares hierarchy. These relaxations have different approximation qualities, and accordingly induce different computational costs. We utilize our framework to generate solutions for, or certify unsolvability of the underlying inverse eigenvalue problem.

We particularly emphasize the computational aspect of our relaxations throughout this thesis. We corroborate the utility of our methods with various numerical experiments.

PUBLISHED CONTENT AND CONTRIBUTIONS

U.O. Candogan and V. Chandrasekaran (2018). "Finding Planted Subgraphs with Few Eigenvalues using the Schur–Horn Relaxation". In: *SIAM Journal on Optimization* 28.1, pp. 735–759. DOI: 10.1137/16M1075144.
U.O.C. participated in the conception of the project, performed the analysis, undertook the numerical experiments, and participated in the writing of the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

Convex optimization is an applied subdiscipline of mathematics with a strong theoretical foundation. Its powerful modeling capabilities and efficient numerical solvers allow for tractably solving various problems arising from numerous areas within the umbrella of computational and mathematical sciences. Diverse applications of convex optimization range from classical fields such as control theory and signal processing to relatively modern areas like machine learning and data analysis. Moreover, convex optimization constitutes an effective tool for solving various real life problems such as scheduling, floor planning, and supply chain management. The list of applications where tools and techniques from convex optimization can provide a useful answer has been steadily expanding. See books [11, 18] and surveys [2, 85, 111, 117] and references therein for numerous examples.

Convex optimization continues to be a useful mathematical framework in application domains where the underlying problem may not readily be expressible as a minimization of a convex function subject to convex constraints. In particular, researchers frequently utilize convex relaxations of non-convex problems by approximating an originally non-convex constraint set from outside with a convex envelope. Oftentimes, such convex relaxations enable solving an underlying difficult and non-convex problem either approximately or exactly. A prominent example from the combinatorial optimization literature is the convex relaxation for the eminent maximum-cut problem, which provably produces an approximate solution that is within a constant multiple of the exact solution [59]. Convex relaxations for graph chromatic number [72] and graph stability number [86] constitute other notable examples that can be utilized for approximately solving related difficult graph problems. Furhermore, solutions for numerous intractable problems arising from operations research (such as the quadratic assignment problem [43]) can be effectively approximated via convex relaxations. Convex relaxations can also be employed for recovering structured solutions in various applications. For instance, the ℓ_1 norm can be utilized as a convex relaxation for the size of the support of a vector with entries in [-1,1] to find sparse solutions to underdetermined linear systems [25].

Similarly, nuclear norm relaxation for the rank of a matrix can be utilized for identifying low rank solutions to linear matrix equations [48], which has important applications in machine learning, such as the PCA.

In this thesis we introduce convex relaxations for exactly or approximately solving three difficult problems that frequently arise in various applications. The first two of these — planted subgraph and graph edit distance problems — are based on graphs, hence they are combinatorial in nature, whereas the third problem we consider is the prominent inverse eigenvalue problem rooted in linear algebra. For the two graph problems, we introduce convex relaxations based on convex graph invariants, which are convex functions of an adjacency matrix of a graph that are invariant to the graph's particular labeling. A prominent convex graph invariant whose utility is investigated in this thesis is the spectrum of an adjacency matrix of the graph. This convex graph invariant gives rise to the Schur-Horn orbitope, which is the convex hull of all matrices with the given spectrum. We later show in this thesis that if adjacency matrices of an underlying graph have few distinct eigenvalues with high multiplicities, then the Schur-Horn orbitope constitutes a suitable outer-approximation to the set of adjacency matrices corresponding to the graph. In addition to the spectrum, in this thesis we also investigate convex relaxations based on other convex graph invariants, such as the inverse stability number of the graph, and the maximum-cut value of the graph. We demonstrate the merit of these convex relaxations in the context of the graph edit distance problem. On the other hand, our convex relaxations for the inverse eigenvalue problem are based on the sum of squares optimization framework. Specifically, we introduce a system of polynomial equations corresponding to affine inverse eigenvalue problem, and investigate two convex relaxations arising from the earlier levels of the corresponding sum of squares hierarchy.

1.1 Main Contributions

We now outline our findings regarding these problems. Details of our work are presented in Chapters 2, 3, and 4 of this thesis. The relevant research correspond to author's papers [26], [27], and [28], respectively.

Finding Planted Subgraphs with Few Eigenvalues using the Schur-Horn Relaxation

In a wide range of application domains, experts wish to extract certain substructures from a larger entity for various purposes. For instance, in biology, researchers frequently attempt to identify repeated substructures in gene and protein networks in order to reveal higher order biological activities. Oftentimes, such problems are cast into a mathematical framework by representing the underlying entities with graphs; the corresponding mathematical problem is the notable planted subgraph problem. This problem is NP-hard in general, and as a result, significant efforts have been directed towards development of tractable procedures that succeed on specific families of problem instances, with a particular focus on planted clique problems.

We propose a new computationally efficient convex relaxation for solving the planted subgraph problem. Our approach is based on the "graph spectrum" convex invariant, as we outer-approximate the (non-convex) set of adjacency matrices representing the hidden graph with the convex hull of all matrices which have the spectrum of an adjacency matrix. Crucially for our purposes, this convex set — referred to as the Schur-Horn orbitope — admits a tractable semidefinite description via majorization inequalities on the spectrum. Our procedure generalizes previous convex relaxation techniques for finding planted cliques. We mathematically demonstrate our procedure's effectiveness in finding planted subgraphs that consist of *few distinct eigenvalues*, and empirically corroborate these findings with numerical experiments. Our analysis is geometric in nature, as we identify conditions of optimality based on the geometry of normal cones at the extreme points of the outer-approximation. The notion of spectrally comonotone matrices, which are pairs of symmetric matrices that can be transformed to diagonal matrices with sorted diagonal entries upon conjugation by the same orthogonal matrix, plays a prominent role in our investigation.

Convex Graph Invariant Relaxations For Graph Edit Distance

In a variety of applications graph similarity frequently arises as a central notion. For instance, in image processing and pattern recognition, graph similarity is frequently used as a primary tool for detecting repetitive substructures. In chemoinformatics and biology, measuring graph similarity allows for deducing unknown functions of entities from similar ones with known functions. The

edit distance between two graphs constitutes a widely used measure of similarity that evaluates the smallest number of vertex and edge deletions/insertions required to transform one graph to another. It is NP-hard to compute in general, and thus exact computation of graph edit distance is only possible for small sized graphs. Consequently, a large number of heuristics have been proposed for approximating this quantity. With few exceptions, these methods generally produce an explicit edit path transforming one graph to the other, and hence they provide upper bounds on the edit distance between two graphs.

We propose a new family of computationally tractable convex relaxations for obtaining *lower bounds* on the graph edit distance between two graphs. As before, we obtain these relaxations by constructing convex outer-envelopes to the set of adjacency matrices representing a graph, based on convex graph invariants. Consequently, these relaxations can be tailored to the structural properties of underlying graphs. The relaxations we investigate include constraints on the graph spectrum (i.e., Schur-Horn orbitope discussed previously) as well as (tractable approximations of) the stability number and the maximum-cut values of graphs. We prove that our relaxation based on the Schur-Horn orbitope succeeds in exactly computing the graph edit distance when one of the graphs consists of few eigenvalues, provided that suitable conditions on graph eigenspaces and true number of edits hold. Furthermore, we identify families of graphs for which our relaxations based on the stability number and the maximum-cut values produce useful lower bounds on the true graph edit distance. We validate the utility of our framework on synthetic problems as well as real applications involving molecular structure comparison problems in chemistry.

Sum of Squares Based Convex Relaxations for Inverse Eigenvalue Problems

The problem of identifying matrices with a desired spectrum which satisfy additional structural constraints — referred to as the inverse eigenvalue problem — is a prominent question in linear algebra. This problem has attracted many researchers in the past several decades due to its inherent difficulty and abundant applications in various fields including pole placement in control theory and frequency identification in vibration theory. Inverse eigenvalue problems are often classified based on the structural constraints they impose. We introduce the *affine inverse eigenvalue problem*, which is the problem of iden-

tifying a symmetric matrix with a desired spectrum in a given affine space. We express this problem in terms of a system of matrix valued polynomial equations. Similar to constraints in combinatorial optimization for enforcing particular scalar variables to be binary, our equations stipulate matrix variables to evaluate to projection matrices. Subsequently, we employ methods from sum of squares optimization and real algebraic geometry literature to obtain two convex semidefinite relaxations to the solution set of affine inverse eigenvalue problem. These convex relaxations incur different computational costs and offer different approximation qualities. In particular, the simpler of these relaxations turn out to be equivalent to the intersection of the constraint subspace with the Schur-Horn orbitope given by the desired vector of eigenvalues. On the other hand, the second convex relaxation imposes additional constraints which essentially correspond to quadratic relations between the decision variables. Our framework is capable of proving unsolvability of a wide range of affine inverse eigenvalue problem instances. Moreover, we demonstrate that if the underlying problem instance is feasible, a simple heuristic of maximizing along random directions across the relaxation sets may be employed for constructing solutions to the affine inverse eigenvalue problem. We corroborate our results with numerous numerical experiments.

Chapter 2

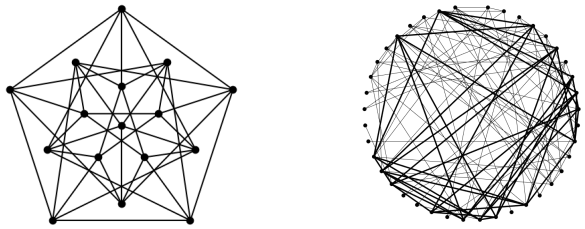
FINDING PLANTED SUBGRAPHS WITH FEW EIGENVALUES USING THE SCHUR-HORN RELAXATION

2.1 Introduction

In application domains ranging from computational biology to social data analysis, graphs are frequently used to model relationships among large numbers of interacting entities. A commonly encountered question across many of these application domains is that of identifying structured subgraphs inside larger graphs. For example, identifying specific motifs or substructures inside gene regulatory networks is useful in revealing higher-order biological function [6, 44, 88]. Similarly, extracting completely connected subgraphs in social networks is useful for determining communities of people that are mutually linked to each other [81, 90, 100]. In this chapter, we propose a new algorithm based on convex optimization for finding structured subgraphs inside large graphs, and we give conditions under which our approach succeeds in performing this task.

Formally, suppose Γ and $\mathfrak{G}$ are graphs[1] on k nodes and n nodes (here $n > k$), respectively, with the following property: there exists a subset of vertices $V \subset \{1, \ldots, n\}$ with $|V| = k$ such that the induced subgraph of $\mathfrak{G}$ corresponding to the vertex set V is isomorphic to Γ. The *planted subgraph* problem is to identify the vertex subset V given the graphs $\mathfrak{G}$ and Γ; see Figure 2.1 for an example. The decision version of the planted subgraph problem is known as the induced subgraph isomorphism problem in the theoretical computer science literature, and it has been shown to be NP-hard [73]. Nevertheless, as this problem arises in a wide range of application domains as described above, significant efforts have been directed towards the development of computationally tractable procedures that succeed on certain families of problem instances. Much of the focus of this attention has been on the special case of the *planted clique* problem in which the subgraph Γ is fully connected. Alon et al. [3] and Feige and Krauthgamer [49] developed a spectral algorithm for the planted clique problem, and subsequently Ames and Vavasis [5] described an approach

[1] Throughout this chapter we consider undirected, unweighted, loopless graphs.



(a) (b)

Figure 2.1: The Clebsch graph (16 nodes) on the left. An example on the right of a 40-node graph containing the Clebsch graph as an induced subgraph; the thick edges correspond to a 16-node induced subgraph that is isomorphic to the Clebsch graph.

based on semidefinite programming with similar performance guarantees to the earlier work based on spectral algorithms. Conceptually, these methods are based on a basic observation about the spectrum of a clique, namely that the adjacency matrix of a clique on k nodes has two distinct eigenvalues, one with multiplicity equal to one and the other with multiplicity equal to $k-1$. We describe a new semidefinite programming technique that generalizes the method of Ames and Vavasis [5] to planted subgraphs Γ that are not fully connected, with the spectral properties of Γ playing a prominent role in our algorithm and our analysis.

Our Contributions

Let $A_\Gamma \in \mathbb{S}^k$ and $A_{\mathfrak{G}} \in \mathbb{S}^n$ represent the adjacency matrices of Γ and of $\mathfrak{G}$, with $\mathbb{S}^q$ denoting the space of $q \times q$ real symmetric matrices. Given any matrix $M \in \mathbb{S}^k$, we let $[M]_{k\to n} \in \mathbb{S}^n$ for $n > k$ denote an $n \times n$ symmetric matrix with the leading principal minor of order k equal to M and all the other entries equal to zero. The following combinatorial optimization problem is a natural first approach to phrase the planted subgraph problem in a variational manner:

$$\begin{aligned} \hat{A}_{co} = \arg\max_{A\in\mathbb{S}^n} \ & \mathrm{Tr}(A \cdot A_{\mathfrak{G}}) \\ & \text{s.t. } A_{i,j} = 0 \text{ if } (A_{\mathfrak{G}})_{i,j} = 0 \text{ and } i \neq j \\ & A \in \{\Pi[A_\Gamma]_{k\to n}\Pi' \mid \Pi \text{ is an } n \times n \text{ permutation matrix}\}. \end{aligned} \tag{2.1}$$

Assuming that there is no other subgraph of $\mathfrak{G}$ that is isomorphic to Γ, one can check that the optimal solution $\hat{A}_{co}$ of this problem identifies the vertices $V \subset \{1,\ldots,n\}$ whose induced subgraph in $\mathfrak{G}$ is isomorphic to Γ, i.e., the unique optimal solution $\hat{A}_{co}$ is equal to zero everywhere except for the principal minor

corresponding to the indices in V and $(\hat{A}_{co})_{V,V} = \tilde{\Pi} A_\Gamma \tilde{\Pi}'$ for some $k \times k$ permutation matrix $\tilde{\Pi}$. However, solving (2.1) is intractable in general. Replacing the combinatorial constraint $A \in \{\Pi[A_\Gamma]_{k\to n}\Pi' \mid \Pi \text{ is an } n\times n \text{ permutation matrix}\}$ with the convex constraint $A \in \text{conv}\{\Pi[A_\Gamma]_{k\to n}\Pi' \mid \Pi \text{ is an } n \times n \text{ permutation matrix}\}$ does not lead to a tractable problem as checking membership in the polytope $\text{conv}\{\Pi[A_\Gamma]_{k\to n}\Pi' \mid \Pi \text{ is an } n \times n \text{ permutation matrix}\}$ is intractable for general planted graphs Γ (unless P = NP).

We describe next a convex outer approximation of the set $\{\Pi[A_\Gamma]_{k\to n}\Pi' \mid \Pi \text{ is an } n \times n \text{ permutation matrix}\}$ that leads to a tractable convex program. For any matrix $M \in \mathbb{S}^n$, the *Schur-Horn orbitope* $\mathcal{SH}(M) \subset \mathbb{S}^n$ is defined as [107]:

$$\mathcal{SH}(M) = \text{conv}\{UMU' \mid U \text{ is an } n \times n \text{ orthogonal matrix}\}. \tag{2.2}$$

The term 'orbitope' was coined by Sanyal, Sottile, and Sturmfels in their work on convex hulls of orbits generated by the actions of groups, and the Schur-Horn orbitope was so named by these authors due to its connection to the Schur-Horn theorem in linear algebra [107]. In combinatorial optimization, approximations based on replacing permutations matrices by orthogonal matrices have also been employed to obtain bounds on the Quadratic Assignment Problem [50]. The set $\mathcal{SH}(M)$ depends only on the eigenvalues of M, and it is clearly an outer approximation of the set $\{\Pi M \Pi' \mid \Pi \text{ is an } n \times n \text{ permutation matrix}\}$. Crucially for our purposes, the Schur-Horn orbitope $\mathcal{SH}(M)$ for any $M \in \mathbb{S}^n$ has a tractable semidefinite description via majorization inequalities on the spectrum of a symmetric matrix [11, 107]; see Section 2.4. Hence, we propose the following tractable semidefinite programming relaxation for the planted subgraph problem:

$$\begin{aligned} \hat{A}_{sh} = \arg\max_{A \in \mathbb{S}^n} \ & \text{Tr}(A \cdot A_{\mathfrak{G}}) \\ \text{s.t. } & A_{i,j} = 0 \text{ if } (A_{\mathfrak{G}})_{i,j} = 0 \text{ and } i \neq j \\ & A \in \mathcal{SH}([A_\Gamma - \gamma I_k]_{k\to n}). \end{aligned} \tag{P}$$

Here $I_k \in \mathbb{S}^k$ is the $k\times k$ identity matrix. We refer to this convex program as the *Schur-Horn relaxation*, and this problem can be solved to a desired precision in polynomial time. This relaxation only requires knowledge of the eigenvalues of the planted graph Γ. The parameter $\gamma \in \mathbb{R}$ is to be specified by the user, and we discuss suitable choices for γ in the sequel. Note that changing A_Γ to

$A_\Gamma - \gamma I_k$ in the constraints of (2.1) essentially leaves that problem unchanged (the nonzero principal minor of the optimal solution simply changes from $\hat{A}_{\text{co}}$ to $\hat{A}_{\text{co}} - \gamma I_k$). However, the additional degree of freedom provided by the parameter γ plays a more significant role in the Schur-Horn relaxation as it allows for shifts of the spectrum of A_Γ to more favorable values, which is essential for the solution of various planted subgraph problems; see Section 2.2 for further details, as well as the experiments in Section 2.4 for numerical illustrations. We say that the Schur-Horn relaxation succeeds in recovering the planted subgraph Γ if the optimal solution $\hat{A}_{sh} \in \mathbb{S}^n$ satisfies the following conditions: the optimal solution $\hat{A}_{sh}$ is unique, the submatrix $(\hat{A}_{sh})_{V,V} = \tilde{\Pi} A_\Gamma \tilde{\Pi}' - \gamma I_k$ for some $k \times k$ permutation matrix $\tilde{\Pi}$, and the remaining entries of $\hat{A}_{sh}$ are equal to zero.

In Section 2.2 we study the geometric properties of the Schur-Horn orbitope as these pertain to the optimality conditions of the Schur-Horn relaxation. Our analysis relies prominently on the notion of *spectrally comonotone* matrices, which refers to a pair of symmetric matrices that can be transformed to diagonal matrices with sorted diagonal entries upon conjugation by the same orthogonal matrix. Spectral comonotonicity is a more restrictive condition than simultaneous diagonalizability, and it enables a precise characterization of the normal cones at extreme points of the Schur-Horn orbitope (Proposition 6). This discussion leads directly to the central observation of this chapter that the Schur-Horn relaxation is useful for finding planted graphs Γ that consist of *few distinct eigenvalues*. Cliques form the simplest examples of such graphs as their spectrum consists of two distinct eigenvalues. There are numerous other graph families whose spectrum consists of few distinct eigenvalues, and the study of such graphs is a significant topic in graph theory [19, 46, 47, 92, 114–116]. For example, strongly regular graphs are (an infinite family of) regular graphs with three distinct eigenvalues; the Clebsch graph of Figure 2.1 is a strongly regular graph on 16 nodes with eigenvalues in the set $\{5, 1, -3\}$ and degree equal to five. For a more extensive list of graphs with few eigenvalues, see Section 2.2.

We state and prove the main theoretical result of this chapter in Section 2.3 – see Theorem 1. If the planted subgraph Γ and its complement are both *symmetric*[2]– Γ and its complement are both vertex- and edge-transitive –

[2]The complement of a symmetric graph is not necessarily symmetric.

and if Γ is connected, then this theorem takes on a simpler form (Corollary 20). Specifically, the success of the Schur-Horn relaxation (P) relies on the existence of a suitable eigenspace $\mathcal{E} \subset \mathbb{R}^k$ of A_Γ. Concretely, let $\mathcal{P}_\mathcal{E} \in \mathbb{S}^k$ denote the projection onto $\mathcal{E}$, and let $\mu(\mathcal{E}) = \max_{i,j,\ i \neq j} \frac{|(\mathcal{P}_\mathcal{E})_{i,j}|}{\sqrt{|(\mathcal{P}_\mathcal{E})_{i,i}||(\mathcal{P}_\mathcal{E})_{j,j}|}}$ denote the coherence of $\mathcal{E}$. Assuming that the edges in $\mathfrak{G}$ outside the induced subgraph Γ are placed independently and uniformly at random with probability $p \in [0, \frac{1}{\mu(\mathcal{E})k})$ (i.e., the Erdős-Rényi random graph model), we show in Corollary 20 that the Schur-Horn relaxation (P) with parameter[3] $\gamma = \lambda_\mathcal{E}$ (the eigenvalue associated to $\mathcal{E}$) succeeds with high probability provided:

$$n \lesssim \min_{\substack{\lambda \text{ eigenvalue} \\ \text{of } A_\Gamma,\, \lambda \neq \lambda_\mathcal{E}}} \min \left\{ |\lambda - \lambda_\mathcal{E}|^2 \, \frac{\dim(\mathcal{E})^2 \big(1 - kp\mu(\mathcal{E})\big)}{k^2\, p}, \big(|\lambda - \lambda_\mathcal{E}| - 2|\lambda_\mathcal{E}|\big)^2 \right\} + k.$$

The coherence parameter $\mu(\mathcal{E})$ lies in $(0, 1]$, and it appears prominently in results on sparse signal recovery via convex optimization [45]. In analogy to that literature, a small value of $\mu(\mathcal{E})$ is useful in our context (informally) to ensure that the planted graph Γ looks sufficiently 'different' from the remainder of $\mathfrak{G}$ (see Section 2.3 for details). Thus, the Schur-Horn relaxation succeeds if the planted graph Γ consists of few distinct eigenvalues that are well-separated, and in which one of the eigenspaces has a small coherence parameter associated to it. For more general non-symmetric graphs, our main result (Theorem 1) is stated in terms of a parameter associated to an eigenspace $\mathcal{E}$ of A_Γ called the *combinatorial width*, which roughly measures the average conditioning over all minors of $\mathcal{P}_\mathcal{E}$ of a certain size.

Specialization to the planted clique problem The sum of the adjacency matrix of a clique and the identity matrix has rank equal to one, and consequently the planted clique problem may be phrased as one of identifying a rank-one submatrix inside a larger matrix (up to shifts of the diagonal by the identity matrix). In her thesis [48], Fazel proposed the nuclear norm as a tractable convex surrogate for identifying low-rank matrices in convex sets, and subsequent efforts provided theoretical support for the effectiveness of this relaxation in a range of rank minimization problems [24, 102]. Building on these ideas, Ames and Vavasis [5] proposed a nuclear norm minimization approach for the planted clique problem. The Schur-Horn relaxation (P) specializes to

[3] In our experiments in Section 2.4, we set γ equal to the eigenvalue of A_Γ with the largest multiplicity. See Section 2.3 for further discussion.

the relaxation in [5] when Γ is the clique. Specifically, letting $A_{\text{clique}} \in \mathbb{S}^k$ denote the adjacency matrix of a k-clique, one can check that:

$$\mathcal{SH}([A_{\text{clique}} + I_k]_{k\to n}) = \{P \in \mathbb{S}^n \mid \text{trace}(P) = k,\ P \succeq 0\}. \tag{2.3}$$

As the nuclear norm of a positive semidefinite matrix is equal to its trace, the Schur-Horn orbitope $\mathcal{SH}([A_\Gamma + I_k]_{k\to n})$ is simply a face of the nuclear norm ball in $\mathbb{S}^n$ scaled by a factor k. Thus, the Schur-Horn relaxation (P) with $\gamma = -1$ is effectively a nuclear norm relaxation when the planted subgraph of interest is the clique.[4] Further, our main result (Theorem 1) can be specialized to the case of a planted clique to obtain the main result in [5]; see Corollary 21.

Chapter Outline

In Section 2.2 we discuss the geometric properties of the Schur-Horn orbitope and their connection to the optimality conditions of the Schur-Horn relaxation, along with an extensive list of families of graphs with few eigenvalues. Section 2.3 contains our main theoretical results, while in Section 2.4 we demonstrate the utility of the Schur-Horn relaxation in practice via numerical experiments. We conclude in Section 2.5 with a discussion of further research directions.

Notation The *normal cone* at a point $x \in \mathcal{C}$ for a closed, convex set $\mathcal{C} \subset \mathbb{R}^n$ is denoted by $\mathcal{N}_{\mathcal{C}}(x)$ and it is the collection of linear functionals that attain their maximal value over $\mathcal{C}$ at x [105]. The projection operator onto a subspace $\mathcal{E} \subset \mathbb{R}^n$ is denoted by $\mathcal{P}_{\mathcal{E}}$. The restriction of a linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ to an invariant subspace $\mathcal{E}$ of A is denoted by $A|_{\mathcal{E}} : \mathcal{E} \to \mathcal{E}$. The orthogonal complement of a subspace $\mathcal{E}$ is denoted by $\mathcal{E}^\perp$. The notation $\dim(\mathcal{E})$ denotes the dimension of a subspace $\mathcal{E}$. The *eigengap* of a symmetric matrix $M \in \mathbb{S}^n$ associated to an invariant subspace $\mathcal{E} \subset \mathbb{R}^n$ of M is defined as:

$$\text{eigengap}(M, \mathcal{E}) = \min\big\{|\lambda_{\mathcal{E}} - \lambda_{\mathcal{E}^\perp}| \;\big|\; \lambda_{\mathcal{E}} \text{ an eigenvalue of } M|_{\mathcal{E}},\ \lambda_{\mathcal{E}^\perp} \text{ an eigenvalue of } M|_{\mathcal{E}^\perp}\big\}.$$

The smallest and largest eigenvalues of a symmetric matrix A are represented by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$, respectively. The norms $\|\cdot\|, \|\cdot\|_2$, and $\|\cdot\|_F$ denote the vector ℓ_2 norm, the matrix operator/spectral norm, and the matrix Frobenius norm, respectively. The vector $1_\ell \in \mathbb{R}^\ell$ denotes the all-ones vector of

[4] The nuclear norm relaxation in [5] is formulated in a slightly different fashion compared to the Schur-Horn relaxation (P) for the case of the planted clique; specifically, we show in Appendix A that our relaxation succeeds whenever the nuclear norm relaxation in [5] succeeds.

length ℓ. We denote the identity matrix of size k by I_k. The matrix $I_\Omega \in \mathbb{R}^{|\Omega| \times k}$ denotes the matrix whose rows are the rows of I_k indexed by $\Omega \subset \{1, \ldots, k\}$, so that the rows of $I_\Omega A$ are the rows of A indexed by Ω for any $A \in \mathbb{R}^{k \times q}$. The matrix $A_{\Omega,\Omega} \in \mathbb{R}^{|\Omega| \times |\Omega|}$ denotes the principal minor of A indexed by the set Ω. The group of $n \times n$ orthogonal matrices is denoted by $\mathcal{O}_n \subset \mathbb{R}^{n \times n}$. The set $\text{relint}(\mathcal{C})$ specifies the relative interior of any convex set $\mathcal{C}$. The column space of a matrix A is denoted by $\text{col}(A)$. The quantity $\mathbb{E}[\cdot]$ denotes the usual expected value, where the distribution is clear from context.

2.2 Geometric Properties of the Schur-Horn Orbitope

In this section, we analyze the optimality conditions of the Schur-Horn relaxation from a geometric perspective. In particular, the notion of a pair of *spectrally comonotone matrices* plays a central role in our development, and we elaborate on this point in the next subsection. Based on this discussion, we observe that the Schur-Horn relaxation is especially useful for finding planted graphs consisting of few distinct eigenvalues, and we give examples of graphs with this property in Section 2.2. The main theoretical results formalizing the utility of the Schur-Horn relaxation are presented in Section 2.3.

Optimality Conditions of the Schur-Horn Relaxation

We state the optimality conditions of the Schur-Horn relaxation in terms of the normal cones at extreme points of the Schur-Horn orbitope:

Lemma 1. *Consider a planted subgraph problem instance in which the nodes of $\mathfrak{G}$ and Γ are labeled so that the leading principal minor of $A_{\mathfrak{G}}$ of order k is equal to A_Γ. Suppose there exists a matrix $M \in \mathbb{S}^n$ with the following properties:*

1. $M_{i,j} = (A_{\mathfrak{G}})_{i,j}$ if $(A_{\mathfrak{G}})_{i,j} = 1$ or if $i = j$,
2. $M \in \text{relint}\left(\mathcal{N}_{\mathcal{SH}([A_\Gamma - \gamma I_k]_{k \to n})}([A_\Gamma - \gamma I_k]_{k \to n})\right)$.

Then the Schur-Horn relaxation succeeds at identifying the planted subgraph Γ inside the larger graph $\mathfrak{G}$, i.e., the unique optimal solution of the convex program (P) is $\hat{A}_{sh} = \begin{pmatrix} A_\Gamma - \gamma I_k & 0 \\ 0 & 0 \end{pmatrix}$.

Proof. From standard results in convex analysis [105], we have that $\begin{pmatrix} A_\Gamma - \gamma I_k & 0 \\ 0 & 0 \end{pmatrix}$ is the unique optimal solution of (P) if $A_{\mathfrak{G}}$ can be decomposed as $A_{\mathfrak{G}} \in K + \operatorname{relint}\left(\mathcal{N}_{\mathcal{SH}([A_\Gamma - \gamma I_k]_{k\to n})}([A_\Gamma - \gamma I_k]_{k\to n})\right)$ for some matrix $K \in \mathbb{S}^n$ that satisfies:

$$K_{i,j} = 0 \text{ if either } (A_{\mathfrak{G}})_{i,j} = 1 \text{ or } i = j.$$

Letting $K = A_{\mathfrak{G}} - M$ we have the desired result. □

The assumption on the node labeling is made purely for the sake of notational convenience in our analysis (to avoid clutter in having to keep track of additional permutations), and our algorithmic methodology does not rely on such a labeling. Based on this characterization of the optimality conditions, the success of the Schur-Horn relaxation relies on the existence of a suitable dual variable $M \in \mathbb{S}^n$ that satisfies two conditions. The first of these conditions relates to the structure of the noise edges in $\mathfrak{G}$, while the second condition relates to the structure of the planted graph Γ via the normal cone $\mathcal{N}_{\mathcal{SH}([A_\Gamma - \gamma I_k]_{k\to n})}([A_\Gamma - \gamma I_k]_{k\to n})$. From the viewpoint of Lemma 1, favorable problem instances for the Schur-Horn relaxation are, informally speaking, those in which there are not too many noise edges in $\mathfrak{G}$ (implying a less restrictive first requirement on M) and in which the normal cone $\mathcal{N}_{\mathcal{SH}([A_\Gamma - \gamma I_k]_{k\to n})}([A_\Gamma - \gamma I_k]_{k\to n})$ is large (entailing a more flexible second condition for M). The interplay between these two conditions forms the basis of our analysis and results presented in Section 2.3. In the remainder of the present section, we investigate spectral properties of planted graphs Γ that result in a large normal cone $\mathcal{N}_{\mathcal{SH}([A_\Gamma - \gamma I_k]_{k\to n})}([A_\Gamma - \gamma I_k]_{k\to n})$.

The normal cones at the extreme points of the Schur-Horn orbitope are conveniently described based on the following notion (see Proposition 6 in the sequel):

Definition 2. *A pair of symmetric matrices $A, B \in \mathbb{S}^n$ is* spectrally comonotone *if there exists an orthogonal matrix $U \in \mathbb{R}^{n\times n}$ such that $U'AU$ and $U'BU$ are both diagonal matrices with the diagonal entries sorted in nonincreasing order.*

The stipulation that two matrices be spectrally comonotone is a stronger condition than the requirement that the matrices be simultaneously diagonalizable,

due to the additional restriction on the ordering of the diagonal entries upon conjugation by an orthogonal matrix.

Example 3. *Consider the matrices* $A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$. *The matrices* A *and* B *are spectrally comonotone, while* A *and* C *are only simultaneously diagonalizable and are not spectrally comonotone.*

As Proposition 1 states the optimality conditions of the Schur-Horn relaxation in terms of the *relative interiors* of normal cones at extreme points of the Schur-Horn orbitope, we need the following "strict" analog of spectral comonotonicity:

Definition 4. *A matrix* $A \in \mathbb{S}^n$ *is* strictly spectrally comonotone *with a matrix* $B \in \mathbb{S}^n$*, if for every* $P \in \mathbb{S}^n$ *that is simultaneously diagonalizable with* B*, there exists* $\epsilon > 0$ *such that* $A + \epsilon P$ *and* B *are spectrally comonotone.*

Strict spectral comonotonicity is more restrictive than spectral comonotonicity. Further, the definition of strict spectral comonotonocity is not a symmetric one, unlike that of spectral comonotonicity, i.e., even if $A \in \mathbb{S}^n$ is strictly spectrally comonotone with $B \in \mathbb{S}^n$, it may be that B is not strictly spectrally comonotone with A.

Example 5. *Consider the matrices* $A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, B = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. *The matrix* A *is strictly spectrally comonotone with the matrix* B*, but* B *is not strictly spectrally comonotone with* A*.*

The following result provides a characterization of normal cones at extreme points of the Schur-Horn orbitope in terms of spectrally comonotone matrices:

Proposition 6. *For any matrix* $M \in \mathbb{S}^n$ *and the associated Schur-Horn orbitope* $\mathcal{SH}(M)$*, the normal cone* $\mathcal{N}_{\mathcal{SH}(M)}(W)$ *and its relative interior at an extreme point* W *of* $\mathcal{SH}(M)$ *are given by:*

$$\mathcal{N}_{\mathcal{SH}(M)}(W) = \{Q \in \mathbb{S}^n \mid Q \text{ and } W \text{ are spectrally comonotone}\}.$$

$$\operatorname{relint}\left(\mathcal{N}_{\mathcal{SH}(M)}(W)\right) = \{Q \in \mathbb{S}^n \mid Q \text{ strictly spectrally comonotone with } W\}.$$

Note For any matrix $M \in \mathbb{S}^n$, the extreme points of $\mathcal{SH}(M)$ are the elements of the set $\{UMU' \mid U \in \mathcal{O}_n\}$, as each of the matrices UMU' for $U \in \mathcal{O}_n$ has the same Frobenius norm.

Proof. Let $W = M$ without loss of generality. We have that:

$$\begin{aligned}\mathcal{N}_{\mathcal{SH}(M)}(M) &= \{Y \in \mathbb{S}^n \mid \sup_{Z \in \mathcal{SH}(M)} \operatorname{Tr}(YZ) \leq \operatorname{Tr}(YM)\} \\ &= \{Y \in \mathbb{S}^n \mid \sup_{Z = UMU' \text{ for } U \in \mathcal{O}_n} \operatorname{Tr}(YZ) \leq \operatorname{Tr}(YM)\} \\ &= \{Y \in \mathbb{S}^n \mid \sup_{U \in \mathcal{O}_n} \operatorname{Tr}(U'YUM) = \operatorname{Tr}(YM)\}.\end{aligned}$$

The last line follows from the inequality $\operatorname{Tr}(YM) \leq \sup_{U \in \mathcal{O}_n} \operatorname{Tr}(U'YUM)$. Considering the case of equality in the Von Neumann trace inequality [119], we have that $\sup_{U \in \mathcal{O}_n} \operatorname{Tr}(U'YUM) = \operatorname{Tr}(YM)$ if and only if Y and M are spectrally comonotone. The claim about the relative interior of the normal cone follows immediately from the definition of strict spectral comonotonicity. □

If a matrix $M \in \mathbb{S}^n$ has few distinct eigenvalues, the normal cone at an extreme point UMU' (for U orthogonal) of $\mathcal{SH}(M)$ is larger as there are many more matrices that are spectrally comonotone with UMU'. Based on Proposition 6, this observation suggests that planted graphs Γ with *few distinct eigenvalues* have large normal cones $\mathcal{N}_{\mathcal{SH}([A_\Gamma - \gamma I_k]_{k \to n})}([A_\Gamma - \gamma I_k]_{k \to n})$, and such graphs are especially amenable to recovery in planted subgraph problems via the Schur-Horn relaxation. We make this insight more precise with our analysis in Section 2.3. Proposition 6 also points to the utility of employing the parameter γ in the Schur-Horn relaxation (P). Specifically, multiplicities in the spectrum of the matrix $[A_\Gamma - \gamma I_k]_{k \to n} \in \mathbb{S}^n$ may be increased via suitable choices of γ, which in turn makes the normal cone $\mathcal{N}_{\mathcal{SH}([A_\Gamma - \gamma I_k]_{k \to n})}([A_\Gamma - \gamma I_k]_{k \to n})$ larger. In particular, setting γ equal to an eigenvalue of A_Γ increases the multiplicity of zero as an eigenvalue of $[A_\Gamma - \gamma I_k]_{k \to n}$. As detailed in Section 2.3, the success of the Schur-Horn relaxation relies on the existence of an eigenspace $\mathcal{E} \subset \mathbb{R}^k$ of A_Γ with small coherence parameter, and the appropriate choice of γ is the eigenvalue $\lambda_\mathcal{E}$ associated to $\mathcal{E}$. In our experiments in Section 2.4, we set γ equal to the eigenvalue of A_Γ with largest multiplicity, so that the multiplicity of zero as an eigenvalue of $[A_\Gamma - \gamma I_k]_{k \to n}$ is as large as possible.

To conclude, we record an observation on spectral comonotonocity that is useful in Section 2.3. The claim is straightforward and therefore we omit the proof.

Lemma 7. *A pair of symmetric matrices $A, B \in \mathbb{S}^n$ is spectrally comonotone if and only if A and B are simultaneously diagonalizable and*

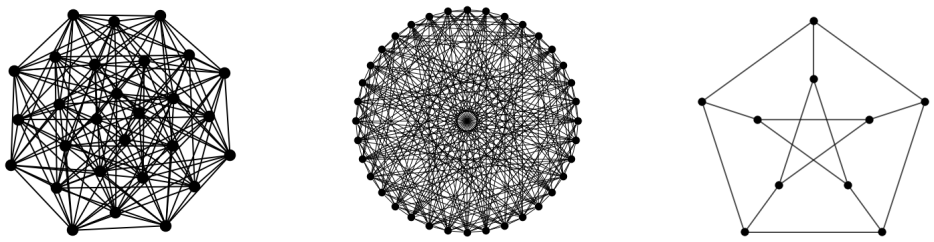
$$\lambda_{\min}(A|_{\mathcal{E}_i}) \geq \lambda_{\max}(A|_{\mathcal{E}_{i+1}}) \quad \forall i \in \{1, \ldots, t-1\}, \tag{2.4}$$

where $\mathcal{E}_i$ for $i \in \{1, \ldots, t\}$ are eigenspaces of B ordered such that the corresponding eigenvalues of B are decreasing. Further, A is strictly spectrally comonotone with B if and only if A and B are simultaneously diagonalizable and each of the inequalities (2.4) *holds strictly.*

Note that if A and B simultaneously diagonalizable, then any eigenspace $\mathcal{E}$ of B is an invariant subspace of A. As a result, the restriction of A to the eigenspaces of B in (2.4) is consistent with the notation described in Section 2.1.

Graphs with Few Eigenvalues

Building on the preceding section, we give examples of families of graphs consisting of few distinct eigenvalues. Such graphs have received much attention due to their connections to topics in combinatorics and design theory such as pseudorandomness [74] and association schemes [9, 58].

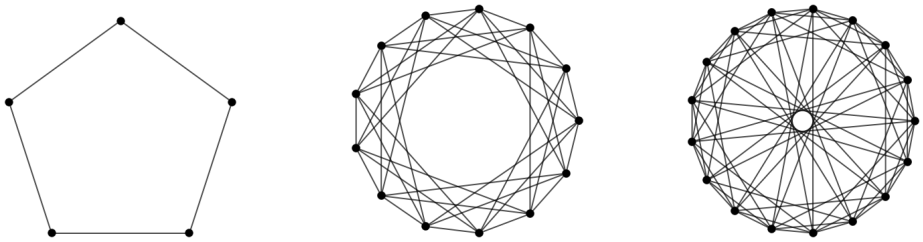


(a) (b) (c)

Figure 2.2: From left to right: 8-triangular graph, 9-triangular graph, and Petersen graph.

Triangular graphs The triangular graph $\mathcal{T}_m$ of order m is the line graph of the complete graph on m nodes. The graph $\mathcal{T}_m$ has $\binom{m}{2}$ nodes and it has the three distinct eigenvalues $2(m-2)$ (with multiplicity 1), $m-4$ (with multiplicity $m-1$), and -2 (with multiplicity $\frac{m(m-3)}{2}$). Figure 2.2 gives two examples.

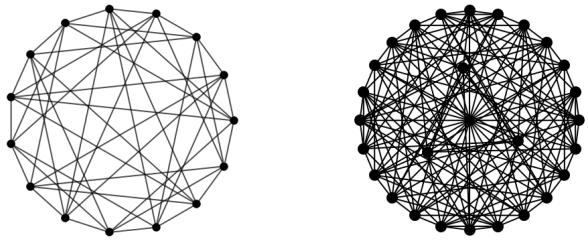
Kneser graphs A Kneser graph $\mathcal{K}(m, \ell)$ is a graph on $\binom{m}{\ell}$ nodes, each corresponding to an ℓ-element subset of m elements, and it consists of edges between those pairs of vertices for which the corresponding subsets are disjoint. The graph $\mathcal{K}(m, 1)$ is the complete graph on m nodes and the graph $\mathcal{K}(5, 2)$ is the Petersen graph (Figure 2.2). The Kneser graph $\mathcal{K}(m, \ell)$ has $\ell + 1$ distinct eigenvalues in general.



(a) (b) (c)

Figure 2.3: From left to right: 5-Paley graph, 13-Paley graph, and 17-Paley graph.

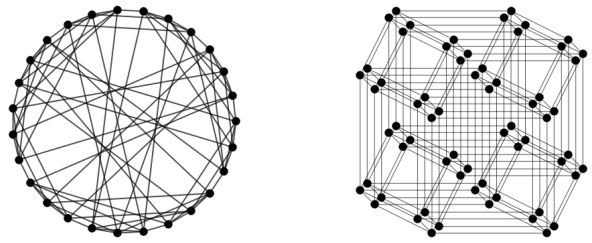
Paley graphs Let q be a prime power such that $q \equiv 1 (\text{mod } 4)$. The Paley graph on q nodes is an undirected graph formed by connecting pairs of nodes $i, j \subset \{0, \ldots, q-1\}$ if the difference $i - j$ is a square in the finite field GF(q). Note that $i - j$ is a square if and only if $j - i$ is a square as -1 is a square in GF(q). Paley graphs have eigenvalues $\frac{1}{2}(q-1)$ (with multiplicity 1), $\frac{1}{2}(-1+\sqrt{q})$ (with multiplicity $\frac{1}{2}(q-1)$), and $\frac{1}{2}(-1-\sqrt{q})$ (with multiplicity $\frac{1}{2}(q-1)$). Paley graphs are also examples of pseudorandom graphs as they exhibit properties similar to random graphs (in the limit of large q) [74]. Figure 2.3 shows the three smallest Paley graphs.



(a) (b)

Figure 2.4: Generalized quadrangle-$(2, 2)$ graph (left) and generalized quadrangle-$(2, 4)$ graph (right).

Strongly regular graphs These are regular graphs with the property that every pair of adjacent vertices has the same number d_a of common neighbors and every pair of non-adjacent vertices has the same number d_{na} of common neighbors, for some integers d_a, d_{na} [16]. Strongly regular graphs that are connected have three distinct eigenvalues; conversely, connected and regular graphs with three distinct eigenvalues are necessarily strongly regular. The triangular graphs, Kneser graphs with parameter $\ell = 2$, and the Paley graphs mentioned above are examples of strongly regular graphs. The Clebsch graph shown in Figure 2.1a in the introduction is also a strongly regular graph with degree 5 and eigenvalues 5 (with multiplicity 1), -3 (with multiplicity 5), and 1 (with multiplicity 10). The generalized quadrangle graphs shown in Figure 2.4 are additional examples of strongly regular graphs. Strongly regular graphs form a significant topic in graph theory due to their many regularity properties [21, 23, 108].



(a) (b)

Figure 2.5: Hamming-$(3, 3)$ graph with 4 distinct eigenvalues (left) and 6-hypercube graph with 7 distinct eigenvalues (right).

Other examples Unlike regular graphs with three distinct eigenvalues, graphs with four (or more) eigenvalues do not appear to have a simple combinatorial characterization [115]. Nonetheless, there are many constructions of such graphs in the literature [62, 114, 115], most notably those derived from distance-regular graphs [20] and from association schemes. Indeed, there are infinitely many graphs with exactly $d+1$ distinct eigenvalues for $d = 3, 4$ which arise from association schemes with d classes. The two graphs shown in Figure 2.5 are obtained from the Hamming scheme.

2.3 Recovering Subgraphs Planted in Erdős-Rényi Random Graphs

In this section we discuss our theoretical results on the performance of the Schur-Horn relaxation in recovering subgraphs planted inside Erdős-Rényi ran-

dom graphs. Formally, suppose without loss of generality as in the previous section that the nodes of $\mathfrak{G}$ and of Γ are labeled so that the leading principal minor of $A_{\mathfrak{G}}$ of order k is equal to A_Γ. The Erdős-Rényi model for the planted subgraph problem specifies a distribution on the edges in the remainder of the graph $\mathfrak{G}$ via a probability parameter $p \in [0,1]$; for each $i, j \in \{1, \ldots, n\}$ with $i < j$ and $k < j$, the graph $\mathfrak{G}$ contains an edge between nodes i and j with probability p (independent of the other edges):

$$(A_{\mathfrak{G}})_{i,j} = (A_{\mathfrak{G}})_{j,i} = \begin{cases} 1, \text{ with probability } p, \\ 0, \text{ with probability } 1-p. \end{cases}$$

We begin with a sufficient condition for the optimality condition described in Lemma 1, which suggests a natural approach for constructing suitable dual variables for certifying optimality. These sufficient conditions point to the importance of the existence of an eigenspace of A_Γ with certain properties to the success of the Schur-Horn relaxation; these properties are discussed in Section 2.3. In Section 2.3 we state and prove the main theorem (Theorem 1) of this chapter, with Section 2.3 giving specializations of this result (e.g., to the planted clique problem).

A Simpler Sufficient Condition for Optimality

The following proposition provides a simpler set of conditions than those in Lemma 1 on dual variables that certify the success of the Schur-Horn relaxation. This result continues to be deterministic in nature, and the probabilistic aspects of our analysis – due to the Erdős-Rényi model – appear in the sequel.

Proposition 8. *Consider a planted subgraph problem instance in which the nodes of $\mathfrak{G}$ and Γ are labeled so that the leading principal minor of $A_{\mathfrak{G}}$ of order k is equal to A_Γ. Suppose there exists an eigenspace $\mathcal{E} \subset \mathbb{R}^k$ of A_Γ with eigenvalue $\lambda_{\mathcal{E}}$, and suppose there exists a matrix $M = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}' & M_{22} \end{pmatrix} \in \mathbb{S}^n$ with submatrices $M_{11} \in \mathbb{S}^k, M_{12} \in \mathbb{R}^{k \times (n-k)}, M_{22} \in \mathbb{S}^{n-k}$ such that the the following conditions are satisfied:*

(i) $M_{i,j} = (A_{\mathfrak{G}})_{i,j}$, *if* $(A_{\mathfrak{G}})_{i,j} = 1$ *or if* $i = j$,

(ii) The submatrix $M_{11} \in \mathbb{S}^k$ is strictly spectrally comonotone with A_Γ,

(iii) $\lambda_{\max}(M_{11}|_{\mathcal{E}}) \geq \lambda_{\mathcal{E}}$ *and* $\lambda_{\min}(M_{11}|_{\mathcal{E}}) \leq \lambda_{\mathcal{E}}$,

(iv) Each column of the submatrix $M_{12} \in \mathbb{R}^{k\times(n-k)}$ *lies in the subspace* $\mathcal{E}$*,*

(v) $\text{eigengap}(M_{11}, \mathcal{E}) > \|M_{12}\|_2 + \|M_{22}\|_2 + |\lambda_{\mathcal{E}}|$.

Then the Schur-Horn relaxation (P) with parameter $\gamma = \lambda_{\mathcal{E}}$ *succeeds at identifying the planted subgraph* Γ *inside the larger graph* $\mathfrak{G}$*.*

Proof. We establish this result by showing that the given matrix $M \in \mathbb{S}^n$ satisfies the requirements of Lemma 1. The first condition of Lemma 1 is identical to that of this proposition, and therefore it is satisfied. We prove next that the remaining conditions of this proposition ensure that the second requirement of Lemma 1 is also satisfied, i.e., $M \in \text{relint}\left(\mathcal{N}_{\mathcal{SH}([A_\Gamma - \lambda_{\mathcal{E}} I_k]_{k\to n})}([A_\Gamma - \lambda_{\mathcal{E}} I_k]_{k\to n})\right)$. Based on Proposition 6, this entails showing that M is strictly spectrally comonotone with $[A_\Gamma - \lambda_{\mathcal{E}} I_k]_{k\to n}$. Our strategy is to employ Lemma 7.

Let $\mathcal{E}_i \subset \mathbb{R}^k,\ i = 1, \ldots, t$ be the eigenspaces of A_Γ ordered such that the corresponding eigenvalues $\lambda_{\mathcal{E}_i}$ are strictly decreasing, and suppose $\mathcal{E}_j = \mathcal{E},\ \lambda_{\mathcal{E}_j} = \lambda_{\mathcal{E}}$ for some $j \in \{1, \ldots, t\}$. As 0 is an eigenvalue of $A_\Gamma - \lambda_{\mathcal{E}} I_k$, one can check that the eigenspaces of $[A_\Gamma - \lambda_{\mathcal{E}} I_k]_{k\to n}$ are $\tilde{\mathcal{E}}_i = \mathcal{E}_i \times \{0\} \subset \mathbb{R}^k \times \mathbb{R}^{n-k},\ i = 1, \ldots, t,\ i \neq j$ (with corresponding eigenvalues $\lambda_{\mathcal{E}_i} - \lambda_{\mathcal{E}}$) and $\tilde{\mathcal{E}}_j = \mathcal{E} \times \mathbb{R}^{n-k} \subset \mathbb{R}^k \times \mathbb{R}^{n-k}$ (with eigenvalue 0). We now need to show that M and $[A_\Gamma - \lambda_{\mathcal{E}} I_k]_{k\to n}$ are simultaneously diagonalizable, and that $\lambda_{\min}(M|_{\tilde{\mathcal{E}}_i}) > \lambda_{\max}(M|_{\tilde{\mathcal{E}}_{i+1}})$ for $i \in \{1, \ldots, t-1\}$.

First, as $\mathcal{E}$ is an eigenspace of $A_\Gamma - \lambda_{\mathcal{E}} I_k$ with eigenvalue 0 and as every column of M_{12} belongs to $\mathcal{E}$, one can check that $(A_\Gamma - \lambda_{\mathcal{E}} I_k) \cdot M_{12} = 0 \in \mathbb{R}^{k\times(n-k)}$. Further, from Lemma 7 we note that M_{11} and $A_\Gamma - \lambda_{\mathcal{E}} I_k$ are simultaneously diagonalizable because M_{11} is strictly spectrally comonotone with A_Γ (and hence with $A_\Gamma - \lambda_{\mathcal{E}} I_k$). From these two observations one can check that M and $[A_\Gamma - \lambda_{\mathcal{E}} I_k]_{k\to n}$ commute with each other, and therefore are simultaneously diagonalizable.

As M and $[A_\Gamma - \lambda_{\mathcal{E}} I_k]_{k\to n}$ are simultaneously diagonalizable, we have that the eigenspaces $\tilde{\mathcal{E}}_i,\ i = 1, \ldots, t$ of $[A_\Gamma - \lambda_{\mathcal{E}} I_k]_{k\to n}$ are invariant subspaces of M. Similarly, as M_{11} is strictly spectrally comonotone with A_Γ, the eigenspaces $\mathcal{E}_i$ are invariant subspaces of M_{11}. Based on the structure of these eigenspaces as described above, one can check that the eigenvalues of $M|_{\tilde{\mathcal{E}}_i}$ are equal to those of $M_{11}|_{\mathcal{E}_i}$ for each $i = 1, \ldots, t,\ i \neq j$. Hence, $\lambda_{\min}(M|_{\tilde{\mathcal{E}}_i}) > \lambda_{\max}(M|_{\tilde{\mathcal{E}}_{i+1}})$ for $i > j$ and for $i < j - 1$.

All that remains to be verified is that $\lambda_{\min}(M|_{\tilde{\mathcal{E}}_j}) > \lambda_{\max}(M|_{\tilde{\mathcal{E}}_{j+1}})$ and that $\lambda_{\min}(M|_{\tilde{\mathcal{E}}_{j-1}}) > \lambda_{\max}(M|_{\tilde{\mathcal{E}}_j})$. As each column of M_{12} belongs to $\mathcal{E}$ and as $\tilde{\mathcal{E}}_j = \mathcal{E} \times \mathbb{R}^{n-k} \subset \mathbb{R}^k \times \mathbb{R}^{n-k}$, we have for $x \in \mathcal{E}, y \in \mathbb{R}^{n-k}$ that:

$$M|_{\tilde{\mathcal{E}}_j} \begin{pmatrix} x \\ y \end{pmatrix} = \left[\begin{pmatrix} M_{11}|_{\mathcal{E}} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & M_{12} \\ M_{12}' & M_{22} \end{pmatrix} \right] \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} M_{11}|_{\mathcal{E}} x + M_{12} y \\ M_{12}' x + M_{22} y \end{pmatrix} \in \tilde{\mathcal{E}}_j. \tag{2.5}$$

Consequently, recalling that $\mathcal{E}_j = \mathcal{E}$ we have:

$$\begin{aligned}
\lambda_{\max}(M|_{\tilde{\mathcal{E}}_j}) &\leq \max\{\lambda_{\max}(M_{11}|_{\mathcal{E}}), 0\} + \|M_{12}\|_2 + \|M_{22}\|_2 \\
&< \max\{\lambda_{\max}(M_{11}|_{\mathcal{E}}), 0\} - |\lambda_{\mathcal{E}}| + \text{eigengap}(M_{11}, \mathcal{E}) \\
&\leq \max\{\lambda_{\max}(M_{11}|_{\mathcal{E}}), 0\} - |\lambda_{\mathcal{E}}| + \lambda_{\min}(M_{11}|_{\mathcal{E}_{j-1}}) - \lambda_{\max}(M_{11}|_{\mathcal{E}}) \\
&= \max\{0, -\lambda_{\max}(M_{11}|_{\mathcal{E}})\} - |\lambda_{\mathcal{E}}| + \lambda_{\min}(M_{11}|_{\mathcal{E}_{j-1}}) \\
&\leq \max\{0, -\lambda_{\mathcal{E}}\} - |\lambda_{\mathcal{E}}| + \lambda_{\min}(M_{11}|_{\mathcal{E}_{j-1}}) \\
&\leq \lambda_{\min}(M_{11}|_{\mathcal{E}_{j-1}}) \\
&= \lambda_{\min}(M|_{\tilde{\mathcal{E}}_{j-1}}).
\end{aligned}$$

The first inequality follows from (2.5), the second inequality from condition (v), the third inequality from the definition of eigengap (see Section 2.1) as $\mathcal{E}_j = \mathcal{E}$, the fourth inequality from condition (iii), and the second equality from the fact that the eigenvalues of $M|_{\tilde{\mathcal{E}}_i}$ are equal to those of $M_{11}|_{\mathcal{E}_i}$ for each $i = 1, \ldots, t,\ i \neq j$. Similarly, one can check that $\lambda_{\min}(M|_{\tilde{\mathcal{E}}_j}) > \lambda_{\max}(M|_{\tilde{\mathcal{E}}_{j+1}})$. This concludes the proof. □

This result provides a concrete approach for constructing dual variables to certify the optimality of the Schur-Horn relaxation (P) at the desired solution. In the remainder of this section, we give conditions on the eigenstructure of the planted graph Γ, the probability p of the Erdős-Rényi model, and the size n of the larger graph $\mathfrak{G}$ under which the Schur-Horn relaxation (P) succeeds with high probability.

Invariants of Graph Eigenspaces

In this section, we investigate properties of eigenspaces of graphs which ensure that the conditions of Proposition 8 can be satisfied. For notational clarity in the discussion in this section, we let $\Omega_j \subset \{1, \ldots, k\}$ for $j = 1, \ldots, n-k$ denote the locations of the entries equal to one in the submatrix $(A_{\mathfrak{G}})_{i,j+k},\ i = 1, \ldots, k; j = 1, \ldots, n-k$, i.e., $(A_{\mathfrak{G}})_{i,j+k} = 1 \Leftrightarrow i \in \Omega_j$.

A requirement of Proposition 8 is the existence of a suitable eigenspace $\mathcal{E} \subset \mathbb{R}^k$ of A_Γ such that one can obtain a matrix $M_{12} \in \mathbb{R}^{k\times(n-k)}$ (a submatrix of a larger dual certificate) that satisfies three conditions: (i) Every column of M_{12} lies in $\mathcal{E}$, (ii) For each $i = 1, \ldots, k$ and $j = 1, \ldots, n-k$ we have that $(M_{12})_{i,j} = 1$ if $(A_{\mathfrak{G}})_{i,j+k} = 1$, and (iii) The operator norm $\|M_{12}\|_2$ is as small as possible.

We begin by analyzing the first two conditions and the restrictions they impose on $\mathcal{E}$. Consider the j'th column of M_{12} for a fixed $j \in \{1, \ldots, n-k\}$ as an illustration. Then conditions (i) and (ii) are simultaneously satisfied if the coordinate subspace of vectors in $\mathbb{R}^k$ with support on the indices in Ω_j has a *transverse intersection* with $\mathcal{E}^\perp$. More generally, a natural sufficient condition for the first two requirements on M_{12} to be satisfied (for every column) is for $\mathcal{E}^\perp$ to have a transverse intersection with the coordinate subspaces specified by each of the subsets Ω_j for $j = 1, \ldots, k$. This observation leads to the following invariant that characterizes the transversality of a subspace with all coordinate subspaces of a certain dimension:

Definition 9. *[76] The* Kruskal rank *of a subspace $\mathcal{S} \subseteq \mathbb{R}^k$, denoted* kruskal($\mathcal{S}$)*, is the largest $m \in \mathbb{Z}$ such that for any $\Omega \subseteq \{1, \ldots, k\}$ with $|\Omega| = m$ we have:*

$$\mathcal{S}^\perp \cap \{v \in \mathbb{R}^k \mid v_i = 0 \text{ if } i \notin \Omega\} = \{0\}.$$

In other words, the Kruskal rank of a subspace $\mathcal{S} \subset \mathbb{R}^k$ is one less than the size of the support of the sparsest nonzero vector in $\mathbb{R}^k$ that is orthogonal to $\mathcal{S}$. The Kruskal rank of a matrix – the largest m such that all subsets of m columns of the matrix are linearly independent – was first introduced in [76] in the context of tensor decompositions. This version in terms of matrices is equivalent to our definition in terms of subspaces. One can check that all principal minors of $\mathcal{P}_\mathcal{S}$ of size up to kruskal($\mathcal{S}$) are non-singular.

Recall that the entries $(A_{\mathfrak{G}})_{i,j+k}$ for $i = 1, \ldots, k$ and $j = 1, \ldots, n-k$ correspond to edges (or lack thereof) between nodes in $\mathfrak{G}$ outside the induced subgraph corresponding to Γ and those of Γ. Therefore, if we employ the Schur-Horn relaxation with parameter $\gamma = \lambda_\mathcal{E}$ (the eigenvalue associated to $\mathcal{E}$), then the Kruskal rank of $\mathcal{E}$ provides a bound on the number of noise edges that can be tolerated between these two sets of nodes. As such kruskal($\mathcal{E}$) plays a central

role in our main result (see Theorem 1) in providing an upper bound on the probability of a noise edge in $\mathfrak{G}$ under the Erdős-Rényi model.

Returning to the three conditions on M_{12} stated at the beginning of this section, if an eigenspace $\mathcal{E}$ of A_Γ has large Kruskal rank and if the size of each $\Omega_j,\ j = 1, \ldots, n-k$ is smaller than kruskal($\mathcal{E}$), then there is an affine space (of dimension potentially larger than zero) of matrices in $\mathbb{R}^{k\times(n-k)}$ that satisfy the first two requirements on M_{12}. The third condition on M_{12} requires that we find the element of this affine space with the smallest spectral norm:

$$\begin{aligned}\hat{M}_{12}^{\text{spectral}} = \underset{X\in\mathbb{R}^{k\times(n-k)}}{\arg\min}\ & \|X\|_2 \\ \text{s.t. } & X_{i,j} = 1 \text{ if } i \in \Omega_j \text{ for } j = 1, \ldots, n-k \\ & \operatorname{col}(X) \subseteq \mathcal{E}.\end{aligned}$$

As long as $|\Omega_j| \leq$ kruskal($\mathcal{E}$) for each $j = 1, \ldots, n-k$, this problem is feasible. However, analytically characterizing the optimal value and solution of this problem is challenging, especially in the context of problem instances that arise from the Erdős-Rényi model, as the subsets $\Omega_j,\ j = 1, \ldots, n-k$, are random. As a result, a common approach is to replace the objective in the above problem with the Frobenius norm:

$$\begin{aligned}\hat{M}_{12}^{\text{frobenius}} = \underset{X\in\mathbb{R}^{k\times(n-k)}}{\arg\min}\ & \|X\|_F \\ \text{s.t. } & X_{i,j} = 1 \text{ if } i \in \Omega_j \text{ for } j = 1, \ldots, n-k \\ & \operatorname{col}(X) \subseteq \mathcal{E}.\end{aligned} \tag{2.6}$$

One of the virtues of this latter formulation in comparison to the earlier one is that the spectral norm of the optimal solution $\|\hat{M}_{12}^{\text{frobenius}}\|_2$ is more tractable to bound, primarily since the optimization problem (2.6) decomposes into $n-k$ separable problems, one for each column of the decision variable X. In particular, for any subspace $\mathcal{S} \subseteq \mathbb{R}^k$ and any $\Omega \subset \{1, \ldots, k\}$ with $|\Omega| \leq$ kruskal($\mathcal{S}$), consider the following minimum Euclidean-norm completion:

$$\begin{aligned}q_\Omega(\mathcal{S}) &\triangleq \underset{q\in\mathbb{R}^k}{\arg\min}\ \|q\| \quad \text{s.t. } q \in \mathcal{S} \text{ and } q_i = 1 \text{ for } i \in \Omega \\ &= \mathcal{P}_\mathcal{S} I_\Omega{}' ((\mathcal{P}_\mathcal{S})_{\Omega,\Omega})^{-1} 1_{|\Omega|}.\end{aligned} \tag{2.7}$$

With this notation, the j'th column of $\hat{M}_{12}^{\text{frobenius}}$ is given by $q_{\Omega_j}(\mathcal{E})$. Further, under the Erdős-Rényi model, the entries $(A_\mathfrak{G})_{i,j+k},\ i = 1, \ldots, k; j =$

$1, \ldots, n-k$ are independent and identically distributed Bernoulli random variables. In such a family of problem instances, the columns of $\hat{M}_{12}^{\text{frobenius}}$, i.e., $q_{\Omega_j}(\mathcal{E}) \in \mathbb{R}^k,\ j = 1, \ldots, k$, are independently and identically distributed random vectors. These observations in conjunction with the following tail bound on the spectral norm of a random matrix suggest a natural invariant of $\mathcal{E}$ that leads to bounds on $\|\hat{M}_{12}^{\text{frobenius}}\|_2$:

Lemma 10. *[118] Let A be a $d \times N$ matrix ($d < N$) with columns A_i and let $\Sigma = \mathbb{E}[A_i A_i^T]$ denote the correlation matrix of the A_i's. Further, suppose there exists $m \in \mathbb{R}$ such that $\|A_i\| \leq \sqrt{m}$ almost surely for all i. Then $\forall x \geq (N\|\Sigma\|_2)^{1/2}$ we have that*

$$\mathbb{P}(\|A\|_2 \geq x) \leq 2d \exp\Big(- \frac{3(x^2 - N\|\Sigma\|_2)^2}{4m(x^2 + 2N\|\Sigma\|_2)} \Big). \tag{2.8}$$

Proof. The proof follows that of Theorem 5.41 in [118] with minor modifications. We apply the non-commutative Bernstein Inequality to $\frac{1}{N}x^2 - \|\Sigma\|_2$ rather than to $\max(\delta, \delta^2)$ on p.27 of [118], and we don't make the isotropy assumption. □

To apply Lemma 10 to obtain a bound on $\|\hat{M}_{12}^{\text{frobenius}}\|_2$, we describe next the second key invariant of $\mathcal{E}$, which is essentially the correlation matrix in Lemma 10.

Definition 11. *Let $\mathcal{S} \subseteq \mathbb{R}^k$ be a subspace. Then the* combinatorial width *of $\mathcal{S}$ for each $\ell = 1, \ldots, \text{kruskal}(\mathcal{S})$ and $p \in [0, 1)$ is defined as:*

$$\omega(\mathcal{S}, \ell, p) \triangleq \Big\| \mathbb{E}[q_\Omega(\mathcal{S})\, q_\Omega(\mathcal{S})' \,\big|\, |\Omega| \leq \ell] \Big\|_2,$$

with the expectation taken over Ω, where each element of $\{1, \ldots, k\}$ is contained in Ω independently with probability p.

The conditioning in the definition ensures that $q_\Omega(\mathcal{S})$ is well-defined as $|\Omega| \leq \text{kruskal}(\mathcal{S})$. We utilize this terminology as a parallel to analogous notions such as 'mean width' that are prominent in the convex geometry literature. The explicit appearance of ℓ in this definition allows for a more fine-grained analysis in our main result Theorem 1; see Section 2.3. Based on the following result, the Kruskal rank and the combinatorial width play a central role in Theorem 1 as the success of the Schur-Horn relaxation (P) relies on the existence of an eigenspace $\mathcal{E}$ of A_Γ that has large Kruskal rank and small combinatorial width.

Proposition 12. *Consider a planted subgraph problem instance in which the nodes of $\mathfrak{G}$ and Γ are labeled so that the leading principal minor of $A_{\mathfrak{G}}$ of order k is equal to A_Γ, and the remaining edges in $\mathfrak{G}$ are drawn according to the Erdős-Rényi model with probability $p \in [0, \frac{\text{kruskal}(\mathcal{E})}{k})$. Fix any $\ell \in \mathbb{Z}$ satisfying $kp < \ell \leq \text{kruskal}(\mathcal{E})$, and denote $\zeta := \min_{\substack{\Omega \subset \{1,\ldots,k\} \\ |\Omega| \leq \ell}} \lambda_{\min}\big((\mathcal{P}_{\mathcal{E}})_{\Omega,\Omega}\big)$. For any $\delta \geq \sqrt{(n-k)\,\omega(\mathcal{E},\ell,p)}$, there exists a matrix $M_{12} \in \mathbb{R}^{k\times(n-k)}$ satisfying the following properties:*

1. *Each column of M_{12} lies in $\mathcal{E}$,*
2. *$(M_{12})_{i,j} = (A_{\mathfrak{G}})_{i,\,j+k}$ if $(A_{\mathfrak{G}})_{i,\,j+k} = 1$,*
3. *$\|M_{12}\|_2 < \delta$,*

with probability at least $\Big(1 - 2k\exp\big(-\frac{3\zeta(\delta^2-(n-k)\,\omega(\mathcal{E},\ell,p))^2}{4\,\ell(\delta^2+2(n-k)\omega(\mathcal{E},\ell,p))}\big)\Big)\big(1 - \exp\big(-\frac{(\ell-kp)^2}{\ell+kp}\big)\big)^{n-k}$.

Proof. We bound the probability that $\hat{M}_{12}^{\text{frobenius}}$ obtained as the optimal solution of (2.6) satisfies the requirements of this proposition.

We begin by bounding the cardinality of each Ω_j for $j = 1, \ldots, n-k$. Under the Erdős-Rényi model, each $|\Omega_j|$ follows a binomial distribution. Consequently, using the Chernoff bound we have for each $j = 1, \ldots, n-k$ that:

$$\begin{aligned}\mathbb{P}(|\Omega_j| \geq \ell+1) \leq \mathbb{P}(|\Omega_j| \geq \ell) &= \mathbb{P}\left(|\Omega_j| \geq \Big(1 + \frac{\ell - kp}{kp}\Big)kp\right) \\ &\leq \exp\left(-\frac{(\ell-kp)^2}{\ell+kp}\right).\end{aligned}$$

The first inequality is not essential and it is simply used to avoid notational clutter. Based on the independence of the Ω_j's,

$$\mathbb{P}(|\Omega_j| \leq \ell,\ j = 1, \ldots, n-k) \geq \left(1 - \exp\Big(-\frac{(\ell-kp)^2}{\ell+kp}\Big)\right)^{n-k}. \tag{2.9}$$

This inequality provides a bound on the probability that the optimization problem (2.6) is feasible.

In our next step we bound $\|\hat{M}_{12}^{\text{frobenius}}\|_2$ via Lemma 10. As $\ell \leq \text{kruskal}(\mathcal{E})$ one can check that $\zeta > 0$. Further, from (2.7) we have that $\|q_{\Omega_j}\|^2 \leq \frac{|\Omega_j|}{\zeta}$. Thus, by applying Lemma 10, we deduce that

$$\mathbb{P}(\|M_{12}\|_2 < \delta \;\big|\; |\Omega_j| \leq \ell \;\forall j) \geq 1 - 2k \exp\left(- \frac{3\zeta(\delta^2 - (n-k)\omega(\mathcal{E},\ell,p))^2}{4\,\ell(\delta^2 + 2(n-k)\omega(\mathcal{E},\ell,p))} \right). \tag{2.10}$$

The final result follows by combining (2.9) and (2.10). □

Properties of Kruskal Rank and Combinatorial Width

Beyond the utility of the Kruskal rank and combinatorial width in characterizing the performance of the Schur-Horn relaxation, these graph parameters are also of intrinsic interest and we discuss next their relationship to structural properties of Γ.

Invariance under Complements for Regular Graphs

Both the Kruskal rank and the combinatorial width are preserved under graph complements for connected regular graphs. Suppose Γ is a connected regular graph on k vertices, and let $A_\Gamma \in \mathbb{S}^k$ be an adjacency matrix representing Γ for some labeling of the nodes. Then the eigenspaces of A_Γ are the same as those of the adjacency matrix A_{Γ^c} of the complement Γ^c based on the following relation:

$$A_{\Gamma^c} = 1_k 1_k' - I_k - A_\Gamma. \tag{2.11}$$

As Γ is connected and regular, the vector 1_k is an eigenvector of A_Γ. Thus, the Kruskal ranks and the combinatorial widths associated to the eigenspaces of A_Γ are the same as those associated to the eigenspaces of A_{Γ^c}.

Combinatorial Width for Symmetric Graphs

For graphs Γ that are symmetric – vertex- end edge-transitive – and also have symmetric complements Γ^c, the combinatorial width of any eigenspace $\mathcal{E}$ of A_Γ can be characterized in terms of the minimum singular values of minors of $\mathcal{P}_\mathcal{E}$. In particular, we establish our result by demonstrating that the correlation matrix $\mathbb{E}[q_\Omega(\mathcal{S})\, q_\Omega(\mathcal{S})' \,\big|\, |\Omega| \leq \ell]$ in the definition of the combinatorial width has the property that all its nonzero eigenvalues are equal to each other, which leads to bounds on the combinatorial width via bounds on the trace of the correlation matrix.

Proposition 13. *Let $A_\Gamma \in \mathbb{S}^k$ be an adjacency matrix of a (connected) symmetric graph Γ with a symmetric complement Γ^c, and let $\mathcal{E} \subset \mathbb{R}^k$ be an eigenspace of A_Γ. Fix any $\ell \in \mathbb{Z}$ and $p \in [0,1)$ such that $kp \leq \ell \leq \text{kruskal}(\mathcal{E})$, and let $\zeta := \min_{\substack{\Omega \subset \{1,\ldots,k\} \\ |\Omega| \leq \ell}} \lambda_{\min}\big((\mathcal{P}_\mathcal{E})_{\Omega,\Omega}\big)$. Then,*

$$\omega(\mathcal{E}, \ell, p) \;\leq\; \frac{2kp}{\zeta \dim(\mathcal{E})}.$$

Proof. Denote the correlation matrix in the definition of the combinatorial width as follows:

$$\Sigma = \mathbb{E}\left[q_\Omega(\mathcal{E}) q_\Omega(\mathcal{E})' \,\middle|\, |\Omega| \leq \ell\right] = \sum_{i=0}^{\ell} c_{p,\ell}\, p^i (1-p)^{k-i} \sum_{|\Omega|=i} q_\Omega(\mathcal{E}) q_\Omega(\mathcal{E})', \quad (2.12)$$

where the term $c_{p,\ell} = \big(\sum_{i=0}^{\ell} \binom{k}{i} p^i (1-p)^{k-i}\big)^{-1}$ is the normalization constant.

The main element of the proof is to show that the rank of $\Sigma \in \mathbb{S}^k$ is equal to $\dim(\mathcal{E})$ (it is easily seen that $\text{col}(\Sigma) \subseteq \mathcal{E}$) and that all the nonzero eigenvalues of Σ are equal to each other. After this step is completed, one can bound the combinatorial width using the following relation:

$$\omega(\mathcal{E}, \ell, p) = \frac{\text{Tr}(\Sigma)}{\dim(\mathcal{E})}. \quad (2.13)$$

In particular, we have $q_\Omega(\mathcal{E}) = \mathcal{P}_\mathcal{E} I_\Omega{}' ((\mathcal{P}_\mathcal{E})_{\Omega,\Omega})^{-1} 1_{|\Omega|}$ with $|\Omega| \leq \text{kruskal}(\mathcal{E})$. One can check that $\|q_\Omega(\mathcal{E})\|^2 \leq \frac{|\Omega|}{\zeta}$, and then obtain that:

$$\text{Tr}(\Sigma) = \sum_{i=0}^{\ell} c_{p,\ell}\; p^i (1-p)^{k-i} \sum_{|\Omega|=i} \|q_\Omega\|^2 \leq 2 \sum_{i=0}^{\ell} \binom{k}{i} p^i (1-p)^{k-i} \frac{i}{\zeta} \leq \frac{2\,k\,p}{\zeta}. \quad (2.14)$$

The first inequality follows from the implication that $kp \leq \ell \Rightarrow c_{p,\ell} \leq 2$. The second inequality is obtained by bounding the sum from above with the expectation of a binomial random variable with parameters k and p. Combining (2.13) and (2.14) we have the desired result.

To complete the proof, we need to show that $\text{rank}(\Sigma) = \dim(\mathcal{E})$ and that all the nonzero eigenvalues of Σ are equal to each other. For each $i \leq \ell$ denote $S^{(i)} := \sum_{|\Omega|=i} I_\Omega{}' ((\mathcal{P}_\mathcal{E})_{\Omega,\Omega})^{-1} 1_{|\Omega|} 1_{|\Omega|}{}' ((\mathcal{P}_\mathcal{E})_{\Omega,\Omega})^{-1} I_\Omega$ so that $\sum_{|\Omega|=i} q_\Omega(\mathcal{E}) q_\Omega(\mathcal{E})' = \mathcal{P}_\mathcal{E} S^{(i)} \mathcal{P}_\mathcal{E}$. Let $\Pi \in \mathbb{R}^{k \times k}$ be a permutation matrix such that $\Pi A_\Gamma \Pi' = A_\Gamma$, i.e., Π corresponds to an element of the automorphism group of Γ. It is easily seen that

$\Pi\mathcal{P}_{\mathcal{E}}\Pi = \mathcal{P}_{\mathcal{E}}$. Consequently, if a vertex subset $\Omega \subset \{1, \ldots, k\}$ is mapped to $\hat{\Omega}$ under the automorphism represented by Π, then we have that $(\mathcal{P}_{\mathcal{E}})_{\Omega,\Omega} = (\mathcal{P}_{\mathcal{E}})_{\hat{\Omega},\hat{\Omega}}$. In turn, one can check that $\Pi\, {I_\Omega}'((\mathcal{P}_{\mathcal{E}})_{\Omega,\Omega})^{-1}1_{|\Omega|}{1_{|\Omega|}}'((\mathcal{P}_{\mathcal{E}})_{\Omega,\Omega})^{-1}I_\Omega\, \Pi'$ $= {I_{\hat{\Omega}}}'((\mathcal{P}_{\mathcal{E}})_{\hat{\Omega},\hat{\Omega}})^{-1}1_{|\hat{\Omega}|}{1_{|\hat{\Omega}|}}'((\mathcal{P}_{\mathcal{E}})_{\hat{\Omega},\hat{\Omega}})^{-1}I_{\hat{\Omega}}$. Based on these observations and the fact that $|\Omega| = i \Leftrightarrow |\hat{\Omega}| = i$, we note that a summand of $S^{(i)}$ gets mapped to another summand of $S^{(i)}$ under conjugation by Π. Moreover, automorphisms are injective functions, and hence distinct summands of $S^{(i)}$ must be mapped to distinct summands of $S^{(i)}$. Thus, we conclude that $\Pi S^{(i)}\Pi' = S^{(i)}$ for each $i \leq \ell$ and for any permutation matrix $\Pi \in \mathbb{R}^{k\times k}$ representing an automorphism of Γ.

As Γ is vertex- and edge-transitive, and as Γ^c is also edge-transitive, each $S^{(i)}$ is of the following form:

$$(S^{(i)})_{p,q} = \begin{cases} \alpha_1, \text{ if } (A_\Gamma)_{p,q} = 1 \text{ and } p \neq q \\ \alpha_2, \text{ if } (A_\Gamma)_{p,q} = 0 \text{ and } p \neq q \\ \alpha_3, \text{ if } p = q \end{cases} \implies S^{(i)} = \alpha_1 A_\Gamma + \alpha_2 A_{\Gamma^c} + \alpha_3 I_k, \tag{2.15}$$

for some $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$. Since Γ is vertex-transitive it is also a regular graph, and consequently the discussion from Section 2.3 implies that the eigenspaces of A_Γ and A_{Γ^c} are the same. As Γ is assumed to be connected, we have from equations (2.11), (2.15) and from the equality $\sum\limits_{|\Omega|=i} q_\Omega(\mathcal{E})q_\Omega(\mathcal{E})' = \mathcal{P}_{\mathcal{E}} S^{(i)} \mathcal{P}_{\mathcal{E}}$ that:

$$\sum_{|\Omega|=i} q_\Omega(\mathcal{E})q_\Omega(\mathcal{E})' = \begin{cases} [\alpha_1\lambda_{\mathcal{E}} + \alpha_2(k - \lambda_{\mathcal{E}} - 1) + \alpha_3]\frac{1_k 1_k{}^T}{k}, \text{ if } \mathcal{E} = \text{span}\{1_k 1_k'\}, \\ [\alpha_1\lambda_{\mathcal{E}} - \alpha_2(\lambda_{\mathcal{E}} + 1) + \alpha_3]\mathcal{P}_{\mathcal{E}}, \text{ otherwise.} \end{cases}$$

Since this holds for each index $i \leq \ell$, we deduce that every summand of Σ in equation (2.12) is a scalar multiple of $\mathcal{P}_{\mathcal{E}}$, and consequently, so is Σ. Furthermore, by construction, Σ cannot be the all zeros matrix. Therefore, $\text{rank}(\Sigma) = \dim(\mathcal{E})$ and all nonzero eigenvalues of Σ are equal to each other. The result follows immediately. □

Simplifications based on Coherence

The Kruskal rank of a subspace is intractable to compute in general; as a result, a number of subspace parameters have been considered in the literature to obtain tractable bounds on the Kruskal rank. The most prominent

of these is the coherence parameter of a subspace. In our context, the additional analytical simplification provided by the coherence of a subspace along with Proposition 13 lead to simple performance guarantees on the Schur-Horn relaxation for symmetric planted graphs.

Definition 14. *Let $\mathcal{S} \subseteq \mathbb{R}^k$ be a subspace. The* coherence *of $\mathcal{S}$, denoted $\mu(\mathcal{S})$, is defined as:*

$$\mu(\mathcal{S}) := \max_{\substack{1 \le i,j \le k \\ i \neq j}} \frac{|(\mathcal{P}_{\mathcal{S}})_{i,j}|}{((\mathcal{P}_{\mathcal{S}})_{i,i})^{1/2}((\mathcal{P}_{\mathcal{S}})_{j,j})^{1/2}}.$$

The coherence parameter of a subspace can be computed efficiently, and it can be used to bound the Kruskal rank from below:

Proposition 15. *[45] For any subspace $\mathcal{S} \in \mathbb{R}^k$,* $\operatorname{kruskal}(\mathcal{S}) \geq \frac{1}{\mu(\mathcal{S})}$.

Further, for symmetric planted graphs Γ, the following result provides a bound on the minimum eigenvalue of minors of $\mathcal{P}_{\mathcal{E}}$ for eigenspaces $\mathcal{E}$ of A_Γ. Recall that this result is directly relevant in the context of Proposition 13.

Proposition 16. *Suppose Γ is a vertex-transitive graph with adjacency matrix $A_\Gamma \in \mathbb{S}^k$, and let $\mathcal{E}$ denote an eigenspace of A_Γ. For any $\ell \in \mathbb{Z}$ with $\ell < \frac{1}{\mu(\mathcal{E})} + 1$, we have that* $\min_{\substack{\Omega \subset \{1,\ldots,k\} \\ |\Omega| \leq \ell}} \lambda_{\min}\big((\mathcal{P}_{\mathcal{E}})_{\Omega,\Omega}\big) \geq \frac{\dim(\mathcal{E})}{k}\big(1 - (\ell - 1)\mu(\mathcal{E})\big)$.

Proof. One can check that $\Pi \mathcal{P}_{\mathcal{E}} \Pi' = \mathcal{P}_{\mathcal{E}}$ for permutation matrices $\Pi \in \mathbb{R}^{k \times k}$ that correspond to automorphisms of Γ. Therefore, by vertex transitivity, the diagonal entries of $\mathcal{P}_{\mathcal{E}}$ are all equal to each other. As $\operatorname{Tr}(\mathcal{P}_{\mathcal{E}}) = \dim(\mathcal{E})$, we conclude that $(\mathcal{P}_{\mathcal{E}})_{i,i} = \frac{\dim(\mathcal{E})}{k}$ for each $i = 1, \ldots, k$. Every row of $(\mathcal{P}_{\mathcal{E}})_{\Omega,\Omega}$ has at most $\ell - 1$ off-diagonal entries, and each of these entries is bounded above by $\frac{\dim(\mathcal{E})}{k}\mu(\mathcal{E})$. We obtain the desired result by applying the Gershgorin circle theorem. □

Main Result

Building on the preceding discussion, we state and prove our main result Theorem (1). The proof of this result relies on an intermediate step regarding the M_{22} submatrix of the dual variable $M = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}' & M_{22} \end{pmatrix}$ from Proposition 8. From that result, we are required to obtain an $M_{22} \in \mathbb{S}^{n-k}$ such that (i) for

each $i, j = 1, \ldots, n-k$ we have $(M_{22})_{i,j} = 1$ if $(A_{\mathfrak{G}})_{i+k,j+k} = 1$ or if $i = j$, and (ii) the operator norm $\|M_{22}\|_2$ is as small as possible.

We present the following result from [8], which we utilize subsequently in Lemma 18 to establish a bound on $\|M_{22}\|_2$:

Lemma 17. *[8] Let $X \in \mathbb{S}^d$ be a symmetric matrix whose entries $X_{i,j}$ are independent and centered random variables. For each $\epsilon \in (0, 1/2]$, there exists a constant $\tilde{c}_\epsilon$ such that for all $x \geq 0$:*

$$\mathbb{P}(\|X\|_2 \geq (1+\epsilon)2\tilde{\sigma} + x) \leq d \exp\left(-\frac{x^2}{\tilde{c}_\epsilon \tilde{\sigma}_*^2}\right),$$

where $\tilde{\sigma} := \max_i \sqrt{\sum_j \mathbb{E}[X_{i,j}^2]}$ and each $|X_{i,j}| \leq \tilde{\sigma}_*$ almost surely.

Lemma 18. *Consider a planted subgraph problem instance in which the nodes of $\mathfrak{G}$ and Γ are labeled so that the leading principal minor of $A_{\mathfrak{G}}$ of order k is equal to A_Γ, and the remaining edges in $\mathfrak{G}$ are drawn according to the Erdős-Rényi model with probability $p \in [0, 1)$. For constants $c_1 = \sqrt{\frac{9p}{1-p}}$ and c_2 depending only on p and for $\alpha \geq c_1\sqrt{n-k}$, there exists $M_{22} \in \mathbb{S}^{n-k}$ satisfying*

1. *$(M_{22})_{i,j} = 1$ if $(A_{\mathfrak{G}})_{i+k,j+k} = 1$ or $i = j$,*
2. *$\|M_{22}\|_2 < \alpha$,*

with probability at least $1 - (n-k)\exp\Big(-c_2\big(\alpha - c_1\sqrt{n-k}\big)^2\Big)$.

Proof. Our proof is inspired by the approach in [5]. Consider the following matrix $M_{22} \in \mathbb{S}^{n-k}$:

$$(M_{22})_{i,j} = \begin{cases} 1, \text{ if } (A_{\mathfrak{G}})_{i+k,j+k} = 1, i \neq j \\ \frac{-p}{1-p}, \text{ if } (A_{\mathfrak{G}})_{i+k,j+k} = 0, i \neq j \\ 0, \text{ if } i = j. \end{cases} \tag{2.16}$$

As the submatrix $(A_{\mathfrak{G}})_{i+k,j+k},\ i, j = 1, \ldots, n-k$ consists of independent and centered entries (in the off-diagonal locations) and zeros on the diagonal, one can check that M_{22} is a random matrix that satisfies the requirements of Lemma 17. Further, the first part of the present lemma is satisfied. The second claim follows from an application of Lemma 17 with $\epsilon = 1/2$. □

Combining Proposition 8, Proposition 12, and Lemma 18, we now state and prove the main result of this chapter:

Theorem 1. *Consider a planted subgraph problem instance in which the nodes of $\mathfrak{G}$ and Γ are labeled so that the leading principal minor of $A_{\mathfrak{G}}$ of order k is equal to A_Γ, and the remaining edges in $\mathfrak{G}$ are drawn according to the Erdős-Rényi model. Suppose $\mathcal{E} \subset \mathbb{R}^k$ is an eigenspace of A_Γ with associated eigenvalue $\lambda_{\mathcal{E}}$, and we employ the Schur-Horn relaxation (P) with parameter $\gamma = \lambda_{\mathcal{E}}$. Further suppose that:*

1. $p \in [0, \frac{\text{kruskal}(\mathcal{E})}{k})$,

and for some $\ell \in \mathbb{Z}$ satisfying $kp < \ell \leq \text{kruskal}(\mathcal{E})$,

2. $n < \min\left(\frac{\text{eigengap}(A_\Gamma,\mathcal{E})^2}{4\omega(\mathcal{E},\ell,p)}, \frac{(\text{eigengap}(A_\Gamma,\mathcal{E})-2|\lambda_{\mathcal{E}}|)^2}{4c_1^2}\right) + k.$

Then the Schur-Horn relaxation succeeds at identifying the planted subgraph Γ inside $\mathfrak{G}$ with probability at least $1 - p_1 - p_2$, where $p_1 = 1 - \left[\left(1 - \exp\left(-\frac{(\ell-kp)^2}{\ell+kp}\right)\right)^{n-k} \times \left(1 - 2k\exp\left(-\frac{3\zeta\left(\frac{1}{4}\text{eigengap}(A_\Gamma,\mathcal{E})^2-(n-k)\omega(\mathcal{E},\ell,p)\right)^2}{4\ell\left(\frac{1}{4}\text{eigengap}(A_\Gamma,\mathcal{E})^2+2(n-k)\omega(\mathcal{E},\ell,p)\right)}\right)\right)\right]$ and $p_2 = (n-k) \times \exp\left(-c_2\left(\frac{1}{2}\text{eigengap}(\text{A}_\Gamma,\mathcal{E}) - |\lambda_{\mathcal{E}}| - c_1\sqrt{n-k}\right)^2\right)$. Here $\zeta = \min_{\substack{\Omega\subset\{1,\ldots,k\}\\|\Omega|\leq\ell}} \lambda_{\min}\left((\mathcal{P}_{\mathcal{E}})_{\Omega,\Omega}\right)$, and the constants $c_1 = \sqrt{\frac{9p}{1-p}}$ and c_2 depend only on p.

Proof. As discussed previously, since $\ell \leq \text{kruskal}(\mathcal{E})$ we have that $\zeta > 0$. We establish the result by constructing a dual certificate M satisfying the conditions of Proposition 8.

We start by setting $M_{11} = A_\Gamma$. This ensures that conditions (ii) and (iii) of Proposition 8 are satisfied. Next, we choose M_{12} as discussed in Proposition 12, with the parameter $\delta = \frac{1}{2}\,\text{eigengap}(\Gamma,\mathcal{E})$, which satisfies $\delta \geq \sqrt{(n-k)\omega(\mathcal{E},\ell,p)}$ due to the upper bound on n. Such an M_{12} exists with probability at least $1-p_1$, and it satisfies condition (iv) of Proposition 8 as well as the bound $\|M_{12}\|_2 < \frac{1}{2}\text{eigengap}(A_\Gamma,\mathcal{E})$. Finally, we set M_{22} as discussed in Lemma 18, with $\alpha = \frac{1}{2}\text{eigengap}(A_\Gamma,\mathcal{E}) - |\lambda_{\mathcal{E}}|$, which satisfies $\alpha \geq c_1\sqrt{n-k}$

due to the upper bound on n. Such an M_{22} exists with probability at least $1 - p_2$ and satisfies the bound $\|M_{22}\|_2 < \frac{1}{2}\text{eigengap}(A_\Gamma, \mathcal{E}) - |\lambda_\mathcal{E}|$.

Based on this construction, the matrix $M = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}' & M_{22} \end{pmatrix}$ satisfies conditions (i) and (v) of Proposition 8. Thus, if M_{12} and M_{22} with the stated properties exist, then all the conditions of Proposition 8 are satisfied. By the union bound, the desired M_{12} and M_{22} exist concurrently with probability at least $1 - p_1 - p_2$. □

Remark 19. *The parameter ℓ arises in multiple aspects of this result. We discuss specific choices of ℓ in the corollaries in the next section.*

Theorem 1 provides a non-asymptotic bound on the performance of the Schur-Horn relaxation (P). In words, this relaxation succeeds with high probability in identifying a subgraph Γ planted inside a larger graph $\mathfrak{G}$ (under the Erdős-Rényi model) provided A_Γ has an eigenspace $\mathcal{E}$ satisfying four conditions: (i) The eigenspace $\mathcal{E}$ has large Kruskal rank, (ii) The eigenspace $\mathcal{E}$ has small combinatorial width, (iii) A_Γ has a large eigengap with respect to $\mathcal{E}$, and (iv) The projection matrix $\mathcal{P}_\mathcal{E}$ has the property that all sufficiently large principal minors are well-conditioned. In practice, larger dimensional eigenspaces of A_Γ may be expected to satisfy these conditions more easily, and therefore we set γ equal to the eigenvalue of A_Γ of largest multiplicity in our experimental demonstrations in Section 2.4.

Specializations of Theorem 1

We appeal to the discussion in Section 2.3 on the properties of the Kruskal rank and the combinatorial width to obtain specializations of Theorem 1 to certain graph families. We begin by considering the case of symmetric planted graphs with symmetric complements:

Corollary 20. *Consider a planted subgraph problem instance in which the nodes of $\mathfrak{G}$ and Γ are labeled so that the leading principal minor of $A_\mathfrak{G}$ of order k is equal to A_Γ, and the remaining edges in $\mathfrak{G}$ are drawn according to the Erdős-Rényi model. Suppose $\mathcal{E} \subset \mathbb{R}^k$ is an eigenspace of A_Γ with associated eigenvalue $\lambda_\mathcal{E}$, and we employ the Schur-Horn relaxation (P) with parameter $\gamma = \lambda_\mathcal{E}$. Further suppose that the following three conditions hold:*

1. *Γ is a connected symmetric graph with a symmetric complement,*

2. $p \in [0, \frac{1}{\mu(\mathcal{E})k})$,

3. $n < \min\left(\frac{\text{eigengap}(A_\Gamma,\mathcal{E})^2 \dim(\mathcal{E})^2 \left(1-kp\mu(\mathcal{E})\right)}{16k^2p}, \frac{(\text{eigengap}(A_\Gamma,\mathcal{E})-2|\lambda_\mathcal{E}|)^2}{4c_1^2}\right) + k.$

Then the Schur-Horn relaxation succeeds in identifying the planted subgraph Γ *inside the larger graph* $\mathfrak{G}$ *with probability at least* $1 - p_1 - p_2$*, where* p_1 *and* p_2 *are as stated in Theorem 1 (one can substitute* $\frac{4k^2p}{\dim(\mathcal{E})^2\left(1-kp\mu(\mathcal{E})\right)}$ *for the* $\omega(\mathcal{E}, \ell, p)$ *term appearing in* p_1*).*

Proof. This result follows by a combination of Theorem 1, and Propositions 13, 15, 16. Set $\ell = \lceil \frac{1}{2}(kp + \frac{1}{\mu(\mathcal{E})})\rceil$. This choice satisfies $kp < \ell \leq \text{kruskal}(\mathcal{E})$ based on Proposition 15. One can also check that the inequality $\ell < \frac{1}{\mu(\mathcal{E})} + 1$ holds. The vertex transitivity of Γ implies that one can appeal to Proposition 16 to conclude that

$$\min_{\substack{\Omega\subset\{1,\ldots,k\}\\|\Omega|\leq\ell}} \lambda_{\min}\big((\mathcal{P}_\mathcal{E})_{\Omega,\Omega}\big) \geq \frac{\dim(\mathcal{E})}{k}\big(1 - (\ell-1)\mu(\mathcal{E})\big) > \frac{\dim(\mathcal{E})}{2k}(1 - kp\mu(\mathcal{E})). \tag{2.17}$$

Based on the condition on p, this lower bound is strictly positive. As Γ is symmetric and has a symmetric complement (and is connected), we conclude from Proposition 13 that $\omega(\mathcal{E}, \ell, p) \leq \frac{4k^2p}{\dim(\mathcal{E})^2(1-kp\mu(\mathcal{E}))}$.

Finally, one can check that conditions (2) and (3) of the corollary imply that both of the requirements of Theorem 1 are met, and hence the Schur-Horn relaxation succeeds in identifying the planted subgraph Γ with probability at least $1-p_1-p_2$, where p_1 and p_2 are as stated in Theorem 1 – one can substitute $\frac{4k^2p}{\dim(\mathcal{E})^2\left(1-kp\mu(\mathcal{E})\right)}$ as an upper bound for $\omega(\mathcal{E}, \ell, p)$ and $\frac{\dim(\mathcal{E})}{2k}(1 - kp\mu(\mathcal{E}))$ as a lower bound for ζ, which yields a lower bound on $1 - p_1 - p_2$ from Theorem 1. □

As the coherence parameter of an eigenspace is more tractable to compute than the Kruskal rank, this result provides an efficiently verifiable set of conditions that guarantee the success of the Schur-Horn relaxation (P) for symmetric planted graphs Γ. This result specialized to the case of the planted clique problem yields the result of Ames and Vavasis [5].

Corollary 21. *Fix* $p \in [0, 1)$ *and consider a family of planted clique problem instances* $\{\Gamma_k, \mathfrak{G}_k\}_{k=1}^\infty$ *generated according to the Erdős-Rényi model, where* Γ_k

is the k-clique and $\mathfrak{G}_k$ is a graph on n_k nodes. There exists a constant $\beta > 0$ only depending on p such that if $n_k \leq \beta k^2$, the Schur-Horn relaxation with $\gamma = -1$ succeeds in identifying Γ_k inside $\mathfrak{G}_k$ with probability approaching one exponentially fast in k.

Proof. The k-clique is a connected symmetric graph with a complement that is also symmetric; hence the first condition of Corollary 20 is satisfied. Each $A_{\Gamma_k} \in \mathbb{S}^k$ has a $(k-1)$-dimensional eigenspace $\mathcal{E}$ such that $\mu(\mathcal{E}) = \frac{1}{k-1}$, $\dim(\mathcal{E}) = k-1$, and $\text{eigengap}(A_\Gamma, \mathcal{E}) = k$.

Based on the choice $\ell = \lceil \frac{1}{2}(kp + \frac{1}{\mu(\mathcal{E})}) \rceil$ as in Corollary (20), one can check that $\frac{\dim(\mathcal{E})}{2k}\big(1 - kp\mu(\mathcal{E})\big) = \Theta(1)$, that $\omega(\mathcal{E}, \ell, p) \leq \frac{4k^2 p}{\dim(\mathcal{E})^2\left(1-kp\mu(\mathcal{E})\right)} = \Theta(1)$, and that $\ell - kp = \Theta(k)$.

Set $n_k = \frac{k^2}{32}\min(\frac{1-kp\mu(\mathcal{E})}{2p}, \frac{1}{{c_1}^2}) + k$. One can check that the third condition of Corollary 20 is satisfied with this choice. Moreover, this value of n_k (or any smaller value) yields $\frac{1}{4}\text{eigengap}(A_\Gamma, \mathcal{E})^2 - (n_k - k)\omega(\mathcal{E}, \ell, p) = \Theta(k^2)$ and $\frac{1}{2}\text{eigengap}(A_\Gamma, \mathcal{E}) - 1 - c_1\sqrt{n_k - k} = \Theta(k)$.

By Corollary 20, we conclude that the Schur-Horn relaxation (P) with parameter $\gamma = -1$ identifies a hidden k-clique with probability $1 - p_1 - p_2$, where

$$p_1 = 1 - \big(1 - \exp(-c_4 k)\big)^{n-k}\big(1 - 2k\exp(-c_3 k)\big) \longrightarrow 0, \text{ as } n,\ k \to \infty, \text{ and}$$
$$p_2 = (n-k)\exp(-c_5 k^2) \longrightarrow 0, \text{ as } n,\ k \to \infty,$$

for some constants $c_3 > 0$, $c_4 > 0$, and $c_5 > 0$. □

Thus, Theorem 1 can be specialized to obtain the result presented in [5].

Remark 22. *Theorem 1 can also be specialized to obtain recovery results for other families of planted subgraph problems. As an example, for some fixed integer $t > 1$ and probability p, consider a family of planted subgraph problem instances $\{\Gamma_k, \mathfrak{G}_k\}_{k=t,2t,\ldots}$ in which Γ_k is the complete t-partite graph on k vertices whose independent sets each contain k/t vertices – this graph is the complement of a disjoint union of t cliques with size k/t – and $\mathfrak{G}_k$ is a graph on n_k vertices. By suitably specializing Theorem 1, one can show that the Schur-Horn relaxation succeeds in identifying the planted complete t-partite graph with probability approaching 1 exponentially fast as $k \to \infty$ provided that $p \in [0, \frac{1}{t})$ and $n_k \lesssim k^2$.*

2.4 Numerical Experiments

Semidefinite Descriptions of the Schur-Horn Orbitope

We begin with a discussion of semidefinite representations of the Schur-Horn orbitope $\mathcal{SH}(M)$ for $M \in \mathbb{S}^n$. Specifically, suppose $s_\ell : \mathbb{S}^n \to \mathbb{R}$ denotes the sum of the ℓ-largest eigenvalues of a symmetric matrix for $\ell = 1, \ldots, n$. Then the Schur-Horn orbitope $\mathcal{SH}(M)$ can be described via majorization inequalities on the spectrum [107]:

$$\mathcal{SH}(M) = \{N \in \mathbb{S}^n \mid s_\ell(N) \leq s_\ell(M) \text{ for } 1 \leq \ell \leq n-1, \text{ and } \operatorname{Tr}(N) = \operatorname{Tr}(M)\}. \tag{2.18}$$

As the sublevel sets of the convex functions s_ℓ have tractable semidefinite descriptions [11], one can obtain a lifted polynomial-sized semidefinite representation of $\mathcal{SH}(M)$ for arbitrary $M \in \mathbb{S}^n$. However, specifications of $\mathcal{SH}(M)$ via semidefinite representations of the sublevels sets of s_ℓ involve a total of $O(n)$ additional matrix variables in $\mathbb{S}^n$ and $O(n)$ semidefinite constraints (one for each of the majorization inequalities in (2.18)); in particular, these do not take advantage of any structure in the spectrum of M, such as multiplicities in the eigenvalues.

We discuss next an alternative semidefinite representation of $\mathcal{SH}(M)$ that is based on a modification of the description of $\mathcal{SH}(M)$ presented in [43], and it exploits the multiplicities in the eigenvalues of M so that both the number of additional matrix variables and semidefinite constraints scale with the number of distinct eigenvalues of M rather than the ambient size n of M. Suppose M has q distinct eigenvalues $\lambda_1, \ldots, \lambda_q$ with multiplicities $m_1, \ldots, m_q$. Then one can check that [43]:

$$\begin{aligned}\mathcal{SH}(M) = \Big\{ N \in \mathbb{S}^n \mid \exists Y_i \in \mathbb{S}^n, Y_i \succeq 0,\ i = 1, \ldots, q \text{ such that} \\ N = \sum_{i=1}^{q} \lambda_i Y_i,\ \sum_{i=1}^{q} Y_i = I_n,\ \operatorname{Tr}(Y_i) = m_i \text{ for } i = 1, \ldots, q \Big\}.\end{aligned} \tag{2.19}$$

In this latter description of the Schur-Horn orbitope, both the number of additional matrix variables in $\mathbb{S}^n$ and the number of semidefinite constraints are on the order of the number of distinct eigenvalues of M, which can be far smaller than n for the adjacency matrices of graphs considered in this chapter. In the numerical experiments presented next, we employ the description (2.19) of the Schur-Horn orbitope.

Planted graph Γ [with # vertices]	**Eigenvalues [with multiplicity]**	**Kruskal rank of the largest eigenspace**
Clebsch $[k = 16]$ (Figure 2.1a)	$5[\times 1], -3[\times 5], 1[\times 10]$	5
Generalized quadrangular-$(2, 4)$ $[k = 27]$ (Figure 2.4b)	$10[\times 1], -5[\times 6], 1[\times 20]$	8
8-Triangular $[k = 28]$ (Figure 2.2a)	$12[\times 1], 4[\times 7], -2[\times 20]$	6
9-Triangular $[k = 36]$ (Figure 2.2b)	$14[\times 1], 5[\times 8], -2[\times 27]$	7

Figure 2.6: Planted subgraphs (and associated parameters) for which we demonstrate the utility of the Schur-Horn relaxation. See Figure 2.7 for the associated phase transitions.

Experimental Results

We investigate the performance of the Schur-Horn relaxation (P) in planted subgraph problems with the four planted subgraphs Γ listed in Figure 2.6. For each of these graphs, we set γ equal to the eigenvalue corresponding to the largest eigenspace of the corresponding graph. We vary n (the size of the larger graph $\mathfrak{G}$ inside which Γ is planted) and p (the probability of a noise edge in $\mathfrak{G}$), and we obtain 10 random instances of planted subgraph problems for each value of n and p. In Figure 2.7, we plot the empirical probability of success of the Schur-Horn relaxation for these random trials; the white cells represent a probability of success of one and the black cells represent a probability of success of zero. Our results were obtained using the CVX parser [60, 61] and the SDPT3 solver [112]. In each of the four cases, the Schur-Horn relaxation (P) succeeds in solving the underlying planted subgraph problem for suitably small n and p.

2.5 Discussion

In this chapter, we introduce a new convex relaxation approach for the planted subgraph problem, and we describe families of problem instances for which our method succeeds. Our method generalizes previous convex optimization techniques for identifying planted cliques based on nuclear norm minimization [5], and it is useful for identifying planted subgraphs consisting of few distinct eigenvalues. There are several further directions that arise from our investigation, and we mention a few of these here.

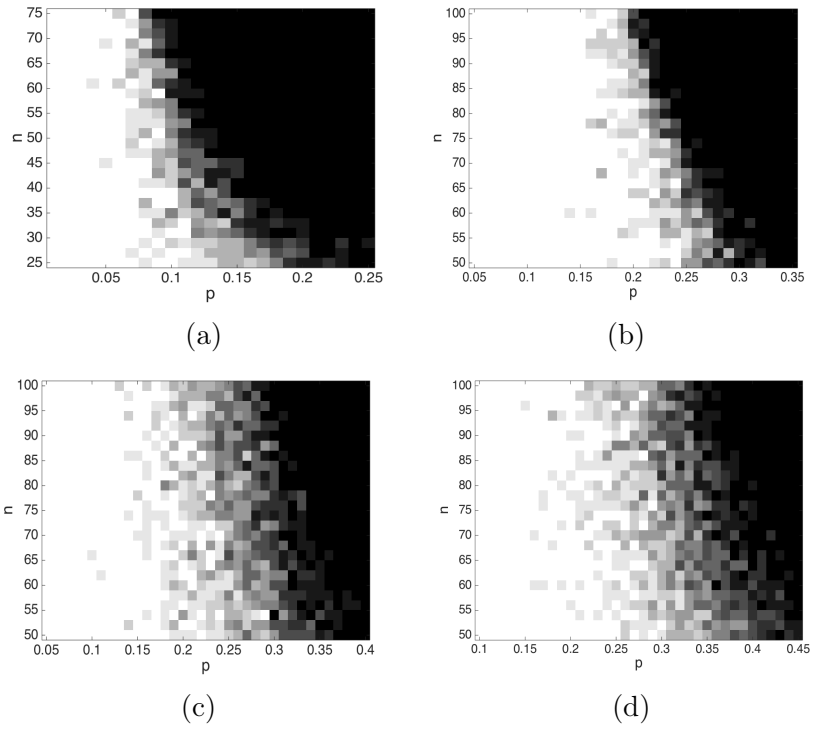


Figure 2.7: Phase transition plots based on the experiment described in Section 2.4 for the (a) Clebsch graph, (b) Generalized quadrangle-$(2, 4)$ graph, (c) 8-Triangular graph, and (d) 9-Triangular graph.

Spectrally comonotone matrices with sparsity constraints One of the ingredients in the proof of our main result Theorem 1 is to find a matrix $M_{11} \in \mathbb{S}^k$ that is spectrally comonotone with $A_\Gamma \in \mathbb{S}^k$, and that further satisfies the condition that $(M_{11})_{i,j} = 1$ whenever $(A_\Gamma)_{i,j} = 1$. In the proof of Theorem 1 we simply choose $M_{11} = A_\Gamma$. This choice does not exploit the fact that the entries of M_{11} corresponding to those where $(A_\Gamma)_{i,j} = 0$ are not constrained (and in particular can be nonzero). With a different choice of M_{11}, one could replace eigengap$(A_\Gamma, \mathcal{E})$ in Theorem 1 by eigengap$(M_{11}, \mathcal{E})$ (recall that $\mathcal{E}$ is an eigenspace of A_Γ). Consequently, our main result could be improved via principled constructions of matrices $M_{11} \in \mathbb{S}^k$ that satisfy the conditions of Theorem 1 and for which eigengap$(M_{11}, \mathcal{E}) >$ eigengap$(A_\Gamma, \mathcal{E})$.

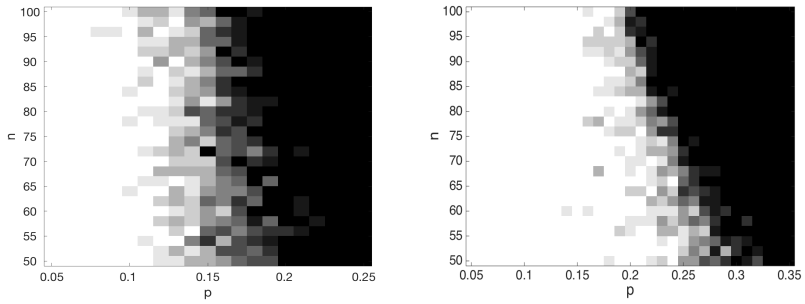
Sparse graphs with eigenspaces with large Kruskal rank One of the central questions concerning the planted subgraph problem is the possibility of identifying 'sparse' planted subgraphs inside 'dense' noise via computationally tractable approaches. Concretely, suppose Γ is a regular graph with degree d. Under the Erdős-Rényi model for the noise, the average degree of any k-node

subgraph of the larger graph $\mathfrak{G}$ is about $(k-1)p$. From Theorem 1, we have that the Schur-Horn relaxation succeeds (with high probability) in identifying Γ if $p \in [0, \frac{\text{kruskal}(\mathcal{E})}{k})$, where $\mathcal{E} \subset \mathbb{R}^k$ is one of the eigenspaces of Γ. In other words, (for suitably large k) if $d < \text{kruskal}(\mathcal{E})$ then the Schur-Horn relaxation succeeds in identifying Γ inside $\mathfrak{G}$ despite the fact that Γ is sparser than a typical k-node subgraph in $\mathfrak{G}$. Of the graphs we have investigated in this chapter, the Clebsch graph from Figure 2.1a is an example in which both the degree and the Kruskal rank of the largest subspace are equal to 5. For some of the other small graphs discussed in this chapter, the degree is larger than the Kruskal ranks of the eigenspaces. For larger graphs, the computation of the Kruskal rank of the large eigenspaces quickly becomes computationally intractable. Therefore, it is of interest to identify graph families in which (by construction) the degree is smaller than the Kruskal rank of one of the eigenspaces.

Convex geometry and graph theory In developing convex relaxations for the planted subgraph problem (based on the formulation (2.1)) as well as other inverse problems involving unlabeled graphs, the key challenge is one of obtaining tractable convex outer approximations of the set $\mathcal{A}(B) = \{\Pi B \Pi' \mid \Pi \text{ is an } n \times n \text{ permutation matrix}\}$ for some given adjacency matrix $B \in \mathbb{S}^n$. In particular, a convex approximation $\mathcal{C}$ that contains $\mathcal{A}(B)$ is useful if the normal cone $\mathcal{N}_{\mathcal{C}}(B)$ is large; as an example, the Schur-Horn relaxation $\mathcal{SH}(B)$ has this property for adjacency matrices B with few distinct eigenvalues. More generally, what is an appropriate convex relaxation for other structured graph families such as low-treewidth graphs (arising in inference in statistical graphical models), or graphs with a specified degree distribution (arising in social network analysis)? Recent work [30] provides a catalog of *convex graph invariants* that are useful for obtaining computationally tractable convex relaxations of $\mathcal{A}(B)$. A deeper investigation of the interaction between convex-geometric aspects of these invariants (such as the normal cones of the associated convex relaxations) and the structural properties of the graph specified by the adjacency matrix B has the potential to yield new convex relaxations for general inverse problems on graphs.

Finding subgraphs vs. induced subgraphs A planted clique is both a subgraph and an induced subgraph of an underlying graph $\mathfrak{G}$. However, for

other planted graphs, this is not necessarily the case, as there could be additional noise edges between the vertices of the planted graph Γ. In this chapter we focus on identifying planted subgraphs that are induced subgraphs, and we investigate the theoretical properties as well as the empirical utility of the Schur-Horn relaxation (P) for this task. For identifying planted subgraphs that may not be induced subgraphs, the experimental results in Figure 2.8 suggest that Schur-Horn relaxation (P) continues to be useful. Specifically, Figure 2.8a gives the performance of the Schur-Horn relaxation in a setting in which the generalized quadrangle-$(2, 4)$ graph is a subgraph of a larger graph on n nodes and noise edges (both outside the planted graph and between disconnected vertices of the planted graph) occur with probability p. In comparison with the result of Section 2.4 (when the generalized quadrangle-$(2, 4)$ graph is an induced subgraph), we notice that the Schur-Horn relaxation performs slightly worse. A reason for this is that identifying structured subgraphs is inherently harder than identifying induced subgraphs due to the additional noise edges. Analyzing the performance of the Schur-Horn relaxation (P) – both theoretically and empirically – in a more extensive manner for identifying subgraphs that may not be induced remains an open question.



(a) (b)

Figure 2.8: Phase transition plot for the problem of identifying the generalized quadrangle-$(2, 4)$ graph: On the left is a plot in which this graph is a subgraph and on the right is a plot in which this graph is an induced subgraph (as in Section 2.4).

Chapter 3

CONVEX GRAPH INVARIANT RELAXATIONS FOR GRAPH EDIT DISTANCE

3.1 Introduction

Graphs are widely used to represent the structure underlying a collection of interacting entities. A common computational question arising in many contexts is that of measuring the similarity between two graphs. For example, the unknown functions of biological structures such as proteins, RNAs, and genes are often deduced from structures which have similar sequences with known functions [67, 69, 89, 109, 110]. Evaluating graph similarity also plays a central role in various pattern recognition applications [35, 95], specifically in areas such as handwriting recognition [51, 87], fingerprint classification [68, 94], and face recognition [121].

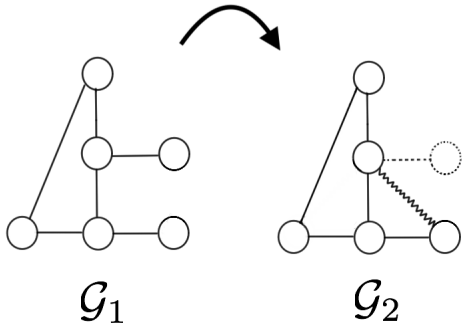


Figure 3.1: An instance of a graph edit distance problem in which we wish to calculate minimum number of edit operations required for transforming graph $\mathcal{G}_1$ to graph $\mathcal{G}_2$. The edit operations are encoded by line style: the dashed graph elements are to be removed from $\mathcal{G}_1$ and the zigzagged graph elements are to be added to $\mathcal{G}_1$ for transforming $\mathcal{G}_1$ to $\mathcal{G}_2$. Assuming a cost of 1 for every edit operation, we conclude that the graph edit distance between $\mathcal{G}_1$ and $\mathcal{G}_2$ is 3.

The notion of similarity that is the most commonly considered is the *graph edit distance* [106]. The edit distance $\mathrm{GED}(\mathcal{G}_1, \mathcal{G}_2)$ between two graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ is the smallest number of operations required to transform $\mathcal{G}_1$ into $\mathcal{G}_2$ by a sequence of edits or changes applied to the vertices and edges of $\mathcal{G}_1$. A particu-

lar sequence of edit operations transforming $\mathcal{G}_1$ into $\mathcal{G}_2$ is usually referred to as an edit path. For unlabeled graphs, the permissible set of edit operations are usually insertions/deletions of vertices/edges. For labeled graphs, the set of permissible edit operations can also include vertex/edge relabelings. In some situations, certain types of edits are considered more 'severe' than others and different edits have different costs associated to them; in such cases, the edit distance is the smallest cost over all edit paths that transform one graph to another, where the cost of an edit path is the sum of the costs of the edits that compose the path. See Figure 3.1 for an illustration of a simple graph edit distance problem.

The problem of computing the graph edit distance is NP-hard in general [55], and in practice exact calculation of the edit distance is only feasible for small-sized graphs. Thus, significant efforts have been directed towards developing computationally tractable heuristics for approximating the edit distance [1, 17, 38, 71, 80, 84, 103, 104, 122] or for exactly computing the edit distance for graphs from specific families such as planar graphs [93]. These methods are largely combinatorial in nature, and most of them aim at identifying an edit path that transforms one graph to the other. Consequently, much of the prior literature on this topic provides techniques that lead to upper bounds on the edit distance between two graphs. In contrast, far fewer approaches have been proposed for obtaining lower bounds on the edit distance. We are aware of three notable examples; in [71] and [104] the authors propose tractable linear programming relaxations of intractable combinatorial optimization formulations of the edit distance based on 0/1 optimization and the quadratic assignment problem, respectively, while in [122] the authors propose a combinatorial algorithm based on a multiset representation of the graph that enables the efficient computation of upper and lower bounds of the graph edit distance.

In this chapter, we develop a computationally efficient framework for obtaining lower bounds on the graph edit distance. Our contributions differ qualitatively from the prior literature in two aspects. First, our approach can be tailored to the structural properties of the specific graphs at hand based on the notion of a *convex graph invariant*. These lead to useful lower bounds on the edit distance if one of the graphs is 'suitably structured'. Second, we give a theoretical analysis of conditions on pairs of graphs under which a tractable semidefinite relaxation based on the spectral properties of a graph provably computes the

edit distance, i.e., our lower bound is tight.

Our Framework

Much of the focus of our development and our analysis is on the edit distance between two unlabeled graphs on the same number of vertices, with edge deletions and insertions being the allowed edit operations. We discuss in Section 3.5 an extension of our framework to settings in which the number of vertices in the two graphs may be different and in which vertex deletions and insertions are also allowed. Let $A_1 \in \mathbb{S}^n$ and $A_2 \in \mathbb{S}^n$ represent the adjacency matrices of two unweighted, unlabeled, simple and loopless graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ on n vertices. (Here $\mathbb{S}^n$ denotes the space of $n \times n$ real symmetric matrices.) The following optimization problem gives a combinatorial formulation of the computation of the edit distance between $\mathcal{G}_1$ and $\mathcal{G}_2$:

$$\begin{aligned} \mathrm{GED}(\mathcal{G}_1, \mathcal{G}_2) = \min_{X, E \in \mathbb{S}^n} \quad & \sum_{1 \leq i < j \leq n} 1_{E_{ij} \neq 0} \\ s.t. \quad & X + E = A_2 \\ & X \in \{\Pi A_1 \Pi^T \ : \ \Pi \text{ is an } n \times n \text{ permutation matrix.}\} \\ & E_{ij} \ \in \ \{-1, 0, 1\} \ \forall \, i, j \in \{1, \ldots, n\}. \end{aligned} \tag{3.1}$$

The function $1_{\{.\}}$ denotes the usual indicator function, the decision variable X is an adjacency matrix representing $\mathcal{G}_1$, and the decision variable E specifies the edge deletions and insertions made to $\mathcal{G}_1$ to obtain $\mathcal{G}_2$. One aspect of the problem (3.1) is that its formulation is not symmetric with respect to $\mathcal{G}_1$ and $\mathcal{G}_2$, although the optimal value remains unchanged if A_1 and A_2 are swapped in the problem description, i.e., $\mathrm{GED}(\mathcal{G}_1, \mathcal{G}_2) = \mathrm{GED}(\mathcal{G}_2, \mathcal{G}_1)$; we revisit this point in the sequel. Unsurprisingly, solving (3.1) is intractable in general, as calculating the graph edit distance is an NP-hard problem. Our approach in this chapter is to obtain tractable convex relaxations of (3.1). Relaxing the objective to a convex function is a straightforward matter; specifically, as the absolute value function constitutes a lower bound on the indicator function in the range $[-1, 1]$; we replace the objective of (3.1) with the convex function $\frac{1}{2}\|E\|_{\ell_1}$, where $\|\cdot\|_{\ell_1}$ denotes the (entrywise) ℓ_1 norm that sums the absolute values of the entries of a matrix.

The main remaining source of difficulty with obtaining a convex relaxation of (3.1) is to identify a tractable convex approximation of a set of the form $\{\Pi A \Pi^T \ : \ \Pi \text{ is an } n \times n \text{ permutation matrix}\}$ for a given matrix $A \in \mathbb{S}^n$. When

A specifies an adjacency matrix of a graph, this set consists of all the adjacency matrices that describe the graph, thus highlighting the *structural* attributes of the graph that remain invariant to vertex relabeling. Consequently, we seek a convex approximation that similarly remains invariant to vertex relabeling. We describe next a notion from [30] that aims to address this challenge:

Definition 23. *[30] A set $\mathcal{C} \subset \mathbb{S}^n$ is an* invariant convex set *if it is convex and if $M \in \mathcal{C}$ implies that $\Pi M \Pi^T \in \mathcal{C}$ for all $n \times n$ permutation matrices Π.*

Invariant convex sets provide a useful convex modeling framework to constrain graph properties that remain invariant to vertex relabeling. In particular, suppose that $\mathcal{C}_{\mathcal{G}_1} \subset \mathbb{S}^n$ is an invariant convex set that contains $\{\Pi A_1 \Pi^T : \Pi \text{ is an } n \times n \text{ permutation matrix}\}$ and has an efficient description. Then, the following *convex program* provides a lower bound on $\mathrm{GED}(\mathcal{G}_1, \mathcal{G}_2)$:

$$\begin{aligned} \mathrm{GED}_{LB}(\mathcal{G}_1, \mathcal{G}_2; \mathcal{C}_{\mathcal{G}_1}) = \min_{X, E \in \mathbb{S}^n} \ & \tfrac{1}{2} \|E\|_1 \\ s.t. \quad & X + E = A_2 \\ & X \in \mathcal{C}_{\mathcal{G}_1} \end{aligned} \tag{P}$$

It is evident that this problem provides a lower bound on $\mathrm{GED}(\mathcal{G}_1, \mathcal{G}_2)$ as the objective function here is a lower bound of the objective of (3.1) over the constraint set of (3.1), and further the constraint set of (P) is an outer approximation of the constraint set of (3.1). Unlike the optimal value $\mathrm{GED}(\mathcal{G}_1, \mathcal{G}_2)$ of (3.1), the optimal value $\mathrm{GED}_{LB}(\mathcal{G}_1, \mathcal{G}_2; \mathcal{C}_{\mathcal{G}_1})$ of (P) is not symmetric; specifically, if $\mathcal{C}_{\mathcal{G}_2}$ is some invariant convex set containing $\{\Pi A_2 \Pi^T : \Pi \text{ is an } n \times n \text{ permutation matrix}\}$, then in general $\mathrm{GED}_{LB}(\mathcal{G}_1, \mathcal{G}_2; \mathcal{C}_{\mathcal{G}_1}) \neq \mathrm{GED}_{LB}(\mathcal{G}_2, \mathcal{G}_1; \mathcal{C}_{\mathcal{G}_2})$. Therefore, in practice we propose computing both $\mathrm{GED}_{LB}(\mathcal{G}_1, \mathcal{G}_2; \mathcal{C}_{\mathcal{G}_1})$ and $\mathrm{GED}_{LB}(\mathcal{G}_2, \mathcal{G}_1; \mathcal{C}_{\mathcal{G}_2})$ for some invariant convex sets $\mathcal{C}_{\mathcal{G}_1}$ and $\mathcal{C}_{\mathcal{G}_2}$ corresponding to $\mathcal{G}_1$ and $\mathcal{G}_2$ respectively, and taking the larger of these quantities as both constitute lower bounds on $\mathrm{GED}(\mathcal{G}_1, \mathcal{G}_2)$.

This discussion leads naturally to the following question – which invariant convex set $\mathcal{C}_{\mathcal{G}}$ best captures the structural properties of a graph $\mathcal{G}$? Employing such a set in the relaxation (P) would provide better lower bounds on the edit distance. Letting $A \in \mathbb{S}^n$ be an adjacency matrix of $\mathcal{G}$, the 'tightest' invariant convex set that contains the collection $\{\Pi A \Pi^T : \Pi \text{ is an } n \times n \text{ permutation matrix}\}$ is simply the convex hull of this collection. However, this convex hull is intractable to describe for general graphs $\mathcal{G}$ (unless P=NP).

As a result, it is of interest to obtain computationally tractable invariant convex relaxations that reflect the structure in $\mathcal{G}$. In the next subsection, we give a list of invariant convex sets that are tractable to compute and that can be 'tuned' to the structure of $\mathcal{G}$. These invariants can either be used individually or combined (at increased computational expense), thus yielding a flexible and powerful framework for obtaining bounds on the graph edit distance. The focus of the rest of the chapter is on investigating the utility of these invariants theoretically as well as via numerical experiments; we give a summary of our main contributions in Section 3.1.

Convexity and Graph Invariants

We list here a few examples of invariant convex sets that play a prominent role in this chapter; we refer the interested reader to [30] for a more exhaustive list as well as additional properties of invariant convex sets.

Loopless and edge weight constraints Looplessness and edge weight bounds are not especially powerful constraints, but they nonetheless serve as simple examples of invariant convex sets. Looplessness corresponds to the constraint set $\{M \in \mathbb{S}^n \mid M_{ii} = 0 \text{ for } i = 1, \ldots, n\}$, and bounds on the edge weights for unweighted graphs (for example) can be specified via the set $\{M \in \mathbb{S}^n \mid 0 \leq M_{ij} \leq 1 \text{ for } i, j = 1, \ldots, n\}$.

Spectral invariants Let $\mathcal{G}$ be a graph represented by an adjacency matrix $A \in \mathbb{S}^n$ with eigenvalues $\lambda(A) \in \mathbb{R}^n$. The smallest convex set containing all graphs that are isospectral to $\mathcal{G}$ is given by the following *Schur-Horn orbitope* associated to A [107]:

$$\mathcal{C}_{\mathcal{SH}(\mathcal{G})} = \text{conv}\{M \in \mathbb{S}^n \mid \lambda(M) = \lambda(A)\}.$$

This set consists precisely of those matrices whose spectra are majorized by $\lambda(A)$. One can replace the list of eigenvalues in this example with the degree sequence of a graph, and in a similar vein, consider the convex hull of all adjacency matrices representing graphs with the same degree sequence; see [30] for more details.

A prominent way in which invariant convex sets can be constructed is via sublevel sets of convex graph invariants:

Definition 24. *[30] A function $f : \mathbb{S}^n \to \mathbb{R}$ is a* convex graph invariant *if it is convex and if $f(M) = f(\Pi M \Pi^T)$ for all $M \in \mathbb{S}^n$ and all $n \times n$ permutation*

matrices Π.

As with invariant convex sets, convex graph invariants characterize structural properties of a graph that are invariant to vertex relabeling. The following convex graph invariants play a prominent role in our work:

Inverse of the stability number A stable set (also known as independent set) of a graph $\mathcal{G}$ is a subset of vertices of $\mathcal{G}$ such that no two vertices in the subset are connected. The stability number of a graph $\mathcal{G}$ is a graph invariant that is equal to the size of the largest stable set of $\mathcal{G}$. It was shown by Motzkin and Straus [91] that the inverse of the stability number admits the following variational description, where $A \in \mathbb{S}^n$ is an adjacency matrix representing $\mathcal{G}$:

$$\begin{aligned} \text{inv-stab-number}(A) = \min_{x \in \mathbb{R}^n} \; & x'(I+A)x \\ s.t. \quad & \sum_i x_i = 1, \; x_i \geq 0 \text{ for } i = 1, \ldots, n. \end{aligned}$$

As the stability number of a graph is NP-hard to compute for general graphs (the above program may be reformulated as a conic program with respect to the completely positive cone), the following tractable relaxation based on doubly nonnegative matrices is widely employed:

$$\begin{aligned} f(A) = \min_{X \in \mathbb{S}^n} \; & \text{Tr}(X(I+A)) \\ s.t. \quad & X \succeq 0, \; \mathbf{1}'X\mathbf{1} = 1, X_{ij} \geq 0 \text{ for } i, j = 1, \ldots, n. \end{aligned} \tag{3.2}$$

One can check that both inv-stab-number(A) and $f(A)$ are concave graph invariants.

Maximum cut The maximum cut value of a graph is the maximum over all partitions of the vertices of the sum of the weights of the edges between the partitions. For a graph $\mathcal{G}$ specified by adjacency matrix $A \in \mathbb{S}^n$, the maximum cut value is given as:

$$\text{max-cut}(A) = \max_{y \in \{-1,1\}^n} \tfrac{1}{4} \sum_{i,j} A_{i,j}(1 - y_i y_j).$$

As this value is NP-hard to compute for general graphs, the following celebrated efficiently-computable relaxation is commonly used [59]:

$$\begin{aligned} g(A) = \max_{X \in \mathbb{S}^n} \; & \tfrac{1}{4} \text{Tr}(A\,(\mathbf{1}\mathbf{1}^T - X)) \\ s.t. \quad & X \succeq 0, \; X_{ii} = 1 \text{ for } i = 1, \ldots, n. \end{aligned} \tag{3.3}$$

Both max-cut(A) and $g(A)$ are convex graph invariants as they are each invariant under conjugation of the argument by a permutation matrix and they are each expressed as a pointwise maximum of affine functions.

Our Contributions

The invariant convex sets listed in the previous section when used in the context of the optimization problem (P) all lead to valid lower bounds on the edit distance between two graphs. The question then is whether certain invariants are more naturally suited to particular structural properties of graphs. The main focus of this chapter is on identifying attributes of graphs for which the invariants described above are well-suited, and evaluating in these contexts the quality of the bounds obtained via (P) both theoretically and through numerical experiments. Specifically, we say that an invariant convex constraint set $\mathcal{C}_{\mathcal{G}}$ is well-suited to the structure of a graph $\mathcal{G}$ if $\mathrm{GED}_{LB}(\mathcal{G}, \mathcal{G}'; \mathcal{C}_{\mathcal{G}})$ provides a tight (or high-quality) lower bound of $\mathrm{GED}(\mathcal{G}, \mathcal{G}')$ for all graphs $\mathcal{G}'$ that are obtained via a small number of edge deletions and insertions applied to $\mathcal{G}$ (here 'small' is interpreted relative to the total number of edges in $\mathcal{G}$).

In Section 3.2, we investigate theoretically the effectiveness of the Schur-Horn orbitope as an invariant convex set in providing lower bounds on the graph edit distance via (P). We consider a stylized setting in which a graph $\mathcal{G}$ on n vertices is modified to a graph $\mathcal{G}'$ by adding or removing at most d edges incident to each vertex of $\mathcal{G}$. We prove in Theorem 2 (see Section 3.2) that the optimal value of the convex program (P) with a Schur-Horn orbitope constraint set equals the graph edit distance between $\mathcal{G}$ and $\mathcal{G}'$, i.e., $\mathrm{GED}_{LB}(\mathcal{G}, \mathcal{G}'; \mathcal{C}_{SH(\mathcal{G})}) = \mathrm{GED}(\mathcal{G}, \mathcal{G}')$ provided: 1) d is sufficiently small, 2) $\mathcal{G}$ has eigenspaces with the property that there exists a linear combination of the associated projection operators onto these spaces so that the largest entry in magnitude is suitably bounded, and 3) any matrix supported on entries corresponding to edits of $\mathcal{G}$ has only a small amount of its energy on each of the eigenspaces of $\mathcal{G}$; see Theorem 2 for precise details. Conditions similar to the third requirement appear in the authors' earlier work on employing the Schur-Horn orbitope in the context of the planted subgraph problem [26]. However, the second condition is novel and is motivated by the context of the present chapter on graph edit distance. Under the additional assumption that $\mathcal{G}$ is vertex-transitive, Corollary 31 provides a simple formula on the maximum allowable number d of edge additions/deletions per vertex of $\mathcal{G}$; we illustrate the

utility of this formula by computing bounds on d for many graph families such as Johnson graphs, Kneser graphs, Hamming graphs and other strongly regular graphs. Indeed, for some of these families, our results are 'order-optimal' in the sense that our bounds on d are on the order of the degree of $\mathcal{G}$. The proofs of the main results of Section 3.2 are given in Section 3.3.

In Section 3.4, we conduct a detailed numerical evaluation of the power and limitations of convex invariant relaxations based on the inverse stability number (via the tractable approximation (3.2)) and the maximum cut value (via the tractable approximation (3.3)). We do not provide precise theoretical guarantees due to a lack of a detailed characterization of the facial structure of the associated convex sets. Nonetheless, we identify classes of graph edit distance problems for which these constraints produce high-quality lower bounds. Specifically, we observe that a convex relaxation based on the Motzkin-Straus approximation of the inverse of the stability number provides useful lower bounds on graph edit distance if one of the graphs has the property that the removal of any edge increases the graph's stability number; graphs with such a property have been studied in the extremal graph theory literature [70] and we refer the reader to Section 3.4 for further details. Similarly, in Section 3.4, we observe that a convex relaxation based on the Goemans-Williamson approximation of the maximum cut value produces effective lower bounds on the graph edit distance if the addition of any edge to one of the graphs increases that graph's maximum cut value; windmill graphs are a prominent example that possess such a property. In both Sections 3.4 and 3.4, we present empirical results that corroborate our observations.

In Section 3.5 we demonstrate the utility of our framework in providing lower bounds on the average pairwise graph edit distance in two chemistry datasets consisting of a collection of molecules known as Alkanes and Polycyclic Aromatic Hydrocarbons (PAH). The PAH dataset in particular consists of large structures for which exact computation of graph edit distance is prohibitively expensive. The best-known upper bound on the average graph edit distance over all pairs of graphs in this dataset is 29.8, and to the best of our knowledge, the exact value of this quantity is not known [17]. Indeed, much of the literature featuring the PAH dataset aims at providing an upper bound on the average pairwise graph edit distance. Our framework provides a lower bound of 21.6 on the average pairwise graph edit distance of PAH, which appears to

be the best available bound to date. In obtaining these results, we combine invariant convex sets based on the Schur-Horn orbitope, the Motzkin-Straus approximation of the inverse stability number, and the Goemans-Williamson approximation of the maximum-cut value.

Notation We denote the normal cone at a point $x \in \mathcal{C}$ of a closed convex set $\mathcal{C}$ by $\mathcal{N}_{\mathcal{C}}(x)$. The projection map onto a subspace $\mathcal{E} \subset \mathbb{R}^n$ is denoted by $P_{\mathcal{E}} : \mathbb{R}^n \to \mathbb{R}^n$. For a collection of subspaces $\mathcal{E}_i \subset \mathbb{R}^n,\ i \in \{1, \ldots, m\}$, the operator $\mathcal{P}_{ij} : \mathbb{S}^n \to \mathbb{S}^n$ is defined as $\mathcal{P}_{ij} := P_{\mathcal{E}_j} \otimes P_{\mathcal{E}_i}$, i.e., $\mathcal{P}_{ij}(A) = P_{\mathcal{E}_i} A P_{\mathcal{E}_j}$. The restriction of a (usually self-adjoint) linear map $f : \mathbb{R}^n \to \mathbb{R}^n$ to an invariant subspace $\mathcal{E}$ of f is denoted by $f|_{\mathcal{E}} : \mathcal{E} \to \mathcal{E}$.

3.2 Theoretical Guarantees for the Schur-Horn Orbitope Constraint

In this section, we give theoretical guarantees that describe conditions under which employing the Schur-Horn orbitope as an invariant convex constraint set in (P) leads to the associated lower bound on the graph edit distance being *tight*, i.e., the optimal value of (P) equals the graph edit distance. Concretely, we consider conditions on a graph $\mathcal{G}$ and the structure of the edits that transform $\mathcal{G}$ to another graph $\mathcal{G}'$ so that $\mathrm{GED}(\mathcal{G}, \mathcal{G}') = \mathrm{GED}_{LB}(\mathcal{G}, \mathcal{G}'; \mathcal{C}_{\mathcal{SH}(\mathcal{G})})$. We begin with a description of our main theoretical results in Section 3.2, some consequences of these results for specific graph families in Section 3.2, and finally an experimental demonstration on the utility of the Schur-Horn orbitope on stylized problems in Section 3.2. The proofs of the results of this section are deferred to Section 3.3.

As the normal cones at extreme points of the Schur-Horn orbitope play a prominent role in the optimality conditions of (P), we state the relevant result here:

Lemma 25. *[26] Let $\mathcal{G}$ be any unweighted graph with m eigenvalues. A matrix W is an extreme point of $\mathcal{SH}(\mathcal{G})$ if and only if it has the same eigenvalues as $\mathcal{G}$ (counting multiplicity). Further, the relative interior of the normal cone relint($\mathcal{N}_{\mathcal{SH}(\mathcal{G})}(W)$) at an extreme point W consists of those matrices Q that satisfy the following conditions:*

1. *Q is simultaneously diagonalizable with W,*

2. $\lambda_{\min}(Q|_{\mathcal{E}_i}) > \lambda_{\max}(Q|_{\mathcal{E}_{i+1}})\ \forall\, i \in \{1, \ldots, m-1\}$,

where $\mathcal{E}_i$ for $i \in \{1, \ldots, m\}$ are eigenspaces of W ordered such that the corresponding eigenvalues are sorted in a decreasing order.

From Lemma 25 we observe that the relative interior of the normal cones of the Schur-Horn orbitope at extreme points are 'larger' if the underlying graph $\mathcal{G}$ consists of few distinct eigenvalues. This observation along with various properties of the eigenspaces of $\mathcal{G}$ play a prominent role in the analysis in this section.

Main Results

We present here the statements of our main theoretical results concerning the performance of the Schur-Horn orbitope as a constraint set in (P). In addition to various structural properties of $\mathcal{G}$, our results are described in terms of a parameter d that denotes the maximum number of deletions/additions of edges that are incident to any vertex of $\mathcal{G}$. Informally, we should expect the Schur-Horn orbitope constraint to be effective in exactly computing the graph edit distance if a matrix representing the edits from $\mathcal{G}$ to $\mathcal{G}'$ has only a small amount of its energy on each of the eigenspaces of $\mathcal{G}$. The reason behind this observation is that if the edits were largely concentrated in the eigenspaces of $\mathcal{G}$, then the eigenspaces of $\mathcal{G}'$ would be close to those of $\mathcal{G}$. This would result in an identifiability problem from the perspective of the Schur-Horn orbitope, which is based purely on the spectral properties of $\mathcal{G}$. To formalize this notion, we present the following definition which plays a key role in our analysis:

Definition 26. *Let $\mathcal{G}$ be a graph on n vertices with m distinct eigenvalues. Let $P_i,\ i = 1, \ldots, m$ represent projection maps onto the eigenspaces of $\mathcal{G}$ indexed by decreasing order of the corresponding eigenvalues and let $\mathcal{P}_{ii} = P_i \otimes P_i$. Fix a positive integer d and $\alpha \in [0,1]^m$. Define the parameter $\xi(\alpha, d, \mathcal{G})$ to be the smallest value*

$$\left\| [I - \sum_{i=1}^{m} \alpha_i \mathcal{P}_{ii}](W) \right\|_{\infty} \leq \xi(\alpha, d, \mathcal{G}) \left\| W \right\|_{\infty}$$

for all $W \in \mathbb{S}^n$ with at most d nonzero entries per row/column.

Remark 27. *The maps P_i represent projections onto eigenspaces of an adjacency matrix representing $\mathcal{G}$, but we simply refer to these as eigenspaces of $\mathcal{G}$*

with an abuse of terminology. The reason is that the quantity $\xi(\alpha, d, \mathcal{G})$ is a graph parameter (for each fixed α, d) and does not depend on a specific labeling of the vertices of $\mathcal{G}$.

Remark 28. *The parameter $\xi(\alpha, d, \mathcal{G})$ is a restricted version of the induced (entrywise) infinity norm $\|I - \sum_i^m \alpha_i \mathcal{P}_{ii}\|_{\infty\to\infty}$, with the key difference being that $\xi(\alpha, d, \mathcal{G})$ computes the induced gain of the operator $I - \sum_i^m \alpha_i \mathcal{P}_{ii}$ restricted to inputs that have at most d nonzeros per row/column.*

The quantity $\xi(\alpha, d, \mathcal{G})$ helps quantify the idea described previously about the energy of the edits not being confined excessively to the eigenspaces of $\mathcal{G}$. As the specific edit pattern is not known in advance, this quantity is agnostic to the particular edits and is parametrized only in terms of the maximum number of edge deletions/additions that are incident to any vertex. In our main results described next, larger values of ξ make it harder to satisfy our sufficient conditions on tightness of our lower bounds. As the value of ξ depends on the selection of the parameter α, our main results allow for flexibility in the choice of this parameter, and we describe in Section 3.2 how specific choices lead to concrete consequences on the exactness of the relaxation (P) with the Schur-Horn orbitope constraint for various graph families. We present next a result that establishes basic optimality conditions of the convex program (P):

Lemma 29. *Let $\mathcal{G}$ be a graph on n vertices with m distinct eigenvalues, and let $\mathcal{G}'$ be a graph that is obtained from $\mathcal{G}$ via edge deletions/additions such that each vertex is incident to at most d edits. Let $A, A + E^* \in \mathbb{S}^n$ represent the graphs $\mathcal{G}$ and $\mathcal{G}'$, respectively; that is, E^* consists of at most d nonzeros per row/column. Let $\Omega \subset \mathbb{S}^n$ denote the subspace consisting of all matrices with nonzeros contained within the support of E^*. Suppose a vector $\alpha \in [0, 1]^m$ and a matrix $Q \in \mathbb{S}^n$ satisfy the following conditions:*

1. $\mathcal{P}_\Omega(Q) = \operatorname{sign}(E^*)$,
2. $||\mathcal{P}_{\Omega^c}(Q)||_\infty < 1$,
3. $Q \in relint(\mathcal{N}_{\mathcal{SH}(\mathcal{G})}(A))$,
4. $\xi(\alpha, d, \mathcal{G}) < 1$.

Then we have that the convex relaxation (P) with the Schur-Horn orbitope constraint exactly computes the edit distance between $\mathcal{G}$ and $\mathcal{G}'$, i.e., $GED(\mathcal{G}, \mathcal{G}') = GED_{LB}(\mathcal{G}, \mathcal{G}'; \mathcal{C}_{\mathcal{SH}(\mathcal{G})})$, with the optimal solution being unique and achieved at a matrix that specifies an optimal set of edits.

Proof. The proof is presented in Appendix B.1. □

Conditions 1, 2, and 3 of this lemma essentially require that the subdifferential at a matrix specifying the edits with respect to the ℓ_1 norm has a nonempty intersection with the relative interior of the normal cone at an adjacency matrix representing $\mathcal{G}$ with respect to the Schur-Horn orbitope. In papers on the topic of low-rank matrix completion and matrix decomposition [24, 29], a convenient approach to ensuring that such types of conic intersection conditions can be satisfied is based on requiring that nullspace (the eigenspace corresponding to an eigenvalue of zero) of the low-rank matrix is suitably 'incoherent', i.e., that there are no elements of this nullspace with energy concentrated in a single location. In our context, all of the eigenspaces of $\mathcal{G}$ play a role rather than just a single distinguished eigenspace, and accordingly we describe next a weighted form of an incoherence-type condition:

Definition 30. *Let $\mathcal{G}$ be a graph on n nodes with m distinct eigenvalues and let $P_1, \ldots, P_m \in \mathbb{S}^n$ denote the projection matrices onto the associated eigenspaces indexed by decreasing order of the corresponding eigenvalues. Fix any $\gamma \in \mathbb{R}^m$. We define the parameter $\rho(\gamma, \mathcal{G})$ as follows:*

$$\rho(\gamma, \mathcal{G}) := \left\| \sum_{i=1}^{m} \gamma_i P_i \right\|_{\infty}.$$

Here the matrix $\|\cdot\|_\infty$ norm is the largest entry of the argument in magnitude. In the literature on inverse problems involving low-rank matrices, one typically considers the infinity norm of the projection map onto the nullspace as well as variants of this quantity. Thus, in this sense the parameter $\rho(\gamma, \mathcal{G})$ is a weighted generalization that is more suited to our setup. We state next our main theorem in terms of sufficient conditions involving the two parameters we have introduced in this section:

Theorem 2. *Let $\mathcal{G}$ be a graph on n vertices with m distinct eigenvalues, and let $\mathcal{G}'$ be a graph that is obtained from $\mathcal{G}$ via edge deletions/additions such that*

each vertex is incident to at most d edits. Suppose the following two conditions are satisfied for some $\gamma \in \mathbb{R}^m$ and $\alpha \in [0,1]^m$:

1. $2\,\xi(\alpha, d, \mathcal{G}) + \rho(\gamma, \mathcal{G}) < 1$,
2. $\frac{(\alpha_i + \alpha_{i+1})\,(1+\rho(\gamma,\mathcal{G}))\,d}{1-\xi(\alpha,d,\mathcal{G})} < \gamma_{i+1} - \gamma_i,\ \forall\, i \in \{1, \ldots, m-1\}$.

Then the convex relaxation (P) with the Schur-Horn orbitope constraint exactly computes the edit distance between $\mathcal{G}$ and $\mathcal{G}'$, i.e., $GED(\mathcal{G}, \mathcal{G}') = GED_{LB}(\mathcal{G}, \mathcal{G}'; \mathcal{C}_{\mathcal{SH}(\mathcal{G})})$, with the optimal solution being unique and achieved at a matrix that specifies an optimal set of edits.

This theorem states that the relaxation (P) with the Schur-Horn orbitope constraint set succeeds in calculating the graph edit distance exactly if 1) d is small enough, 2) there exists a vector α with small entries such that $\xi(\alpha, d, \mathcal{G})$ is also suitably small, and 3) there exists an ordered vector γ with well-separated entries that yields a small value of $\rho(\gamma, \mathcal{G})$. As discussed in the next subsection, graphs with a small number of well-separated eigenvalues offer an ideal candidate. Specifically, for such graph families, we give concrete consequences in terms of bounds on the maximum number d of edits per vertex via particular choices for α and γ in Theorem 2.

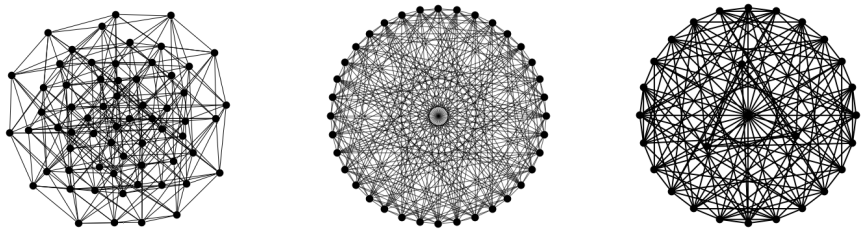
Consequences for Graph Families with Few Eigenvalues

Theorem 2 constitutes our most general result on the tightness of the Schur-Horn orbitope constraint in computing the graph edit distance when employed as a constraint set in the context of (P). The generality of the result stems from the wide range of flexibility provided by the vectors γ and α. In Corollary 31, we consider specific choices of these parameters to obtain concrete bounds in terms of graph parameters that can be computed easily:

Corollary 31. *Let $\mathcal{G}$ be a vertex-transitive graph on n vertices consisting of m distinct eigenvalues, and let κ denote the multiplicity of the eigenvalue with the second-highest multiplicity. Suppose $\mathcal{G}'$ is a graph on n vertices that can be obtained from $\mathcal{G}$ with the addition or removal of at most d edges incident to each vertex of $\mathcal{G}$. Then there exists a constant c depending only on m so that the optimal value $GED_{LB}(\mathcal{G}, \mathcal{G}'; \mathcal{C}_{\mathcal{SH}(\mathcal{G})})$ of (P) equals $GED(\mathcal{G}, \mathcal{G}')$ provided*

$$d \leq c\frac{n}{\kappa}.$$

The particular dependence on the multiplicity of the eigenvalue with second-largest multiplicity is due to the choices of α and γ in Theorem 2 that we have employed in our proof; see Section 3.3 for more details. In the sequel we give consequences of this result for specific graph families in which the number of distinct eigenvalues is small (for example, three or four). In the context of such graphs, the relaxation (P) is tight even when the number of edits per vertex is large so long as the value of κ is suitably small. Indeed, for several graph families we observe that Corollary 31 produces 'order-optimal' bounds as the largest value of d that is allowed is on the same order as the degree of the underlying graphs.



(a) (b) (c)

Figure 3.2: From left to right: Hamming graph H(3,4), 9-Triangular graph, generalized quadrangle-(2,4) graph.

Johnson Graphs A Johnson graph $J(k, \ell)$ with $\ell > 0$ is a graph on $n = \binom{k}{\ell}$ vertices that correspond to the ℓ-element subsets of a set of k elements. Two vertices of a Johnson graph are connected if the corresponding subsets of these vertices contain $\ell - 1$ common elements. The Johnson graph $J(k, \ell)$ is vertex-transitive and contains $\ell+1$ distinct eigenvalues. For $k \geq 2\ell$ and $j \in \{0, \ldots, \ell\}$, the multiplicity of its j'th eigenvalue is $\binom{k}{j}$ -$\binom{k}{j-1}$ for $j > 0$ and one for $j = 0$. For small values of ℓ, the multiplicity of the second most repeated eigenvalue is about $k^{\ell-1}$. As a result, for small fixed values of ℓ, Corollary 31 states that the convex relaxation (P) is tight provided

$$d \lesssim n^{\frac{1}{\ell}}.$$

Kneser Graphs A Kneser graph $K(k, \ell)$ with $\ell > 0$ shares certain aspects with Johnson graphs. Specifically, the vertices of $K(k, \ell)$ coincide with the ℓ-element subsets of a set of k elements, as with Johnson graphs. However, two vertices of a Kneser graph are connected if the subsets corresponding to these vertices are disjoint. Kneser graphs are vertex-transitive, and their eigenvalues

exhibit the same multiplicities as those of the Johnson graphs $J(k, \ell)$. As a result, for small fixed values of ℓ, Corollary 31 implies that the relaxation (P) is tight provided:

$$d \lesssim n^{\frac{1}{\ell}}.$$

Hamming Graphs A Hamming graph $H(\ell, q)$ consists of q^ℓ vertices (see Figure 3.2a for a depiction of $H(3, 4)$). Each vertex of $H(\ell, q)$ corresponds to a sequence of length ℓ from a set with q distinct elements. Two vertices are connected if their associated sequences differ in exactly one coordinate, i.e., their Hamming distance is equal to 1. Hamming graphs are vertex-transitive, and the spectrum of $H(\ell, q)$ consists of $\ell + 1$ distinct eigenvalues with multiplicities $\binom{\ell}{i}(q-1)^i$, $i \in \{0, \ldots, \ell\}$. Therefore, for a small fixed value of ℓ, Corollary 31 states that the relaxation (P) is tight provided:

$$d \lesssim n^{\frac{1}{\ell}}.$$

Vertex-Transitive Strongly Regular Graphs A strongly regular graph on n vertices with degree r is defined by the property that every pair of adjacent vertices has d_a common neighbors and every pair of nonadjacent vertices has d_{na} common neighbors. Such graphs are generally denoted $srg(n, r, d_a, d_{na})$. Due to their rich algebraic structure, strongly regular graphs have only three distinct eigenvalues with multiplicities equal to one and $\frac{1}{2}\left[(n-1) \pm \frac{2r+(n-1)(d_a-d_{na})}{\sqrt{(d_a-d_{na})^2+4(r-d_{na})}}\right]$. Furthermore, many strongly regular graphs are also vertex-transitive and as a result, our Corollary 31 is applicable. We highlight two prominent examples:

- A k-Triangular graph T_k on $n = \binom{k}{2}$ vertices is a vertex-transitive strongly regular graph with parameters $srg(k(k-1)/2, 2(k-2), k-2, 4)$ (in fact, T_k is also isomorphic to the Johnson graph $J(k, 2)$); see Figure 3.2b for the 9-Triangular graph. Corollary 31 states that the convex relaxation (P) is tight provided:

  $$d \lesssim n^{\frac{1}{2}}.$$

  Incidentally, the degree of T_k also scales as $n^{\frac{1}{2}}$; as a result, Corollary 31 is tight for this family up to constant factors.

- A generalized quadrangle is an incidence relation satisfying certain geometric axioms on points and lines. A generalized quadrangle of order (s,t) gives rise to a strongly regular graph with parameters $srg((s+1)(st+1), s(t+1), s-1, t+1)$ denoted by $GQ(s,t)$ on $n=(s+1)(st+1)$ vertices – see Figure 3.2c for an illustration of the vertex-transitive graph $GQ(2,4)$. Considering generalized quadrangle graphs $GQ(s,s^2)$ when they are vertex-transitive, Corollary 31 implies that the relaxation (P) is tight provided

$$d \lesssim n^{\frac{1}{4}}.$$

In Section 3.2 we demonstrate the utility of our framework via numerical experiments on edit distance problems involving the graphs T_9 and $GQ(2,4)$.

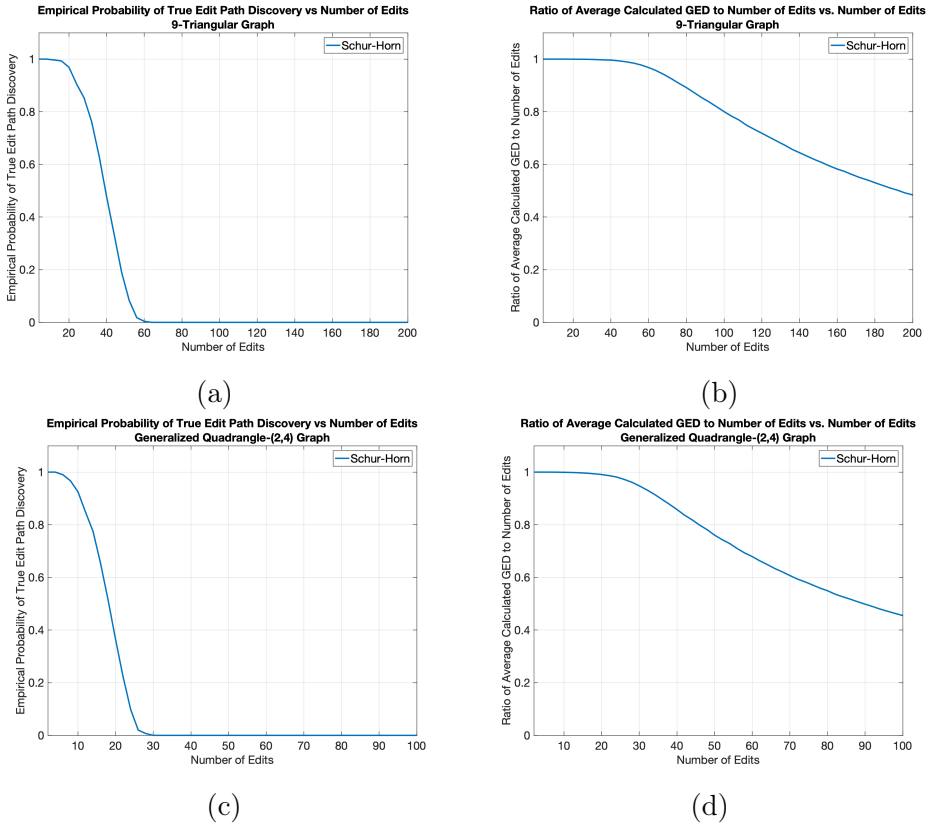


Figure 3.3: Performance of our framework (P) with the Schur-Horn constraint. Left: empirical probability of discovering the true graph edit path. Right: ratio of average calculated graph edit distance to number of edit operations. The first row corresponds to the 9-triangular graph and the second row to the generalized quadrangle-(2,4) graph.

Numerical Experiments

We demonstrate the utility of the Schur-Horn orbitope as a constraint set in (P) in obtaining bounds on the graph edit distance between graphs $\mathcal{G}$ and $\mathcal{G}'$. In our experiments, we fix $\mathcal{G}$ to be either the 9-triangular graph T_9 (Figure 3.2b) or the generalized quadrangle-(2,4) graph $GQ(2,4)$ (Figure 3.2c) introduced previously. The graph T_9 consists of 36 vertices and 252 edges and the graph GQ(2,4) consists of 27 vertices and 135 edges. Both of these are strongly regular graphs. In each case, the corresponding graph $\mathcal{G}'$ is obtained by adding/deleting edges randomly (both addition and deletion occur with equal probability) to achieve a desired number of edits. When $\mathcal{G}$ is T_9 we vary the number of edits from four to 200 in increments of four, and when $\mathcal{G}$ is $GQ(2,4)$ we vary the number of edits from two to 100 in increments of two. For each number of edits, we consider 1000 random trials and we report the probability that $\mathrm{GED}(\mathcal{G},\mathcal{G}') = \mathrm{GED}_{LB}(\mathcal{G},\mathcal{G}';\mathcal{C}_{\mathcal{SH}(\mathcal{G})})$ and the ratio of the average computed lower bound $\mathrm{GED}_{LB}(\mathcal{G},\mathcal{G}';\mathcal{C}_{\mathcal{SH}(\mathcal{G})})$ to the number of edits. In particular, we declare that $\mathrm{GED}(\mathcal{G},\mathcal{G}') = \mathrm{GED}_{LB}(\mathcal{G},\mathcal{G}';\mathcal{C}_{\mathcal{SH}(\mathcal{G})})$ if the infinity norm (maximum entrywise magnitude) of the difference between the optimal solution $\hat{E}$ and the true edit matrix E^* is less than 0.01. The results are shown in Figure 3.3 and they were obtained using the CVX parser [60, 61] and the SDPT3 solver [112]. As these plots demonstrate, the convex relaxation (P) with a Schur-Horn orbitope as an invariant convex constraint set is tight when the number of edits is small and leads to effective lower bounds when the number of edits is large.

3.3 Proofs of Results from Section 3.2

Constructing a Dual Certificate

We describe here a method for constructing a suitable dual certificate satisfying the conditions of Lemma 29, and we prove that this construction is valid whenever certain conditions involving the parameters ξ and ρ from Section 3.2 are satisfied. Our proofs are presented in the context of two intermediary lemmas, which are then used to prove Theorem 2. Specifically, our approach to constructing $Q \in \mathbb{S}^n$ that satisfies the requirements of Lemma 29 is to express Q as follows:

$$Q = R + \Delta.$$

Here $R \in \mathbb{S}^n$ plays the role of a 'reference' matrix that depends purely on the underlying graph $\mathcal{G}$, while $\Delta \in \mathbb{S}^n$ is a perturbation that additionally depends on the specific edits that transform $\mathcal{G}$ to $\mathcal{G}'$. We begin by stating an easily-proved result that serves as the basis for our subsequent development:

Lemma 32. *Let $\mathcal{G}$ be a graph on n vertices with m distinct eigenvalues, and let $\mathcal{G}'$ be a graph that is obtained from $\mathcal{G}$ via edge deletions/additions such that each vertex is incident to at most d edits. Let $A, A + E^* \in \mathbb{S}^n$ represent the graphs $\mathcal{G}$ and $\mathcal{G}'$, respectively; that is, E^* consists of at most d nonzeros per row/column. Let $\Omega \subset \mathbb{S}^n$ denote the subspace consisting of all matrices with nonzeros contained within the support of E^*. Let $P_i \in \mathbb{S}^n,\ i = 1, \ldots, m$ denote projection maps onto the eigenspaces of A indexed by decreasing order of the corresponding eigenvalues. Suppose a vector $\alpha \in [0,1]^m$, a vector $\gamma \in \mathbb{R}^m$, and a matrix $\Delta \in \mathbb{S}^n$ satisfy the following conditions with $R = \sum_i \gamma_i P_i$:*

1. $\mathcal{P}_\Omega(\Delta) + \mathcal{P}_\Omega(R) = \text{sign}(E^*)$,
2. $||\mathcal{P}_{\Omega^c}(\Delta)||_\infty + ||\mathcal{P}_{\Omega^c}(R)||_\infty < 1$,
3. $\mathcal{P}_{ij}(\Delta) = 0,\ \forall i, j \in \{1, \ldots, m\},\ i \neq j$,
4. $||\mathcal{P}_{ii}(\Delta)||_2 + ||\mathcal{P}_{i+1,i+1}(\Delta)||_2 < \gamma_{i+1} - \gamma_i,\ \forall i \in \{1, \ldots, m-1\}$,
5. $\xi(\alpha, d, \mathcal{G}) < 1$.

Then the convex relaxation (P) with the Schur-Horn orbitope constraint exactly computes the edit distance between $\mathcal{G}$ and $\mathcal{G}'$, i.e., $GED(\mathcal{G}, \mathcal{G}') = GED_{LB}(\mathcal{G}, \mathcal{G}'; \mathcal{C}_{\mathcal{SH}(\mathcal{G})})$, with the optimal solution being unique and achieved at a matrix that specifies an optimal set of edits.

Proof. One can check that the conditions of Lemma 29 are satisfied by setting $Q = R + \Delta$. □

This lemma highlights the role of the parameter γ, in particular demonstrating that larger separation among the values of γ makes condition 4 easier to satisfy but may also increase the value of $||\mathcal{P}_{\Omega^c}(R)||_\infty$, thus making condition 2 potentially harder to satisfy.

We now move on to the perturbation term Δ. As this matrix must satisfy several of the constraints discussed in Lemma 32, its construction is somewhat

delicate. We build on the ideas developed in [32] in the context of low-rank matrix recovery, but with certain adaptations that are crucial to our setting. We construct Δ as an element in the range of an operator $\mathcal{L}^\alpha : \mathbb{S}^n \to \mathbb{S}^n$ that is parametrized by $\alpha \in [0,1]^m$:

$$\mathcal{L}^\alpha := \left(\sum_{i=1}^m \alpha_i \mathcal{P}_{ii}\right) \mathcal{P}_\Omega \left[I - \left(I - \sum_{i=1}^m \alpha_i \mathcal{P}_{ii}\right) \mathcal{P}_\Omega\right]^{-1}.$$

All of the operators here are as defined before. A point of the departure in the description of this operator relative to the ideas in [32] is that our version allows for 'fractional' contractions (as well as integral ones) based on the choice of α. When it is well-defined (i.e., the term involving the inverse is indeed invertible), the operator $\mathcal{L}^\alpha$ possesses a number of properties that lead to a convenient approach for constructing a suitable dual variable:

(P1) $\mathcal{P}_{ij}\mathcal{L}^\alpha = 0 \ \forall i, j \in \{1, \ldots, m\}$, $i \neq j$

(P2) $\mathcal{P}_\Omega \mathcal{L}^\alpha = \mathcal{P}_\Omega$.

In the context of Lemma 32, property (P1) ensures that that Δ is completely contained in a desired subspace, as stipulated by condition 3 of Lemma 32. Further, property (P2) implies that condition 1 of Lemma 32 is satisfied – in particular, we make use of this property to ensure that Δ takes on a desired value when restricted to Ω. Conditions 2 and 4 of Lemma 32 require that the quantities $\|\mathcal{P}_{ii}(\Delta)\|_2$ and $\|\mathcal{P}_{\Omega^c}(\Delta)\|_\infty$ to be sufficiently small – these conditions are satisfied by the operator $\mathcal{L}^\alpha$ as well, as documented next:

Lemma 33. *Consider the same setup as in Lemma 32. Fix any $\alpha \in [0,1]^m$ such that $\xi(\alpha, d, \mathcal{G}) < 1$. Then the operator $\mathcal{L}^\alpha : \mathbb{S}^n \to \mathbb{S}^n$ is well-defined (i.e., the term containing the inverse is indeed invertible) and the following inequalities hold:*

1. $\|[\mathcal{P}_{\Omega^c}\mathcal{L}^\alpha](X)\|_\infty \leq \frac{\xi(\alpha,d,\mathcal{G})\|X\|_\infty}{1-\xi(\alpha,d,\mathcal{G})}$,
2. $\|[\mathcal{P}_{ii}\mathcal{L}^\alpha](X)\|_2 \leq \frac{\alpha_i d\|X\|_\infty}{1-\xi(\alpha,d,\mathcal{G})}$.

In addition to providing upper bounds that serve as a foundation for the proof of our main theorem, Lemma 33 conveys the significance of the parameter $\xi(\alpha, d, \mathcal{G})$. Specifically, a suitably small value of $\xi(\alpha, d, \mathcal{G})$ guarantees that $\mathcal{L}^\alpha$

is well-defined, along with the conclusion that elements in the range of $\mathcal{L}^\alpha$ have small infinity norm (when restricted to Ω) and small operator norm (restricted to eigenspaces of $\mathcal{G}$). The lemma also suggests that the operator norm of the restriction of $\mathcal{L}^\alpha(X)$ to any eigenspace of $\mathcal{G}$ scales with the corresponding entry of α. Consequently, one can adjust α and γ to ensure that every inequality in Lemma 32 condition 4 is satisfied. Identifying the best values of α and γ to achieve this may be accomplished in special cases based on underlying structure in $\mathcal{G}$, as demonstrated by Corollary 31 and the many concrete consequences that are described in Section 3.2. In particular, in the proof of Corollary 31, we choose γ such that the separation between consecutive γ_i's is proportional to the sum of consecutive α_i's and we demonstrate that this approach yields easily-computable bounds based on properties of the underlying graph $\mathcal{G}$ on the highest number d of tolerable edits per vertex.

Proofs

Proof of Lemma 33

Our proof is analogous to that of [32, Lemma 8]. In order to avoid notational clutter, we denote $(I - \sum_{i=1}^m \alpha_i \mathcal{P}_{ii})$ as $\mathcal{P}_T$. Then from the definition of $\xi(\alpha, d, \mathcal{G})$, we have:

$$\|[\mathcal{P}_T \mathcal{P}_\Omega](X)\|_\infty \leq \xi(\alpha, d, \mathcal{G}) \|X\|_\infty . \tag{3.4}$$

Due to the assumption that $\xi(\alpha, d, \mathcal{G}) < 1$, we have that the series $I + \mathcal{P}_T\mathcal{P}_\Omega + \mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega + \dots$ converges geometrically with rate $\frac{1}{1-\xi(\alpha,d,\mathcal{G})}$ and equals $(I - \mathcal{P}_T\mathcal{P}_\Omega)^{-1}$. As such, the operator $\mathcal{L}^\alpha$ is well-defined.

Next, we proceed to the upper bounds. First we have that:

$$\begin{aligned}
\|[\mathcal{P}_{\Omega^c}\mathcal{L}^\alpha](X)\|_\infty &= \left\|[\mathcal{P}_{\Omega^c}\mathcal{P}_T\mathcal{P}_\Omega(I - \mathcal{P}_T\mathcal{P}_\Omega)^{-1}](X)\right\|_\infty \\
&\leq \left\|[\mathcal{P}_T\mathcal{P}_\Omega(I - \mathcal{P}_T\mathcal{P}_\Omega)^{-1}](X)\right\|_\infty \\
&\leq \xi(\alpha, d, \mathcal{G}) \left\|(I - \mathcal{P}_T\mathcal{P}_\Omega)^{-1}(X)\right\|_\infty \\
&\leq \frac{\xi(\alpha, d, \mathcal{G}) \|X\|_\infty}{1 - \xi(\alpha, d, \mathcal{G})}.
\end{aligned}$$

One can check that the first equality holds based on a term-by-term comparison. The first inequality follows by dropping the projection $\mathcal{P}_{\Omega^c}$. Bounding the resulting quantity from above using $\xi(\alpha, d, \mathcal{G})$ yields the second inequality. The last inequality follows from the geometric convergence of $(I - \mathcal{P}_T\mathcal{P}_\Omega)^{-1}$.

Next we bound the quantity involving the operator norm:

$$\begin{aligned}\left\|[\mathcal{P}_{ii}\mathcal{L}^{\alpha}](X)\right\|_2 &= \left\|[\alpha_i\mathcal{P}_{ii}\mathcal{P}_{\Omega}(I-\mathcal{P}_T\mathcal{P}_{\Omega})^{-1}](X)\right\|_2 \\ &\leq \alpha_i \left\|[\mathcal{P}_{\Omega}(I-\mathcal{P}_T\mathcal{P}_{\Omega})^{-1}](X)\right\|_2 \\ &\leq \alpha_i d \left\|[\mathcal{P}_{\Omega}(I-\mathcal{P}_T\mathcal{P}_{\Omega})^{-1}](X)\right\|_{\infty} \\ &\leq \alpha_i d \left\|[(I-\mathcal{P}_T\mathcal{P}_{\Omega})^{-1}](X)\right\|_{\infty} \\ &\leq \frac{\alpha_i d \left\|X\right\|_{\infty}}{1-\xi(\alpha,d,\mathcal{G})}.\end{aligned}$$

The first inequality holds by dropping the projection $\mathcal{P}_{ii}$. The second inequality holds due to the fact that the operator norm of a matrix with at most d entries per row/column can be bounded above by d times the maximum element in magnitude of the matrix. The third inequality holds by dropping the operator $\mathcal{P}_{\Omega}$. The final inequality follows from geometric convergence, as before.

Proof of Theorem 2

We prove that under the assumptions of this theorem the sufficient conditions of Lemma 32 are satisfied. Set $R = \sum_{i=1}^{m} \gamma_i P_i$ where $P_i \in \mathbb{S}^n$ is the projection matrix corresponding to the i'th eigenspace of $\mathcal{G}$. Denote the edits by a matrix $E^* \in \mathbb{S}^n$, and let the subspace of matrices with nonzero entries contained inside the support of E^* be denoted Ω. Set $M = \text{sign}(E^*) - \mathcal{P}_{\Omega}(R)$ and note that $M \in \Omega$. Condition 5 of Lemma 32 is satisfied based on assumption 1 of Theorem 2. As a result, the operator $\mathcal{L}^{\alpha}$ is well-defined by Lemma 33. Set $\Delta = \mathcal{L}^{\alpha}(M)$. We prove that $Q = R + \Delta$ satisfies the requirements of Lemma 32.

Condition 1 of Lemma 32: One can check that:

$$\mathcal{P}_{\Omega}(\Delta) + \mathcal{P}_{\Omega}(R) = \mathcal{P}_{\Omega}(\mathcal{L}^{\alpha}(M)) + \mathcal{P}_{\Omega}(R) = \mathcal{P}_{\Omega}(M) + \mathcal{P}_{\Omega}(R) = \text{sign}(E^*).$$

Here the second equality holds due to property (P1) of the operator $\mathcal{L}^{\alpha}$.

Condition 2 of Lemma 32: We have that:

$$\begin{aligned}\left\|\mathcal{P}_{\Omega^c}(\Delta)\right\|_{\infty} + \left\|\mathcal{P}_{\Omega^c}(R)\right\|_{\infty} &\leq \frac{\xi(\alpha,d,\mathcal{G})\left\|M\right\|_{\infty}}{1-\xi(\alpha,d,\mathcal{G})} + \left\|R\right\|_{\infty} \\ &\leq \frac{\xi(\alpha,d,\mathcal{G}) + \rho(\gamma,\mathcal{G})}{1-\xi(\alpha,d,\mathcal{G})} < 1.\end{aligned}$$

The first inequality employed assertion 1 of Lemma 33, the second inequality follows from the triangle inequality and the definition of M, and the last inequality holds by assumption 1 of the theorem.

Condition 3 of Lemma 32: Follows from property (P2) of operator $\mathcal{L}^\alpha$.

Condition 4 of Lemma 32: One can check that:

$$\begin{aligned}\|\mathcal{P}_{ii}(\Delta)\|_2 + \|\mathcal{P}_{i+1,i+1}(\Delta)\|_2 &= \|[\mathcal{P}_{ii}\mathcal{L}^\alpha](M)\|_2 + \|[\mathcal{P}_{i+1,i+1}\mathcal{L}^\alpha](M)\|_2 \\ &\leq \frac{(\alpha_i + \alpha_{i+1})(1+\rho(\gamma,\mathcal{G}))\,d}{1-\xi(\alpha,d,\mathcal{G})} \\ &< \gamma_{i+1} - \gamma_i,\ \forall\, i \in \{1,\ldots,m-1\}.\end{aligned}$$

Here the first inequality follows from assertion 2 of Lemma 33 and the triangle inequality, and the second inequality follows from the assumption of the theorem.

Proof of Corollary 31

For this proof we require the notion of incoherence of a subspace, which measures how well the subspace is aligned with the standard basis vectors. This notion appears prominently in results on sparse signal recovery via convex optimization [45].

Definition 34. *Let $S \subseteq \mathbb{R}^n$ be a subspace and let P_S be the corresponding projection onto S. The* incoherence *of S is denoted $\mu(S)$ and is defined as*

$$\mu(S) := \max_i \|P_S e_i\|_2.$$

Here e_i is the i'th standard basis vector.

For any projection matrix P_S, one can check that the inequality $\|P_S\|_\infty \leq \mu(S)^2$ is satisfied.

Remark 35. *For vertex-transitive graphs, the diagonal entries of a projection matrix associated to any eigenspace of the graph are identical. As a result, the incoherence of an eigenspace $\mathcal{E}$ of a vertex-transitive graph on n vertices is equal to*

$$\mu(\mathcal{E}) = \sqrt{\frac{\dim(\mathcal{E})}{n}}.$$

We now proceed to the proof of the corollary. Denote the eigenspaces of $\mathcal{G}$ by $\mathcal{E}_i$ for $i \in \{1, \ldots, m\}$ ordered by decreasing eigenvalue order. Remark 35 implies that $\mu(\mathcal{E}_i) = \sqrt{\frac{\dim(\mathcal{E}_i)}{n}}$. Denote the second largest coherence of the eigenspaces of $\mathcal{G}$ by $\bar{\mu} = \sqrt{\frac{\kappa}{n}}$, and denote the index of the eigenspace with the highest incoherence by ℓ. Set $\alpha_\ell = 1$ and the remaining entries of α to 0. Furthermore, choose γ such that:

$$\gamma_{i+1} - \gamma_i = c_1 \frac{\alpha_i + \alpha_{i+1}}{\bar{\mu}^2} + \epsilon, \quad \forall\, i \in \{1, \ldots, m-1\}, \text{ for some } c_1 > 0,\, \epsilon > 0.$$

Here c_1 and ϵ are positive constants that can be as small as desired. To establish condition 2 of Theorem 2, we prove that the inequality below holds for all $i \in \{1, \ldots, m-1\}$:

$$\begin{aligned} \frac{(\alpha_i + \alpha_{i+1})\,(1 + \rho(\gamma, \mathcal{G}))\,d}{1 - \xi(\alpha, d, \mathcal{G})} &\leq \gamma_{i+1} - \gamma_i - \epsilon \\ &= c_1 \frac{\alpha_i + \alpha_{i+1}}{\bar{\mu}^2},\ \forall\, i \in \{1, \ldots, m-1\}. \end{aligned}$$

Clearly, if $\alpha_i + \alpha_{i+1} = 0$ for some i, then the corresponding inequality is satisfied. On the other hand, all the remaining inequalities can be collapsed to a single one by dividing both sides of all such inequalities by $\alpha_i + \alpha_{i+1}$:

$$\frac{(1 + \rho(\gamma, \mathcal{G}))d}{1 - \xi(\alpha, d, \mathcal{G})} \leq c_1 \frac{n}{\kappa},$$

a sufficient condition for which is:

$$\frac{(1 + \rho(\gamma, \mathcal{G}))c}{1 - \xi(\alpha, d, \mathcal{G})} \leq c_1. \tag{3.5}$$

In the remainder of the proof, we show that our particular choice of γ and α satisfy inequality (3.5) and Theorem 2 condition 1.

We bound $\rho(\gamma, \mathcal{G})$ from above via a change of variable. Setting $\tilde{\gamma} = \gamma_1 \frac{\bar{\mu}^2}{c_1}$, we

have that:

$$\rho(\gamma, \mathcal{G}) = \left\| \sum_{i=1}^{m} \gamma_i P_i \right\|_\infty = \left\| \sum_{i=1}^{m} \Big[\gamma_1 + \sum_{j=1}^{i-1} \big(\frac{c_1(\alpha_j + \alpha_{j+1})}{\bar{\mu}^2} + \epsilon \big) \Big] P_i \right\|_\infty \tag{3.6}$$

$$\leq \left\| \sum_{i=1}^{m} \Big[\gamma_1 + \sum_{j=1}^{i-1} \frac{c_1(\alpha_j + \alpha_{j+1})}{\bar{\mu}^2} \Big] P_i \right\|_\infty + \sum_{i=1}^{m} \sum_{j=1}^{i-1} \epsilon \left\| P_i \right\|_\infty \tag{3.7}$$

$$\leq \frac{c_1}{\bar{\mu}^2} \left\| \sum_{i=1}^{m} \Big[\tilde{\gamma} + \sum_{j=1}^{i-1} (\alpha_j + \alpha_{j+1}) \Big] P_i \right\|_\infty + \epsilon c_3 \tag{3.8}$$

$$\leq \frac{c_1}{\bar{\mu}^2} \sum_{i=1}^{m} \Big[|\tilde{\gamma} + \sum_{j=1}^{i-1} (\alpha_j + \alpha_{j+1})| \left\| P_i \right\|_\infty \Big] + \epsilon c_3 \tag{3.9}$$

$$\leq c_1 c_2 + \epsilon c_3. \tag{3.10}$$

Here (3.7) follows by grouping all terms with ϵ and using the triangle inequality, (3.8) follows by the change of variables described above and bounding all the terms in the right summand from above by one, and (3.9) follows from the triangle inequality. We choose the remaining degree of freedom $\tilde{\gamma}$ to eliminate the contribution of the subspace with the highest incoherence parameter in the left summand. Consequently, (3.10) follows by bounding the infinity norms of the remaining projection matrices from above by $\bar{\mu}^2$. Crucially, the fact that $\alpha \in [0,1]^m$ and m are viewed as fixed enables us to bound the sum from above with positive constants c_2 and c_3 that depend only on m.

Next, we use our particular choice for α to bound $\xi(\alpha, d, \mathcal{G})$ from above. In particular, for any $W \in \mathbb{S}^n$ we have:

$$\begin{aligned}
&\left\| \big[(I - \sum_{i=1}^{m} \alpha_i \mathcal{P}_{ii}) \mathcal{P}_\Omega \big](W) \right\|_\infty \\
&= \left\| \sum_{\substack{i=1 \\ i \neq \ell}}^{m} \Big(P_i \mathcal{P}_\Omega(W) + \mathcal{P}_\Omega(W) P_i - P_i \mathcal{P}_\Omega(W) P_i \Big) - \sum_{\substack{i=1 \\ i \neq \ell}}^{m} \sum_{\substack{j=1 \\ j \neq i,\ell}}^{m} P_i \mathcal{P}_\Omega(W) P_j \right\|_\infty \\
&\leq \sum_{\substack{i=1 \\ i \neq \ell}}^{m} \Big(\left\| P_i \mathcal{P}_\Omega(W) \right\|_\infty + \left\| \mathcal{P}_\Omega(W) P_i \right\|_\infty + \left\| P_i \mathcal{P}_\Omega(W) P_i \right\|_\infty \Big) + \\
&\qquad\qquad \sum_{\substack{i=1 \\ i \neq \ell}}^{m} \sum_{\substack{j=1 \\ j \neq i,\ell}}^{m} \left\| P_i \mathcal{P}_\Omega(W) P_j \right\|_\infty
\end{aligned}$$

$$\leq \left[\sum_{\substack{i=1 \\ i \neq \ell}}^{m} 2\mu(\mathcal{E}_i)\sqrt{d} + \left(\sum_{\substack{i=1 \\ i \neq \ell}}^{m} \mu(\mathcal{E}_i)^2 + \sum_{\substack{i=1 \\ i \neq \ell}}^{m} \sum_{\substack{j=1 \\ j \neq i,\ell}}^{m} \mu(\mathcal{E}_i)\mu(\mathcal{E}_j) \right) d \right] \|W\|_\infty$$

$$\leq (c_4 \bar{\mu}\sqrt{d} + c_5 \bar{\mu}^2 d) \|W\|_\infty$$

$$= (c_4 \sqrt{\frac{\kappa d}{n}} + c_5 \frac{\kappa d}{n}) \|W\|_\infty$$

$$\leq (c_4 \sqrt{c} + c_5 c) \|W\|_\infty , \tag{3.11}$$

for some positive real numbers c_4 and c_5 depending only on m. Here the first equality is obtained by rearranging the sum in terms of the projection matrices P_i, the first inequality is due to the triangle inequality, and the second inequality is a consequence of the following inequalities:

$$\|P_i \mathcal{P}_\Omega(W)\|_\infty \leq \mu(\mathcal{E}_i)\sqrt{d} \|W\|_\infty ,$$
$$\|\mathcal{P}_\Omega(W) P_i\|_\infty \leq \mu(\mathcal{E}_i)\sqrt{d} \|W\|_\infty ,$$
$$\|P_i \mathcal{P}_\Omega(W) P_j\|_\infty \leq \mu(\mathcal{E}_i)\mu(\mathcal{E}_j) d \|W\|_\infty ;$$

which hold for all $i, j \in \{1, \ldots, m\}$.

Equations (3.10) and (3.11) assert that $\rho(\gamma, \mathcal{G})$ and $\xi(\alpha, d, \mathcal{G})$ can lowered as desired by reducing the constants c_1, ϵ and c. Consequently, one can check that both condition 1 of Theorem 2 and equation (3.5) (which implies condition 2 of Theorem 2) can be satisfied by first choosing a sufficiently small c_1 and ϵ (both depending on m) to bound $\rho(\gamma, \mathcal{G})$ from above, and then suitably choosing a sufficiently small c depending on m, c_1 and ϵ.

3.4 Numerical Illustrations with Invariants based on Stable Sets and Cuts

In this section we evaluate the utility of two invariant convex sets based on (tractable relaxations of) the inverse of the stability number and the maximum cut value, both of which are described in Section 3.1. Our investigation is via numerical experiments rather than theoretical bounds as in Section 3.2. The primary reason for this choice is that we do not have a detailed understanding of the face structure of the invariant convex sets considered in this section; in contrast, we have a precise (and convenient for the purposes of analysis) characterization of the geometry of the Schur-Horn orbitope, which played a crucial role in the theoretical results of the previous section. Nonetheless, we pursue a systematic approach in the present section by identifying classes of

graphs that are 'brittle' in the sense that deleting / adding a small number of edges results in large changes in their stability number / maximum cut value. Such graph families present excellent examples for which invariant convex sets based on the inverse of stability number and the maximum cut value are particularly well-suited to obtaining useful bounds on the graph edit distance. More broadly, our discussion in this section highlights the larger point that our framework (P) can be tailored to the particular structural properties of the underlying graphs to yield useful lower bounds on the edit distance.

Constraining the Inverse of the Stability Number

The function $f(A)$ described in Section 3.1 is an efficiently computable lower bound on the inverse of the stability number of a graph, and further it is a concave graph invariant. Consequently, super-level sets of this function provide tractable invariant convex sets that may be employed in our framework (P). Given a graph $\mathcal{G}$, we denote the associated set by $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$:

$$\begin{aligned}\mathcal{C}_{\mathcal{IS}(\mathcal{G})} :&= \{M \in \mathbb{S}^n \mid f(M) \geq f(A)\} \\ &= \{M \in \mathbb{S}^n \mid \exists \mu \in \mathbb{S}^n,\, \mu \geq 0,\; I + M - \mu - f(A)11^T \succeq 0\}. \quad (3.12)\end{aligned}$$

Here A is any adjacency matrix representing $\mathcal{G}$. From this description, it is immediately clear that for any edit to $\mathcal{G}$ that corresponds to an increase in the value of the function f, the constraint $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$ is inactive. Adding edges to a graph can only reduce the stability number, and hence can potentially only increase the inverse of the stability number. Although the function f is only a lower bound on the inverse of the stability number, it satisfies a similar monotonocity property in that the value of f is non-decreasing with the addition of edges to a graph. The following lemma formalizes matters by describing the tangent cone at an adjacency matrix of a graph $\mathcal{G}$ with respect to the set $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$:

Lemma 36. *For any graph $\mathcal{G}$ on n vertices and associated adjacency matrix $A \in \mathbb{S}^n$, let $\alpha^* = f(A)$, i.e., the value corresponding to the Motzkin-Straus relaxation of the inverse of the stability number. Then we have that:*

$$\mathcal{T}_{\mathcal{C}_{\mathcal{IS}(\mathcal{G})}}(A) = \{T \in \mathbb{S}^n | \exists \mu, \Lambda \in \mathbb{S}^n,\, \mu \geq 0, \Lambda \succeq 0,\; T + I + A - \alpha^* 11^T = \mu + \Lambda\}.$$

Proof. The proof of this lemma follows from a direct application of convex duality. □

The description of the tangent cone in Lemma 36 is based on the dual of the cone of doubly nonnegative matrices; see [12] for more details on this connection. In particular, this lemma implies that entrywise nonnegative matrices belong to the tangent cone at an adjacency matrix A representing a graph $\mathcal{G}$ with respect to the set $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$; consequently, edits to $\mathcal{G}$ consisting purely of addition of edges are feasible directions with respect to the set $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$ and for such edits this set does not provide useful lower bounds on the edit distance. Thus, we investigate the utility of the constraint $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$ in settings in which the edits consist mainly of edge deletions. Such problems arise in the context of *graph completion* in which the objective is to add edges to a given graph so that the resulting graph satisfies some desired property.

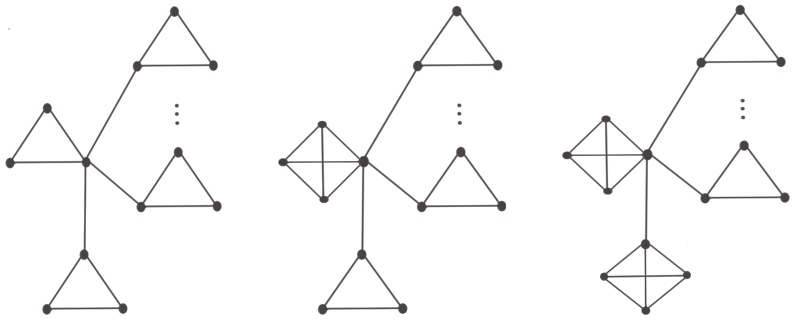
Building on this discussion, the constraint set $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$ is most likely to be useful for graphs $\mathcal{G}$ in settings in which the deletion of even a small number of edges of $\mathcal{G}$ results in an increase in the stability number. Graphs that have a large number of stable sets with cardinality equal to the stability number offer a natural prospect for further exploration. Fortunately, such graphs have been studied in extremal graph theory literature, from which we quote the following result [70]:

Theorem 3. *[70] For $s, n \in \mathbb{N}$ with $n \geq 6$, let*

$$h(n) = \begin{cases} 2 \times 3^{s-1} + 2^{s-1}, & \text{if } n = 3s, \\ 3^s + 2^{s-1}, & \text{if } n = 3s+1, \\ 4 \times 3^{s-1} + 3 \times 2^{s-2}, & \text{if } n = 3s+2. \end{cases}$$

Let $\mathcal{G}$ be any connected graph on n vertices, and denote the cardinality of the set of all maximum independent sets of $\mathcal{G}$ by $\phi(\mathcal{G})$. Then $\phi(\mathcal{G}) \leq h(n)$ with equality if and only if $\mathcal{G}$ is isomorphic to one of the graphs shown in Figure 3.4.

This theorem states that the graphs $E(n)$ shown in Figure 3.4 are precisely the connected graphs that have the largest number of distinct maximum independent sets. As such, they present a natural test case to investigate the utility of the constraint set $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$ in providing bounds on the graph edit distance, at least in settings in which the edits are composed predominantly of edge deletions. We illustrate here the results of numerical experiments conducted on the graph $E(30)$, which is a sparse graph with 39 edges and 396 nonedges.



(a) (b) (c)

Figure 3.4: Left to right, $E(n)$ for $n = 3s$, $n = 3s + 1$, $n = 3s + 2$. For $n = 3s + r$, these graphs are formed by connecting $(s - r)\, K_3$'s and $r\, K_4$'s through edges connecting to a specific vertex.

The setup of this experiment is the same as that described in Section 3.2 with one notable exception: in the present experiment, we assume asymmetric edits rather than symmetric edits so that 80% of the edits are edge deletions while 20% are edge additions. We range the total number of edits from 5 to 45 with increments of 5, and for each number of edits we repeat our experiment 1000 times. In each iteration, we obtain a bound on the graph edit distance between $E(30)$ and the modified graph using our framework (P) with three different constraint sets: the Schur-Horn orbitope, the constraint set $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$, and an invariant convex set based on the Goemans-Williamson relaxation of the maximum cut value (which is discussed in greater detail in the next subsection). Figure 3.5 reports the ratio of the average computed lower bound on the graph edit distance to the number of edit operations for each constraint set. (The number of edits is an upper bound on the true graph edit distance.) As one might expect, the relaxation based on the constraint $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$ yields the best lower bounds of the three approaches. Specifically, even when a majority of the edges of $E(30)$ are removed, the constraint set $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$ continues to provide lower bounds that are at least 40% of the number of edit operations. In contrast, the bounds provided by the Schur-Horn orbitope constraint are much weaker, and those obtained using the Goemans-Williamson relaxation of the maximum cut value are ineffective.

Constraining the Maximum Cut Value

In analogy with the inverse of the stability number, the function $g(A)$ due to Goemans and Williamson [59] that is described in Section 3.1 provides an

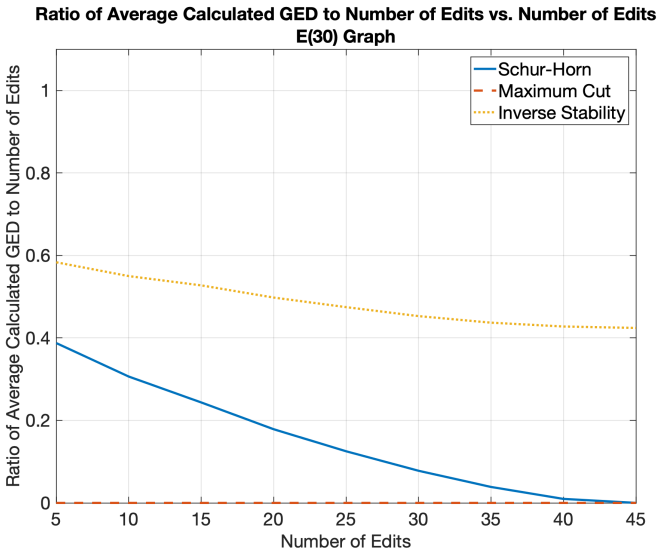
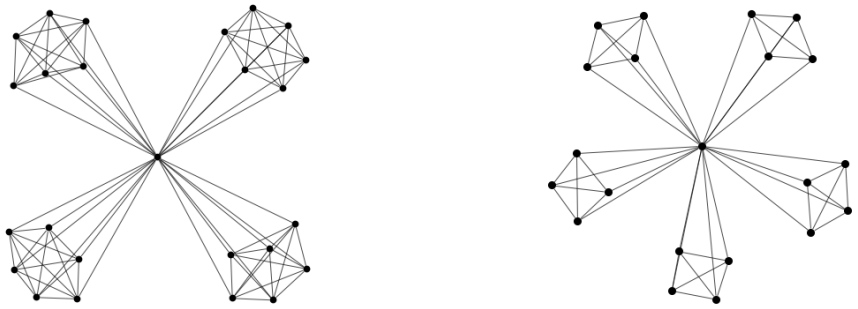


Figure 3.5: Ratio of average computed lower bound on graph edit distance to number of edit operations. Experiment conducted on $E(30)$ graph. The edit operations are 80% edge deletions and 20% edge additions.

efficiently computable upper bound on the maximum cut value of a graph. As this function is invariant to conjugation of its argument by permutation matrices, its sublevel sets are invariant convex sets. For a graph $\mathcal{G}$, we denote the associated set by $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$:

$$\begin{aligned}\mathcal{C}_{\mathcal{MC}(\mathcal{G})} :=& \{M \in \mathbb{S}^n \mid g(M) \leq g(A)\} \\ =& \{M \in \mathbb{S}^n \,|\, \exists D \in \mathbb{S}^n \text{ diagonal}, \; M - D \succeq 0, \frac{1}{4}\mathrm{Tr}(M11' - D) \leq g(A)\},\end{aligned} \tag{3.13}$$

where A is an adjacency matrix representing $\mathcal{G}$. Reasoning in a similar manner as in the previous subsection, we observe that edits corresponding to a decrease in the value of the function g represent feasible directions with respect to the set $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$, and for such edits the constraint $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$ is inactive. Deleting edges from a graph reduces its maximum cut value, and one can check that directions represented by entrywise nonpositive matrices belong to the tangent cone at an adjacency matrix A representing $\mathcal{G}$ with respect to $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$. Consequently, we should only expect the constraint $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$ to potentially provide useful lower bounds on the graph edit distance in settings in which most of the edits to a graph $\mathcal{G}$ correspond to edge additions. In some sense, this type of a graph inverse problem – removing the smallest number of edges from a graph so



(a) Windmill graph $D(4, 7)$ (b) Windmill graph $D(5, 5)$

Figure 3.6: Two sample Windmill graphs.

that it satisfies a desired property – is a complement of the graph completion problem discussed in the previous subsection.

Building further on the preceding discussion, we remark that the constraint set $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$ is most likely to be effective if adding even a small number of edges to $\mathcal{G}$ increases the value of the function g. A prominent example of such graphs are the so-called Windmill graphs shown in Figure 3.6. The Windmill graph $D(m, n)$ is constructed by taking m copies of the complete graph K_n and intersecting them at a single vertex. Due to the ample amount of symmetry in these graphs, there are many partitions of the vertices into two sets that achieve the maximum cut value – the number of such partitions is $\binom{n-1}{n/2}^m$ for even n and $\binom{n}{(n-1)/2}^m$ for odd n. Thus, Windmill graphs present a natural test family to evaluate the power of the constraint set $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$ when the graph edits consist primarily of the addition of edges. We present the results of numerical experiments on the Windmill graph $D(4, 7)$ in a setting that closely mirrors the one in the previous subsection. The Windmill graph $D(4, 7)$ is a graph on 25 nodes with 84 edges and 216 non-edges. The edits made to this graph consist mostly of edge additions – 80% are edge additions and the remaining 20% are edge deletions. We vary the number of edits from 10 to 200 with increments of 5 and consider 1000 random instances of perturbations for each number of edits. For each problem instance, we obtain a lower bound on the edit distance by utilizing our framework (P) with the Schur-Horn orbitope constraint, the Motzkin-Straus relaxation from the previous subsection, and the constraint $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$. We report the average ratio of the computed lower bound on the graph edit distance to the number of edit operations in Figure 3.7. (As before the number of edits is an upper bound on the graph edit distance.)

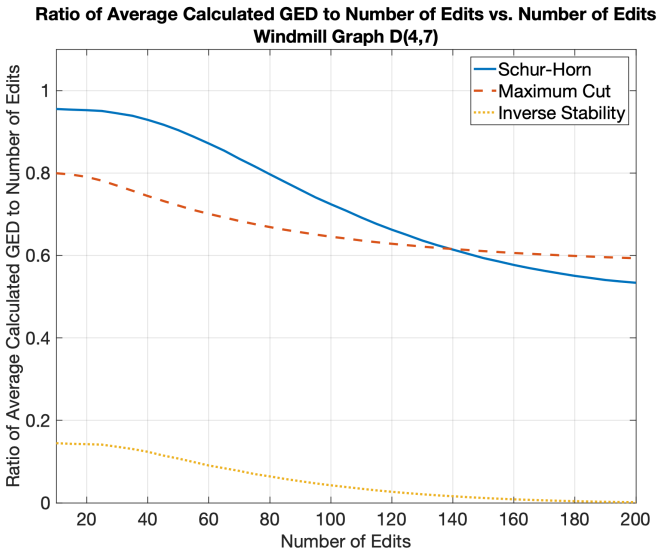


Figure 3.7: Ratio of average computed lower bound on the graph edit distance to number of edit operations. Experiment conducted on Windmill graph $D(4, 7)$. The edit operations are 80% edge additions and 20% edge deletions.

From Figure 3.7 we see that the Schur-Horn orbitope constraint produces the best lower bounds for the graph edit distance when the number of edits is small, whereas the constraint $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$ produces the best lower bounds when the number of edits is large. On average, both of these constraints provide bounds that are consistently better than 50% of the total number of edits. As the edits consist mainly of edge additions, the constraint based on the Motzkin-Straus relaxation of the inverse of the stability number performs poorly.

3.5 Experiments with Real Data

In this section, we present experimental results that demonstrate the utility of our framework on real data. We begin by introducing an extension of our framework to allow for edits that include vertex additions and deletions. We then describe the bounds obtained on two widely studied datasets consisting of molecular structures.

Enabling vertex additions and deletions

In many situations, one wishes to obtain bounds on the edit distance between two graphs consisting of different numbers of vertices. In such cases one allows vertex insertions and deletions in addition to the usual operations of edge insertions and deletions that we've considered thus far. To extend our frame-

work to this setting, we allow an adjacency matrix to take on nonzero values on the diagonal to denote the presence or absence of a vertex. Specifically, we consider a "vertex-indexed adjacency matrix" $A \in \mathbb{S}^n$ with entries equal to either zero or one and in which $A_{ij} = 1,\ j \neq i$ implies that $A_{ii} = 1$. In words, a value of one on the i'th diagonal entry implies that a vertex corresponding to that index is 'present' in the graph, and an edge being incident on a vertex implies that the vertex must be present in the graph. (Note that a value of one on a diagonal entry does not represent a "vertex weight" but instead the presence of a vertex in a graph.) With this notation in hand, we are now in a position to describe a generalization of our framework that allows for vertex deletions and insertions. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two unweighted and unlabeled graphs on n_1 and n_2 vertices, respectively, and let $n := \max\{n_1, n_2\}$. Letting $A_2 \in \mathbb{S}^n$ specify a vertex-indexed adjacency matrix for $\mathcal{G}_2$ with zeros on the diagonal corresponding to those indices that do not correspond to a vertex (when $n_2 < n$), consider the following convex optimization problem:

$$\begin{aligned} \mathrm{GED}_{LB}(\mathcal{G}_1, \mathcal{G}_2; \mathcal{C}_{\mathcal{G}_1}) = \min_{X, E \in \mathbb{S}^n} & \sum_{1 \leq i \leq j \leq n} |E_{ij}| \\ s.t. \quad & X + E = A_2 \\ & X \in \mathcal{C}_{\mathcal{G}_1} \\ & X_{ij} \leq X_{ii},\ X_{ij} \leq X_{jj}\ \forall i, j \in \{1, \ldots, n\}. \end{aligned} \tag{P_{ext}}$$

Here the set $\mathcal{C}_{\mathcal{G}_1}$ is an invariant convex set associated to $\mathcal{G}_1$, and the matrices X, A_1, and A_2 are to be interpreted as vertex-indexed adjacency matrices. There are two main differences between the convex program (P_{ext}) and the convex problem (P). The first is in the objective function in which we only sum the upper triangular elements of the matrix E in the program (P_{ext}), as we do not wish to double-count the edge edits relative to vertex edits. The second modification arises in the constraint in the last line of (P_{ext}) based on the observation that if edges are incident to a vertex, then this vertex must be 'present'. Using a line of reasoning similar to that following the presentation of (P), one can conclude that the optimal value of the convex program (P_{ext}) provides a lower bound on the graph edit distance between $\mathcal{G}_1$ and $\mathcal{G}_2$ with the permissible edit operations being vertex additions/deletions and edge additions/deletions. Finally, we note that our framework can also accommodate situations in which the cost of a vertex edit operation is different from that of an edge edit operation.

Experimental Results on Chemistry Datasets

We employ the convex program (P_{ext}) to obtain lower bounds on graph edit distance problems arising in chemistry. Specifically, we conduct experiments on two datasets known as the Polycyclic Aromatic Hydrocarbons (PAH) dataset and the Alkane dataset.[1] Both of these datasets consist of unlabeled, unweighted graphs representing chemicals, with the vertices of the graphs corresponding to carbon atoms in a molecule and edges specifying bonds between two carbons. These datasets have been used as benchmarks for evaluating the performances of graph edit distance algorithms; for example, see [1] and [38] for comparisons of the performance of various algorithms on these datasets. For each dataset, we compare upper bounds on the average edit distance taken over all pairs of graphs (obtained using other procedures) with lower bounds on the average obtained using our method.

The Alkane dataset consists of 150 unlabeled, acyclic graphs representing alkanes, with the number of vertices ranging from 1 to 10 vertices (the average is 8.9) and an average degree of 1.8. As these graphs are relatively small in size, the average pairwise graph edit distance for this dataset can be calculated exactly using combinatorial algorithms such as the A^* procedure [65]. The PAH dataset consists of 94 graphs representing polycyclic aromatic hydrocarbons. As with the Alkane dataset, the vertices of the graphs in this dataset denote carbon atoms, and two vertices are connected if there exists a bond between the corresponding carbons. Unlike the Alkane dataset, the chemicals in the PAH dataset represent large compounds: the smallest graph in PAH has 10 vertices, the largest graph has 28 vertices, and the average number of vertices is 20.7. The average degree of the graphs in the PAH dataset is 2.4. Due to this larger size, calculating the exact average pairwise graph edit distance of the PAH dataset is prohibitively expensive. In fact, to the best of our knowledge, the exact average pairwise graph edit distance of the PAH dataset is unknown to this date [17]. Consequently, obtaining guaranteed lower bounds on the average graph edit distance of the PAH dataset is especially useful as a way to compare to known average upper bounds.

For each pair of graphs, we employ the convex program (P_{ext}) twice by switching the roles of $\mathcal{G}_1$ and $\mathcal{G}_2$, and take the larger optimal value as our lower bound. In each case we utilize four different types of invariant convex set

[1] Available online at https://brunl01.users.greyc.fr/CHEMISTRY/.

constraints: the Schur-Horn orbitope ($\mathcal{C}_{\mathcal{SH}}$), the Motzkin-Straus bound on the inverse of the stability number ($\mathcal{C}_{\mathcal{IS}}$), the Goemans-Williamson bound on the maximum cut value ($\mathcal{C}_{\mathcal{MC}}$), and finally the intersection of all these three constraints ($\mathcal{C}_{\mathcal{SH}} \cap \mathcal{C}_{\mathcal{IS}} \cap \mathcal{C}_{\mathcal{MC}}$). In our experiments, we follow the convention adopted in the graph edit distance literature with these two datasets, namely that the cost of an edit operation is equal to three.[2] The average pairwise lower bounds obtained using our convex program (P_{ext}) on the Alkane and PAH datasets are given in Table 3.1.

Dataset Name	Best known upper bound on avg. GED	Lower bounds on avg. GED via (P_{ext})			
		$\mathcal{C}_{\mathcal{MC}}$	$\mathcal{C}_{\mathcal{IS}}$	$\mathcal{C}_{\mathcal{SH}}$	$\mathcal{C}_{\mathcal{MC}} \cap \mathcal{C}_{\mathcal{IS}} \cap \mathcal{C}_{\mathcal{SH}}$
Alkane	15.3 (exact) [38]	4.66	6.12	9.58	10.72
PAH	29.8 [38]	12.01	14.52	20.29	21.60

Table 3.1: Average pairwise graph edit distances of the Alkane and PAH datasets. Edit operations are limited to edge and vertex additions and removals. Every edit operation incurs a cost of 3.

There are a number of interesting aspects to these results. For both datasets, a constraint based only on the Goemans-Williamson relaxation seems to produce the worst lower bounds (4.66 and 12.01), while the Schur-Horn orbitope constraint produces the best lower bounds (9.58 and 20.29) when only a single type of invariant convex constraint is employed. As expected, the combination of all three individual constraint sets produces the best overall lower bounds (10.72 and 21.60). More broadly, these results demonstrate the effectiveness of our approach in producing useful lower bounds for graph edit distance problems arising from real data in a computationally tractable manner. Specifically, for the Alkane dataset the average lower bound 10.72 is obtained using our convex programming framework and the exact value of the average graph edit distance is 15.3 (which is obtained via combinatorial approaches). Our results have more interesting implications for the PAH dataset as it is prohibitively expensive to compute the exact average graph edit distance for this dataset due to the large size of its constituents. In particular, the best-known upper bound on the average graph edit distance of PAH is 29.8 [38]. Our convex relaxation framework produces a lower bound of 21.6 on the average graph

[2]The reason for this choice in that community is that vertex/edge deletions/insertions are considered more significant edit operations than vertex/edge label substitutions which have a lower cost of one associated to them (in this chapter, we do not consider such edits based on substitutions).

edit distance over all pairs of graphs in PAH, which provides a floor on the possible improvement that one should expect to obtain via better algorithms for computing graph edit distances.

3.6 Discussion

In this chapter we introduce a framework based on convex graph invariants for obtaining lower bounds on the edit distance between two graphs. Much of the literature on this topic provides methods for computing upper bounds on the edit distance between two graphs by identifying a feasible sequence of edits to transform one graph to the other. Our approach is qualitatively different in that it is based on convex relaxation and it leads to guaranteed lower bounds on the edit distance. Further, our approach can be adapted to the structure underlying the two graphs. We provide both theoretical and empirical support for our method.

There are a number of potential directions for further investigation arising from this chapter. First, our analysis of the performance of the Schur-Horn relaxation could potentially be tightened in order to obtain sharper conditions for the success of our algorithm. For example, Corollary 31 only utilizes information about the second most-repeated eigenvalue, and while this provides order-optimal scaling results for families such as triangular graphs, it may be possible to improve our analysis to obtain order-optimal bounds for other families as well. More broadly, a key step in carrying out a precise theoretical analysis of the power of an invariant convex constraint set is to obtain a full understanding of the facial structure of the set, and it would be of interest to develop such a characterization for a larger suite of invariant convex sets than those presented in this chapter. Finally, a commonly encountered question in many applications is to test whether a given graph is a minor of another graph, and it would be useful to extend the framework described in this chapter to address this problem.

Chapter 4

SUM OF SQUARES BASED CONVEX RELAXATIONS FOR INVERSE EIGENVALUE PROBLEMS

4.1 Introduction

The affine inverse eigenvalue problem (IEP) consists of identifying a real symmetric matrix with a prescribed set of eigenvalues in an affine space. IEPs arise in a range of applications in engineering and physical sciences, such as natural frequency identification in vibrating systems [10, 37, 56, 101], pole placement [22], factor analysis [64], reliability testing [52], estimation of the Earth's conductivity [97], graph partitioning [36], and nuclear and molecular spectroscopy [57, 113]. Further, there are many situations in which a question of interest is to solve a discrete inverse Sturm-Liouville problem [63, 82], which is a special case of an affine IEP. Due to its ubiquity, IEPs have received much attention in the literature over the past several decades (see the surveys [15, 33, 34]). On one end of the spectrum, there have been several efforts aimed at providing necessary and sufficient conditions for the existence of a solution of a given IEP [40–42, 53, 66, 77, 124]. For example, Landau proved that there always exists a symmetric Toeplitz matrix with a desired set of eigenvalues [77]; however, computing such matrices in a tractable manner remains a challenge. At the other end of the spectrum, several efforts have been aimed at developing efficient procedures for numerically finding solutions to particular types of the inverse eigenvalue problems [13, 39, 54, 120], including some recent approaches based on convex optimization [7, 83, 96, 123]. Our work differs from these approaches in two prominent ways. First, our framework is applicable to general affine IEPs, while some of the previous convex approaches are only useful for certain structured problem instances; see Section 4.3 for the broad range of examples to which we apply our methods. Second, we describe a family of convex relaxations for IEPs rather than just a single convex program, and our work allows for a tradeoff between computational cost and solution quality.

We begin by first reformulating the affine IEP as a question of checking the existence of a real solution to a system of polynomial equations. Formally, an instance of an affine IEP may be stated as follows:

Affine Inverse Eigenvalue Problem. *Given* (i) *a desired spectrum* $\Lambda = \{(\lambda_i, m_i)\}_{i=1}^q \subset \mathbb{R} \times \mathbb{Z}_+$ *of eigenvalue-multiplicity pairs with* $\sum_i m_i = n$*, and* (ii) *an affine space* $\mathcal{E} = \{X \in \mathbb{S}^n \ : \ \mathrm{Tr}\,(C_k X) = b_k, \ k = 1, \ldots, \ell\}$*, find an element of* $\mathcal{E}$ *with spectrum given by* Λ *or certify that such a matrix does not exist. Here* $\mathbb{S}^n$ *denotes the space of* $n \times n$ *real symmetric matrices.*

This problem may be reformulated as checking whether the following system has a real solution:

$$\mathcal{S}_{\text{iep}} := \begin{cases} f_1 := \sum_{i=1}^q Z_i - I = 0, \\ f_2^{(i)} := \mathrm{Tr}\,(Z_i) - m_i = 0 \text{ for } i = 1, \ldots, q, \\ f_3^{(i)} := Z_i^2 - Z_i = 0 \text{ for } i = 1, \ldots, q, \\ f_4^{(k)} := \sum_{i=1}^q \lambda_i \mathrm{Tr}\,(Z_i C_k) - b_k = 0 \text{ for } k = 1, \ldots, \ell. \end{cases} \tag{4.1}$$

The variables in this system are the matrices $Z_1, \ldots, Z_q \in \mathbb{S}^n$. The matrix I denotes the $n \times n$ identity. We are concerned with whether the system of polynomials $\mathcal{S}_{\text{iep}} = \{f_1, f_2^{(1)}, \ldots, f_2^{(q)}, f_3^{(1)}, \ldots, f_3^{(q)}, f_4^{(1)}, \ldots, f_4^{(\ell)}\}$ has a common zero over the reals, or in other words checking whether the associated real variety $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$ is empty. It is clear that the system of equations (4.1) encodes the IEP. The equations $f_1, \{f_2^{(i)}\}_{i=1}^q, \{f_3^{(i)}\}_{i=1}^q$ specify that the Z_i's are projection matrices that partition the identity, and the equations $\{f_4^{(k)}\}_{k=1}^\ell$ requires that the matrix $\sum_i \lambda_i Z_i$ belongs to $\mathcal{E}$. As such, an affine IEP is feasible if and only if $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}}) \neq \emptyset$.

The advantage of this polynomial reformulation is that it allows us to leverage results from the optimization literature to systematically obtain convex relaxations for the affine IEP. Specifically, Parrilo [98, 99] and Lasserre [78, 79] developed hierarchies of semidefinite programming relaxations for polynomial optimization problems using results from real algebraic geometry. These relaxations entail the solution of increasingly larger convex optimization problems that search over successively more complex collections of certificates that prove the infeasibility of the system defined by $\mathcal{S}_{\text{iep}}$. From a dual perspective, these relaxations may also be viewed as providing a sequence of convex outer approximations $\mathcal{R}_1(\Lambda, \mathcal{E}) \supseteq \mathcal{R}_2(\Lambda, \mathcal{E}) \supseteq \cdots \supseteq \mathrm{conv}\,(\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}}))$, which leads to a natural heuristic for attempting to obtain solutions of the system $\mathcal{S}_{\text{iep}}$. We describe the mechanism to obtain these relaxations in Section 4.2. As an illustration, searching over a simple class of infeasibility certificates gives the

following convex outer approximation to $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$:

$$\mathcal{R}_1(\Lambda,\mathcal{E}) = \Big\{(Z_1,\ldots,Z_q) \in \otimes^q \mathbb{S}^n \mid \sum_{i=1}^{q} Z_i = I;\ \text{Tr}\,(Z_i) = m_i,\ Z_i \succeq 0\ \forall i; \\ \text{Tr}\left(\sum_{i=1}^{q} \lambda_i Z_i C_k\right) = b_k\ \forall k\Big\}. \tag{4.2}$$

In Section 4.2 we relate the set $\mathcal{R}_1(\Lambda,\mathcal{E})$ to the *Schur-Horn orbitope* [107] associated to the spectrum Λ, which is the convex hull of all real symmetric matrices with spectrum Λ. If $\mathcal{R}_1(\Lambda,\mathcal{E}) = \emptyset$, then it is clear that $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}}) = \emptyset$; otherwise, $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$ may or may not be empty, and one can either attempt to find an element of $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$ or search over a larger family of infeasibility certificates (see Section 4.2). In Section 4.2 we describe a convex outer approximation $\mathcal{R}_2(\Lambda,\mathcal{E})$ to $\text{conv}\,(\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}}))$ that is in general tighter than $\mathcal{R}_1(\Lambda,\mathcal{E})$.

The description of $\mathcal{R}_1(\Lambda,\mathcal{E})$ in (4.2) consists of q semidefinite constraints on matrix variables of size $n \times n$. The description of $\mathcal{R}_2(\Lambda,\mathcal{E})$ in Section 4.2 involves a semidefinite constraint on a matrix variable of size $\binom{n+1}{2}q \times \binom{n+1}{2}q$. Tighter relaxations to $\text{conv}\,(\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}}))$ than $\mathcal{R}_1(\Lambda,\mathcal{E})$ and $\mathcal{R}_2(\Lambda,\mathcal{E})$ require even larger semidefinite descriptions, and they become prohibitively expensive to solve for large n. Consequently, although we describe the general mechanism by which semidefinite relaxations of increasing size may be generated, we restrict our attention in numerical experiments to the performance of the relaxations $\mathcal{R}_1(\Lambda,\mathcal{E})$ and $\mathcal{R}_2(\Lambda,\mathcal{E})$. As the affine IEP includes (co-)NP-hard problems as special cases, these two relaxations generally do not solve every instance of an affine IEP (as expected); nonetheless, we demonstrate their effectiveness in Section 4.3 on stylized problems such as certifying non-existence of planted subgraphs, solving discrete Sturm-Liouville equations, and computing Toeplitz matrices with a desired spectrum.

Connection to combinatorial optimization A number of combinatorial problems such as computing the stability number of a graph or the knapsack problem may be formulated as checking feasibility of a system of equations in a collection of variables that take on values of 0/1. As many of these problems are NP-hard, a prominent approach to developing tractable approximations is to first specify the problems via polynomial equations and to then employ the methods referenced above to obtain semidefinite relaxations [14]. The

polynomial reformulations consist of a system of equations defined by affine polynomials and quadratic equations of the form $x_i^2 - x_i = 0$ for each of the variables x_i to enforce the Boolean constraints. Our system (4.1) for the affine IEP may be viewed as a matricial analog of those arising in the literature on combinatorial problems, as the idempotence constraints $Z_i^2 - Z_i = 0$ represent a generalization of the scalar Boolean constraints $x_i^2 - x_i = 0$. The present note describes promising experimental results of the performance of semidefinite relaxations for the affine IEP. As with the significant prior body of work on combinatorial optimization, it is of interest to investigate structural properties of our relaxations for specific affine spaces $\mathcal{E}$ and spectra Λ. We outline future directions along these lines in Section 4.4.

4.2 Semidefinite Relaxations for Affine IEPs

From Polynomial Formulations to Semidefinite Relaxations

We summarize here the basic aspects of obtaining semidefinite relaxations for certifying infeasibility of a polynomial systems over the reals; we refer the reader to the survey [14] for further details. Let $\mathbb{R}[x]$ denote the ring of polynomials with real coefficients in indeterminates $x = (x_1, \ldots, x_n)$. A *polynomial ideal* $\mathcal{I}$ is a subset of $\mathbb{R}[x]$ that satisfies the following properties: (i) $0 \in \mathcal{I}$, (ii) $f_1, f_2 \in \mathcal{I} \Rightarrow f_1 + f_2 \in \mathcal{I}$, and (iii) $f \in \mathcal{I}, h \in \mathbb{R}[x] \Rightarrow hf \in \mathcal{I}$. The ideal *generated* by a collection of polynomials $f_1, \ldots, f_t \in \mathbb{R}[x]$ is the set $\langle f_1, \ldots, f_t \rangle = \{\sum_{i=1}^t f_i h_i : h_i \in \mathbb{R}[x]\}$ – here, $f_1, \ldots, f_t$ and $h_1, \ldots, h_t$ are referred to as *generators* and *coefficients*, respectively. The *real variety* corresponding to polynomials $g_1, \ldots, g_r \in \mathbb{R}[x]$ is denoted $\mathcal{V}_{\mathbb{R}}(g_1, \ldots, g_r) = \{x \in \mathbb{R}^n : g_i(x) = 0,\ i = 1, \ldots, r\}$. Finally, the set of polynomials that can be expressed as a *sum of squares* of polynomials is denoted $\Sigma := \{p \in \mathbb{R}[x] : p = \sum_i p_i^2,\ p_i \in \mathbb{R}[x]\}$. We state next the *real Nullstellensatz* due to Krivine for certifying infeasibility of a system of a polynomial equations over $\mathbb{R}$:

Theorem 4 (Real Nullstellensatz)**.** *[75] Given any collection of polynomials $f_1, \ldots, f_t \in \mathbb{R}[x]$, we have that:*

$$-1 \in \Sigma + \langle f_1, \ldots, f_t \rangle \iff \mathcal{V}_{\mathbb{R}}(f_1, \ldots, f_t) = \emptyset.$$

Here $-1 \in \mathbb{R}[x]$ refers to the constant polynomial. The implication that $-1 \in \Sigma + \langle f_1, \ldots, f_t \rangle \Rightarrow \mathcal{V}_{\mathbb{R}}(f_1, \ldots, f_t) = \emptyset$ is straightforward. The reverse direction may be proved by appealing to Tarski's transfer principle. In general, the best-known bounds on the size of infeasibility certificates – i.e., the degrees of the

polynomials in $\Sigma, \langle f_1, \ldots, f_t \rangle$ that sum to -1 – are at least triply exponential. This is to be expected as many co-NP-hard problems can be reformulated as certifying infeasibility of a system of polynomial equations.

Obtaining *tractable relaxations* based on the real Nullstellensatz relies on three key observations. First, one fixes a subset $\tilde{\mathcal{I}} \subset \langle f_1, \ldots, f_t \rangle$ by considering polynomials $\sum_i h_i f_i$ in which the coefficients $h_i \in \mathbb{R}[x]$ have bounded degree (sets of the form $\tilde{\mathcal{I}}$ are sometimes called *truncated ideals*, although they are not formally ideals). In searching for infeasibility certificates of the form $-1 = p + q,\ p \in \Sigma, q \in \tilde{I}$, one can check that without loss of generality the search for p can also be restricted to sum-of-squares polynomials of bounded degree; formally, if every element of $\tilde{I}$ has degree at most $2d$, one can restrict the search to elements of Σ with degree at most $2d$. Second, a decomposition $-1 = p + \sum_i h_i f_i$ where the p and the $h_i's$ all have bounded degree is a linear constraint in the coefficients of p and the h_i's. Finally, checking that a polynomial $p \in \mathbb{R}[x]$ in n variables of degree at most $2d$ is an element of Σ can be formulated as a semidefinite feasibility problem; letting $m_{n,d}(x)$ denote the vector of all $\binom{n+d}{d}$ monomials in n variables of degree at most d, we have that:

$$p \in \Sigma \quad \Leftrightarrow \quad \exists P \in \mathbb{S}^{\binom{n+d}{d}},\ P \succeq 0,\ p(x) = m_{n,d}(x)'\ P\ m_{n,d}(x).$$

The relation $p(x) = m_{n,d}(x)'\ P\ m_{n,d}(x)$ is equivalent to a set of linear equations relating the entries of P to the coefficients of p. Thus, the search over a restricted family of infeasibility certificates via bounding the degree of the coefficients of the elements of $\langle f_1, \ldots, f_t \rangle$ is a semidefinite feasibility problem.

By considering a sequence of degree-bounded subsets $\mathcal{I}' \subset \mathcal{I}'' \subset \cdots \subset \langle f_1, \ldots, f_t \rangle$, one can search for more complex infeasibility certificates at the expense of solving increasingly larger semidefinite programs. Associated to this sequence of semidefinite programs is a sequence of *dual optimization problems* that provide successively tighter convex outer approximations to $\mathcal{V}_{\mathbb{R}}(f_1, \ldots, f_t)$ (assuming strong duality holds), i.e., $\mathcal{R}' \supseteq \mathcal{R}'' \supseteq \cdots \supseteq \text{conv}\,(\mathcal{V}_{\mathbb{R}}(f_1, \ldots, f_t))$. This dual perspective is especially interesting for attempting to identify elements of $\mathcal{V}_{\mathbb{R}}(f_1, \ldots, f_t)$. Concretely, fix a subset $\mathcal{I}' \subset \langle f_1, \ldots, f_t \rangle$, and suppose that the search for an infeasibility certificate of the form $-1 \in \Sigma + \mathcal{I}'$ is unsuccessful. Then $\mathcal{V}_{\mathbb{R}}(f_1, \ldots, f_t)$ may or may not be empty. At this stage, one can attempt to find an element of $\mathcal{V}_{\mathbb{R}}(f_1, \ldots, f_t)$ by optimizing a random linear functional over the set $\mathcal{R}'$ (obtained by considering the dual problem associated

to the system $-1 \in \Sigma + \mathcal{I}'$), and checking whether the resulting optimal solution lies in $\mathcal{V}_{\mathbb{R}}(f_1, \ldots, f_t)$; this heuristic is natural as $\mathcal{R}' \supseteq \text{conv}\,(\mathcal{V}_{\mathbb{R}}(f_1, \ldots, f_t))$, and if these sets were equal then the heuristic would generically succeed at identifying an element of $\mathcal{V}_{\mathbb{R}}(f_1, \ldots, f_t)$. If this approach to finding a solution is also unsuccessful, one can consider a larger subset $\mathcal{I}'' \subset \langle f_1, \ldots, f_t \rangle$ and an associated tighter approximation $\mathcal{R}'' \supseteq \text{conv}\,(\mathcal{V}_{\mathbb{R}}(f_1, \ldots, f_t))$ (here $\mathcal{I}' \subset \mathcal{I}''$ and $\mathcal{R}' \supseteq \mathcal{R}''$), and repeat the above procedure at a greater computational expense.

In Sections 4.2 and 4.2, we employ the methodology described above to give concrete descriptions of two convex outer approximations of the variety specified by the system $\mathcal{S}_{\text{iep}}$ associated to the affine IEP.

A First Semidefinite Relaxation

As our first example, we consider the following truncated ideal associated to the system $\mathcal{S}_{\text{iep}}$:

$$\mathcal{I}_1 = \Big\{ \text{Tr}\,(h_1 f_1) + \sum_{i=1}^{q} \left[h_2^{(i)} f_2^{(i)} + \text{Tr}\left(h_3^{(i)} f_3^{(i)} \right) \right] + \sum_{k=1}^{\ell} h_4^{(k)} f_4^{(k)} : h_2^{(i)}, h_4^{(k)} \in \mathbb{R}, \\ h_1, h_3^{(i)} \in \mathbb{S}^n \ \forall i, k h_1, h_2^{(i)}, h_3^{(i)}, h_4^{(k)} \text{ do not depend on } Z_i \Big\}. \tag{4.3}$$

In words, the truncated ideal $\mathcal{I}_1 \subset \langle \mathcal{S}_{\text{iep}} \rangle$ is obtained by restricting the coefficients to be constant polynomials (i.e., degree-zero polynomials). As a result, the elements of $\mathcal{I}_1$ consist of polynomials with degree at most two. Consequently, in searching for infeasibility certificates of the form $-1 \in \mathcal{I}_1 + \Sigma$ one need only consider quadratic polynomials in Σ, which in turn leads to checking feasibility of the following semidefinite program:

$$\begin{aligned} -\text{Tr}\,(A) - \sum_{i=1}^{q} m_i d_i - \sum_{k=1}^{\ell} b_k \xi_k &= 1; \\ A + d_i I + \lambda_i \sum_{k=1}^{\ell} \xi_k C_k - B_{ii} &= 0, \\ B_{ii} \succeq 0, \ i = 1, \ldots, q & \end{aligned} \tag{4.4}$$

in variables $A \in \mathbb{S}^n$, $d_i \in \mathbb{R}$ and $B_{ii} \in \mathbb{S}^n$ for $i = 1, \ldots, q$, and $\xi_k \in \mathbb{R}$ for $k = 1, \ldots, \ell$. The elements of the truncated ideal $\mathcal{I}_1$ can be associated to the above problem via the relations $h_1 = -A$, $h_2^{(i)} = -d_i$, $h_3^{(i)} = -B_{ii}$, $h_4^{(k)} = -\xi_k$, and then observing that the constraints in (4.4) are equivalent to checking that

the polynomial $\operatorname{Tr}(h_1 f_1) + \sum_{i=1}^{q} h_2^{(i)} f_2^{(i)} + \sum_{i=1}^{q} \operatorname{Tr}\left(h_3^{(i)} f_3^{(i)}\right) + \sum_{k=1}^{\ell} f_4^{(k)} h_4^{(k)}$ in variables $(Z_1, \ldots, Z_q)$ can be decomposed as $-1 - \Sigma$. Next we relate $\mathcal{R}_1(\Lambda, \mathcal{E})$ and the system $-1 \in \mathcal{I}_1 + \Sigma$ via strong duality:

Proposition 37. *Consider an affine IEP specified by a spectrum Λ and an affine space $\mathcal{E} \subset \mathbb{S}^n$. Let $\mathcal{I}_1$ be defined as in* (4.3) *and $\mathcal{R}_1(\Lambda, \mathcal{E})$ as in* (4.2)*. Then exactly one of the following two statements is true:*

$$(1)\ \mathcal{R}_1(\Lambda, \mathcal{E}) \text{ is nonempty}, \qquad\qquad (2)\ -1 \in \mathcal{I}_1 + \Sigma.$$

Proof. The feasibility of the system (4.4) is equivalent to the condition $-1 \in \mathcal{I}_1 + \Sigma$. One can check that the system (4.4) and the constraints describing $\mathcal{R}_1(\Lambda, \mathcal{E})$ are *strong alternatives* of each other, which follows from an application of conic duality – strong duality follows from Slater's condition being satisfied. □

As a consequence of this result, it follows that $\mathcal{R}_1(\Lambda, \mathcal{E})$ is a convex outer approximation of the variety $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$. A more direct way to see this is to consider any element $(Z_1, \ldots, Z_q) \in \mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$ and to note that the idempotence constraints $Z_i^2 - Z_i = 0$ in $\mathcal{S}_{\text{iep}}$ imply the semidefinite constraints $Z_i \succeq 0$ in $\mathcal{R}_1(\Lambda, \mathcal{E})$.

The set $\mathcal{R}_1(\Lambda, \mathcal{E})$ is closely related to the Schur-Horn orbitope associated to the spectrum Λ [107]:

$$\begin{aligned} \operatorname{SH}(\Lambda) &= \operatorname{conv}\{M \in \mathbb{S}^n \mid \lambda(M) = \Lambda\} \\ &= \Big\{ X \in \mathbb{S}^n \mid \exists (Z_1, \ldots, Z_q) \in \otimes^q \mathbb{S}^n \text{ s.t. } \sum_{i=1}^{q} Z_i = I; \\ &\qquad \operatorname{Tr}(Z_i) = m_i,\ Z_i \succeq 0\ \forall i;\ X = \sum_{i=1}^{q} \lambda_i Z_i \Big\}. \end{aligned} \tag{4.5}$$

The second equality follows from the characterization in [43]. The Schur-Horn orbitope was so-named by the authors of [107] due to its connection with the Schur-Horn theorem. A subset of the authors of the present note employed the Schur-Horn orbitope in developing efficient convex relaxations for NP-hard combinatorial optimization problems such as finding planted subgraphs [26] and computing edit distances between pairs of graphs [27]. In the context of the present note, the Schur-Horn orbitope provides a precise characterization

of the conditions under which $-1 \in \mathcal{I}_1 + \Sigma$ is successful. Specifically, from Proposition 37 and (4.5) we have that:

$$-1 \in \mathcal{I}_1 + \Sigma \quad \Leftrightarrow \quad \mathcal{R}_1(\Lambda, \mathcal{E}) = \emptyset \quad \Leftrightarrow \quad \mathrm{SH}(\Lambda) \cap \mathcal{E} = \emptyset. \tag{4.6}$$

Hence, if $-1 \notin \mathcal{I}_1 + \Sigma$, we have that $\mathcal{R}_1(\Lambda, \mathcal{E}) \neq \emptyset$. In particular, the variety $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\mathrm{iep}})$ may or may not be empty. At this stage, as discussed in Section 4.2 one can maximize a random linear functional over the set $\mathcal{R}_1(\Lambda, \mathcal{E})$; the resulting optimal solution $(\hat{Z}_1, \ldots, \hat{Z}_q)$ is generically an extreme point of $\mathcal{R}_1(\Lambda, \mathcal{E})$, and one can check if $(\hat{Z}_1, \ldots, \hat{Z}_q)$ satisfies the equations in the system $\mathcal{S}_{\mathrm{iep}}$. If this attempt at finding a feasible point in $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\mathrm{iep}})$ is unsuccessful, one can repeat the preceding steps at attempting to certify infeasibility or to find a feasible point in $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\mathrm{iep}})$ via a larger semidefinite program, which we describe in the next subsection.

We present here a result on guaranteed recovery of a solution to an affine IEP when the affine space is a random subspace:

Proposition 38. *Let $X^\star \in \mathbb{S}^n$ have a spectrum Λ with n distinct eigenvalues, and suppose $\mathcal{E} = \{X \mid \mathrm{Tr}(C_k X) = \mathrm{Tr}(C_k X^\star),\ k = 1, \ldots, \ell\}$ is an affine space with the $C_k \in \mathbb{S}^n$ being random matrices with i.i.d. standard Gaussian entries. If $\ell > \binom{n+1}{2} - H_n$ where $H_n = \sum_{j=1}^{n} \frac{1}{j}$ is the n'th harmonic number, then with high probability the unique element of $\mathcal{R}_1(\Lambda, \mathcal{E})$ is the set of n projection maps onto the eigenspaces of $X^\star$. (Recall that $\log(n) \leq H_n \leq \log(n) + 1$.)*

Proof. As Gaussian random matrices constitute an orthogonally invariant ensemble, we can assume without loss of generality that $X^\star$ is a diagonal matrix with the eigenvalues in descending order on the diagonal. From (4.2), (4.5), and (4.6), we need to ensure that $\mathrm{SH}(\Lambda) \cap \mathcal{E} = \{X^\star\}$. From the results in [4, 31], we have that if η is the expected value of the square of the Euclidean distance of a Gaussian random matrix to the normal cone at $X^\star$ with respect to $\mathrm{SH}(\Lambda)$, then $\mathrm{SH}(\Lambda) \cap \mathcal{E} = \{X^\star\}$ with high probability provided $\ell > \eta$. From [26] we have that the normal cone at $X^\star$ with respect to $\mathrm{SH}(\Lambda)$ is the set of diagonal matrices with the diagonal entries sorted in descending order. From [4] we have that the expected squared Euclidean distance of a random Gaussian matrix to such a cone of sorted entries is equal to $\binom{n+1}{2} - H_n$. $\square$

A Tighter Semidefinite Relaxation

Next we consider a larger truncated ideal $\mathcal{I}_2 \subset \langle \mathcal{S}_{\text{iep}} \rangle$ with larger degree coefficients than in $\mathcal{I}_1$:

$$\begin{aligned}\mathcal{I}_2 = \Big\{ &\operatorname{Tr}\left(h_1\left(Z_1,\ldots,Z_q\right) f_1\right) + \left[\sum_{i=1}^{q} h_2^{(i)}\left(Z_1,\ldots,Z_q\right) f_2^{(i)} + \operatorname{Tr}\left(h_3^{(i)} f_3^{(i)}\right)\right] + \\ &\sum_{k=1}^{l} f_4^{(k)} h_4^{(k)}\left(Z_1,\ldots,Z_q\right) : h_1\left(Z_1,\ldots,Z_q\right),\ h_3^{(i)} \in \mathbb{S}^n,\ h_2^{(i)}\left(Z_1,\ldots,Z_q\right), \\ &h_4^{(k)}\left(Z_1,\ldots,Z_q\right) \in \mathbb{R},\ \forall i,k,\ h_1\left(Z_1,\ldots,Z_q\right),\ h_2^{(i)}\left(Z_1,\ldots,Z_q\right), \\ &h_4^{(k)}\left(Z_1,\ldots,Z_q\right) \text{ affine in } Z_i,\ h_3^{(i)} \text{ does not depend on } Z_i \Big\}.\end{aligned} \tag{4.7}$$

The coefficients $h_3^{(i)}$ are constrained in the same way as in $\mathcal{I}_1$ but the other coefficients $h_1, h_2^{(i)}, h_4^{(k)}$ are allowed to be affine polynomials (in the case of h_1, more precisely a matrix of affine polynomials). The resulting collection $\mathcal{I}_2$ consists of polynomials of degree at most two, and therefore we can restrict our attention to elements of Σ of degree at most two in searching for infeasibility certificates of the form $-1 \in \mathcal{I}_2 + \Sigma$. However, $\mathcal{I}_2$ is in general larger than $\mathcal{I}_1$ so that $\mathcal{I}_2 + \Sigma$ offers a richer family of infeasibility certificates than $\mathcal{I}_1 + \Sigma$. The convex relaxation $\mathcal{R}_2(\Lambda, \mathcal{E})$ obtained as an alternative to the system $-1 \in \mathcal{I}_2 + \Sigma$ in turn provides a tighter approximation in general than $\mathcal{R}_1(\Lambda, \mathcal{E})$ to the convex hull $\operatorname{conv}(\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}}))$. We require some notation to give a precise description of $\mathcal{R}_2(\Lambda, \mathcal{E})$. Let $\delta_{k,l}$ denote the usual delta function which equals one if the arguments are equal and zero otherwise. Additionally, for $s, t = 1, \ldots, n$ let

$$f_{s,t} = \begin{cases} e_s e_t^T, & \text{if } s = t, \\ \frac{1}{2}(e_s e_t^T + e_t e_s^T), & \text{otherwise.} \end{cases}$$

Here, $e_s, e_t \in \mathbb{R}^n$ are the s'th and t'th standard basis vectors in $\mathbb{R}^n$. The set $\mathcal{R}_2(\Lambda, \mathcal{E})$ is then specified as:

$$
\begin{aligned}
&\mathcal{R}_2(\Lambda, \mathcal{E}) = \\
&\Big\{(Z_1, \ldots, Z_q) \in \otimes^q \mathbb{S}^n \mid \exists\, \mathcal{W}_{i,j} : \mathbb{S}^n \to \mathbb{S}^n, i, j = 1, \ldots, q,\ \exists\, \mathfrak{W} : \otimes^q \mathbb{S}^n \to \otimes^q \mathbb{S}^n, \\
&\mathfrak{W} \succeq 0,\ [\mathfrak{W}(X_1, \ldots, X_q)]_i = \sum_{j=1}^{q} \mathcal{W}_{i,j}(X_j)\ i = 1, \ldots, q,\ Z_i \succeq 0\ i = 1, \ldots, q \\
&\sum_{i=1}^{q} Z_i = I,\ \mathrm{Tr}\,(Z_i) = m_i,\ i = 1, \ldots, q,\ \sum_{i=1}^{q} \lambda_i \mathrm{Tr}\,(Z_i C_k) = b_k, k = 1, \ldots, l, \\
&\sum_{j=1}^{q} \mathcal{W}_{i,j}(f_{s,t}) = \delta_{s,t} Z_i,\ s, t = 1, \ldots, n,\ i = 1, \ldots, q, \\
&\sum_{s=1}^{n} \mathcal{W}_{i,j}(f_{s,s}) = m_j Z_i,\ i, j = 1, \ldots, q, \\
&\sum_{r=1}^{n} \mathrm{Tr}\,(f_{s,r} \mathcal{W}_{i,i}(f_{t,r})) = (Z_i)_{s,t},\ i = 1, \ldots, q,\ s, t = 1, \ldots, n, \\
&\sum_{j=1}^{q} \lambda_j \mathcal{W}_{i,j}(C_k) = b_k Z_i,\ i = 1, \ldots, q,\ k = 1, \ldots, \ell\ \Big\}.
\end{aligned}
\tag{4.8}
$$

Our next result records the fact that $\mathcal{R}_2(\Lambda, \mathcal{E})$ does constitute a strong alternative for $-1 \in \mathcal{I}_2 + \Sigma$.

Proposition 39. *Consider an affine IEP specified by a spectrum Λ and an affine space $\mathcal{E} \subset \mathbb{S}^n$. Let $\mathcal{I}_2$ be defined as in* (4.7) *and $\mathcal{R}_2(\Lambda, \mathcal{E})$ as in* (4.8)*. Then exactly one of the following two statements is true:*

$$
\text{(1) } \mathcal{R}_2(\Lambda, \mathcal{E}) \text{ is nonempty,} \qquad\qquad \text{(2) } -1 \in \mathcal{I}_2 + \Sigma.
$$

Proof. The proof is identical to that of Proposition 37, and it follows from an application of conic duality. □

It is clear that $\mathcal{R}_1(\Lambda, \mathcal{E}) \supseteq \mathcal{R}_2(\Lambda, \mathcal{E})$ as the constraints defining $\mathcal{R}_2(\Lambda, \mathcal{E})$ are a superset of those defining $\mathcal{R}_1(\Lambda, \mathcal{E})$. Further, for any $(Z_1, \ldots, Z_q) \in \mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$, one can check that the constraints defining $\mathcal{R}_2(\Lambda, \mathcal{E})$ are satisfied by setting the linear operators $\mathcal{W}_{i,j}(X) = \mathrm{Tr}\,(Z_j X)\, Z_i\ \forall X \in \mathbb{S}^n$. Thus, there are additional *quadratic relations* among the Z_i's that are satisfied by the elements of $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$ and are implied by $\mathcal{R}_2(\Lambda, \mathcal{E})$, but are not captured by the set

$\mathcal{R}_1(\Lambda, \mathcal{E})$. This is the source of the improvement of the relaxation $\mathcal{R}_2(\Lambda, \mathcal{E})$ compared to $\mathcal{R}_1(\Lambda, \mathcal{E})$, although the improvement comes at the expense of solving a substantially larger semidefinite program. In particular, $\mathcal{R}_1(\Lambda, \mathcal{E})$ entails checking q semidefinite constraints on $n \times n$ real symmetric matrices, while the description of $\mathcal{R}_2(\Lambda, \mathcal{E})$ involves a semidefinite constraint on the operator $\mathfrak{W} : \otimes^q \mathbb{S}^n \to \otimes^q \mathbb{S}^n$, which is equivalent to stipulating that a $\binom{n+1}{2}q \times \binom{n+1}{2}q$ real symmetric matrix is positive semidefinite. Thus, optimizing over $\mathcal{R}_2(\Lambda, \mathcal{E})$ is much more computationally expensive than $\mathcal{R}_1(\Lambda, \mathcal{E})$.

4.3 Numerical Illustrations

Here we present experiments illustrating the performance of the relaxations $\mathcal{R}_1(\Lambda, \mathcal{E}), \mathcal{R}_2(\Lambda, \mathcal{E})$ on random problem instances and stylized instances arising in applications. Our results are obtained using the CVX parser [60, 61] and the SDPT3 solver [112]. Before describing these, we present an approach to strengthen the relaxation $\mathcal{R}_2(\Lambda, \mathcal{E})$ by adding valid constraints without increasing the size of the semidefinite inequality.

Strengthening the Relaxations

A prominent approach in the optimization literature for obtaining improved bounds on hard nonconvex problems is to add redundant constraints. The procedure presented in Section 4.2 of considering a sequence of truncated ideals $\mathcal{I}_1 \subseteq \mathcal{I}_2 \subseteq \cdots \subseteq \langle \mathcal{S}_{\text{iep}} \rangle$ is a systematic method to add valid constraints; in particular, the elements of $\mathcal{I}_1$ and $\mathcal{I}_2$ represent polynomials that vanish at all the points in $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$. As $\mathcal{I}_1 \subseteq \mathcal{I}_2$, the relaxation $\mathcal{R}_2(\Lambda, \mathcal{E})$ offers (in general) a tighter convex outer approximation of $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$ than $\mathcal{R}_1(\Lambda, \mathcal{E})$ as $\mathcal{R}_2(\Lambda, \mathcal{E})$ is derived from the incorporation of a larger collection of redundant constraints.

Here we present a simple alternative approach to adding redundant constraints for the affine IEP by augmenting the original system $\mathcal{S}_{\text{iep}}$ with additional polynomials that vanish on $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$, and which are not contained in the truncated ideals $\mathcal{I}_1, \mathcal{I}_2$. Specifically, we consider the following modified system of equations:

$$\mathcal{S}_{\text{iep}}^{+} = \mathcal{S}_{\text{iep}} \cup \{ Z_i Z_j, \ i, j = 1, \ldots, q, \ i \neq j \}. \tag{4.9}$$

The matrix equations $Z_i Z_j = 0$ are satisfied for $i \neq j$ by every solution of $\mathcal{S}_{\text{iep}}$ as a consequence of the vanishing of $f_1, f_2^{(i)}, f_3^{(i)}$. However, despite being of low degree, the matrix polynomials $Z_i Z_j, \ i \neq j$ are not contained in $\mathcal{I}_1, \mathcal{I}_2$. Consequently, incorporating these degree-two equations offers the prospect of

strengthening our relaxations without a significant additional computational expense. We define truncated ideals $\mathcal{I}_1^+, \mathcal{I}_2^+$ corresponding to $\mathcal{S}_{\text{iep}}^+$ in an identical fashion to $\mathcal{I}_1, \mathcal{I}_2$ by restricting the coefficients corresponding to the additional polynomials $Z_i Z_j,\ i \neq j$ to be matrices of constant polynomials (as in the restriction of the coefficients $h_3^{(i)}$ of $f_3^{(i)}$), with the coefficients of the other polynomials $f_1, f_2^{(i)}, f_3^{(i)}, f_4^{(k)}$ being as in $\mathcal{I}_1, \mathcal{I}_2$.

The semidefinite relaxation $\mathcal{R}_1^+(\Lambda, \mathcal{E})$ obtained as a strong alternative to the system $-1 \in \mathcal{I}_1^+ + \Sigma$ is identical to $\mathcal{R}_1(\Lambda, \mathcal{E})$, i.e., the additional redundant constraints do not strengthen the relaxation. However, the strong alternative to the system $-1 \in \mathcal{I}_2^+ + \Sigma$ leads to a convex outer approximation $\mathcal{R}_2^+(\Lambda, \mathcal{E})$ of $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$ that is in general tighter than $\mathcal{R}_2(\Lambda, \mathcal{E})$; in addition to all the constraints that define $\mathcal{R}_2(\Lambda, \mathcal{E})$ in (4.8), the set $\mathcal{R}_2^+(\Lambda, \mathcal{E})$ consists of the additional constraints $\sum_{r=1}^n \operatorname{Tr}(f_{s,r} \mathcal{W}_{i,j}(f_{t,r})) = 0,\ i, j = 1, \ldots, q,\ i \neq j,\ s, t = 1, \ldots, n$ on the variables $\mathcal{W}_{i,j}$. Thus, a notable feature of the relaxation $\mathcal{R}_2^+(\Lambda, \mathcal{E})$ is that it is of the same size as $\mathcal{R}_2(\Lambda, \mathcal{E})$, despite providing a tighter convex outer approximation in general to $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$. We demonstrate the merits of this relaxation in the numerical experiments in this section.

Experiments with Random Affine IEPs

We present two series of experiments on random problems instances in this subsection.

In the first set of results, we compare the relative power of the two relaxations described in Section 4.2 in certifying infeasibility, or from a dual viewpoint, in approximating $\operatorname{conv}(\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}}))$. To provide a visual illustration, we consider affine IEPs involving matrices in $\mathbb{S}^3$, with a desired spectrum of $\{-1, 0, 1\}$. We begin by considering an affine space defined by $\ell = 3$ random linear equations, i.e., $\mathcal{E} = \{X \in \mathbb{S}^3 \mid \operatorname{Tr}(C_k X) = 0,\ C_k \in \mathbb{S}^3,\ k = 1, \ldots, \ell\}$, where the C_k's have random entries. Given the spectrum (which fixes the trace) and the affine space $\mathcal{E}$, the solution set $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$ is constrained to lie in an affine space of dimension at most two in $\mathbb{S}^3$. We then set the entries X_{11}, X_{22} to fixed values in the range $[-1, 1]$, and check whether there exists a matrix in $\mathbb{S}^3$ with these values for X_{11}, X_{22} that can be expressed as $\sum_i \lambda_i Z_i$ for $(Z_1, Z_2, Z_3) \in \mathcal{R}_1(\Lambda, \mathcal{E})$ and for $(Z_1, Z_2, Z_3) \in \mathcal{R}_2(\Lambda, \mathcal{E})$. Figures 4.1a and 4.1b represent two different problem instances obtained by generating two affine spaces $\mathcal{E}$ as described above, and they illustrate graphically when the

relaxations succeed or fail at certifying infeasibility. Evidently, the relaxation $\mathcal{R}_2(\Lambda, \mathcal{E})$ is successful in certifying infeasibility over a larger range of values of X_{11}, X_{22} than the relaxation $\mathcal{R}_1(\Lambda, \mathcal{E})$, thus illustrating its increased power (at a greater computational expense). From a dual perspective, we have in both cases that $\mathcal{R}_1(\Lambda, \mathcal{E}) \supsetneq \mathcal{R}_2(\Lambda, \mathcal{E})$. In particular, the feasibility regions corresponding to $\mathcal{R}_1(\Lambda, \mathcal{E})$ and $\mathcal{R}_2(\Lambda, \mathcal{E})$ in Figures 4.1a and 4.1b represent the projections of these sets onto the (X_{11}, X_{22})-plane of $\mathbb{S}^3$. In each of the two settings, we maximized 1000 random linear functionals over $\mathcal{R}_2(\Lambda, \mathcal{E})$ and in all cases obtained an element of $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$. Consequently, it appears at least based on numerical evidence that $\mathcal{R}_2(\Lambda, \mathcal{E}) = \text{conv}(\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}}))$ in both examples. Figures 4.1c and 4.1d give two examples based on the same setup as above, but with $\ell = 2$ random linear equations defining the affine space $\mathcal{E}$. Here the dimension of the solution set $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$ is at most three, and the feasibility regions corresponding to $\mathcal{R}_1(\Lambda, \mathcal{E})$ and $\mathcal{R}_2(\Lambda, \mathcal{E})$ in Figures 4.1c and 4.1d represent two-dimensional projections (onto the (X_{11}, X_{22})-plane of $\mathbb{S}^3$) of these sets. As with the previous examples, we maximized 1000 random linear functionals over $\mathcal{R}_2(\Lambda, \mathcal{E})$ and in all cases obtained an element of $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$. Consequently, it again appears that $\mathcal{R}_2(\Lambda, \mathcal{E}) = \text{conv}(\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}}))$.

Next, we consider higher-dimensional examples in which the solution set $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$ is, by construction, non-empty. We assess the performance of our heuristic of maximizing random linear functionals over the approximations $\mathcal{R}_1(\Lambda, \mathcal{E}), \mathcal{R}_2(\Lambda, \mathcal{E}), \mathcal{R}_2^+(\Lambda, \mathcal{E})$ in recovering solutions to the underlying IEPs. Concretely, for each $\ell \in \{2, 4, 6, 8, 10\}$ and for a given spectrum Λ, we consider $C_k \in \mathbb{S}^n$ with random entries and we generate a random $X^\star \in \mathbb{S}^n$ with spectrum Λ; with these in hand, we set $\mathcal{E} = \{X \in \mathbb{S}^n \mid \text{Tr}\,(C_k X) = \text{Tr}\,(C_k X^\star)\,,\ C_k \in \mathbb{S}^n,\ k = 1, \ldots, \ell\}$. For each ℓ and for a given Λ, we consider 100 random problem instances and for each instance we maximize 100 random linear functionals over the sets $\mathcal{R}_1(\Lambda, \mathcal{E}), \mathcal{R}_2(\Lambda, \mathcal{E}), \mathcal{R}_2^+(\Lambda, \mathcal{E})$. Table 4.1 gives the average number of successes in obtaining elements of $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$ for three different choices of the spectrum Λ. As expected, the relaxations $\mathcal{R}_1(\Lambda, \mathcal{E}), \mathcal{R}_2(\Lambda, \mathcal{E}), \mathcal{R}_2^+(\Lambda, \mathcal{E})$ are successively better, with $\mathcal{R}_2^+(\Lambda, \mathcal{E})$ being particularly useful in cases in which the eigenvalues have small multiplicities.

Discrete Inverse Sturm-Liouville Problem

Next, we demonstrate an application of our framework to certify infeasibility of, or produce a solution to, the extensively studied discrete inverse

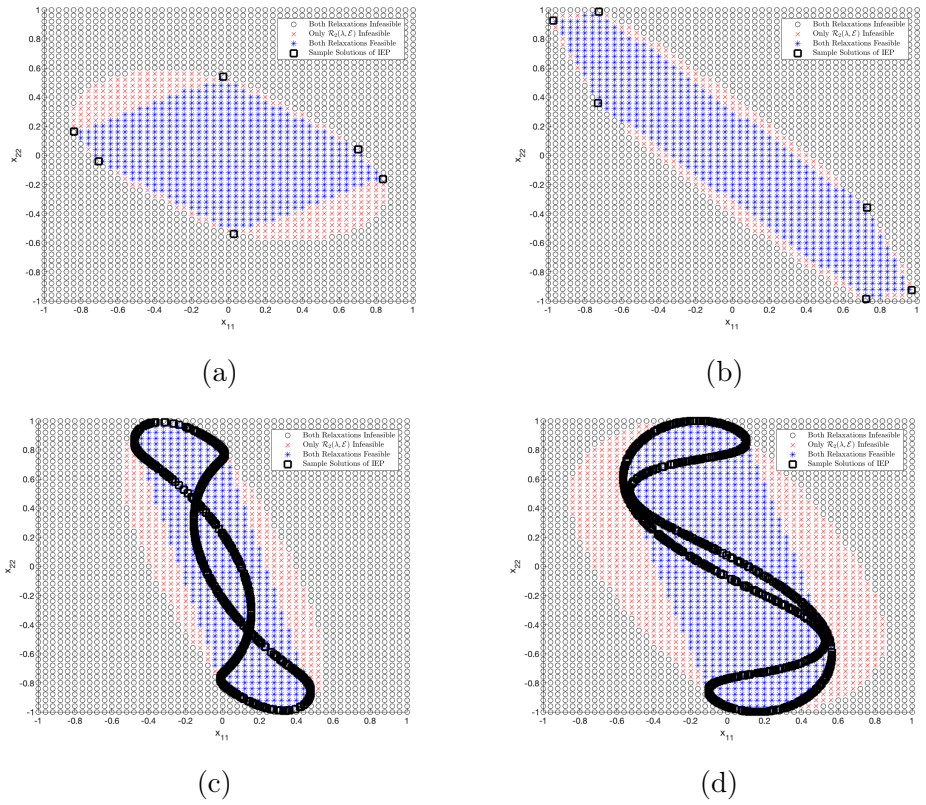


Figure 4.1: Comparison of feasible/infeasible regions of $\mathcal{R}_1(\Lambda, \mathcal{E})$ and $\mathcal{R}_2(\Lambda, \mathcal{E})$ for four random problem instances as described in Section 4.3. The points marked with black circles, red crosses, and blue stars correspond, respectively, to settings in which $\mathcal{R}_1(\Lambda, \mathcal{E})$ and $\mathcal{R}_2(\Lambda, \mathcal{E})$ are both infeasible; $\mathcal{R}_1(\Lambda, \mathcal{E})$ is feasible and $\mathcal{R}_2(\Lambda, \mathcal{E})$ is infeasible; and both $\mathcal{R}_1(\Lambda, \mathcal{E})$ and $\mathcal{R}_2(\Lambda, \mathcal{E})$ are feasible. Thick black squares represent (X_{11}, X_{22}) values of solutions to the affine IEP.

Sturm-Liouville problem [63, 82]. This problem arises as a discretization of a continuous differential boundary problem of the form $-u''(x) + p(x)u(x) = \lambda u(x),\ u(0) = u(\pi) = 0$. Here, $u(x)$ and $p(x)$ are functions, and λ is a parameter that is an eigenvalue of the system. A particular discretization of this differential equation gives rise to the linear system $\left(\frac{(n+1)^2}{\pi^2} J + D\right) u = \lambda u$, where J is a Jacobian matrix with diagonal entries equal to 2 and the nonzero off-diagonal entries equal to -1 [63]. Hence, given a collection $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, one wishes to identify a diagonal matrix D so that this linear system has a solution for each setting $\lambda = \lambda_i$, i.e., $\lambda_1, \ldots, \lambda_n$ are eigenvalues of the matrix $\frac{(n+1)^2}{\pi^2} J + D$. This is clearly an instance of an affine IEP.

We consider two different instantiations of the problem with $n = 5$. First, we consider the set of eigenvalues $\{1, 2, 3, 4, 5\}$. In this instance, there ex-

	$\mathcal{R}_1(\Lambda, \mathcal{E})$	$\mathcal{R}_2(\Lambda, \mathcal{E})$
l=2	94.11	98.91
l=4	69.28	92.56
l=6	37.03	79.52
l=8	12.81	54.36
l=10	2.59	27.05

(a)

	$\mathcal{R}_1(\Lambda, \mathcal{E})$	$\mathcal{R}_2(\Lambda, \mathcal{E})$
l=2	94.30	98.57
l=4	69.32	92.03
l=6	38.02	77.35
l=8	14.03	54.65
l=10	3.90	30.76

(b)

	$\mathcal{R}_1(\Lambda, \mathcal{E})$	$\mathcal{R}_2(\Lambda, \mathcal{E})$	$\mathcal{R}_2^+(\Lambda, \mathcal{E})$
l=2	95.64	97.57	99.68
l=4	75.95	85.97	97.77
l=6	45.87	61.95	91.83
l=8	18.00	30.48	72.18
l=10	3.43	8.28	41.18

(c)

Table 4.1: The average over 100 random problem instances and 100 random linear functionals for which our heuristic succeeds in identifying a feasible solution. The problem settings are (a) $n = 6,\ \Lambda = \{(-1, 2), (0, 2), (1, 2)\}$, (b) $n = 8, \Lambda = \{(-1, 4), (1, 4)\}$, and (c) $n = 5, \Lambda = \{(1, 1), (2, 1), (3, 1), (4, 1), (5, 1)\}$.

ists a decomposition $-1 \in \mathcal{I}_1 + \Sigma$ which certifies that the discrete inverse Sturm-Liouville problem is infeasible with the given spectrum. Next, we consider eigenvalues in the set $\{1, 4, 9, 16, 25\}$. In this case, the discrete inverse Sturm-Liouville problem turns out to be feasible. Specifically, we attempt to produce a solution to the inverse discrete Sturm-Lioville problem by maximizing 100 random linear functionals over the convex sets $\mathcal{R}_1(\Lambda, \mathcal{E})$, $\mathcal{R}_2(\Lambda, \mathcal{E})$, and $\mathcal{R}_2^+(\Lambda, \mathcal{E})$; our approach succeeds 14 out of 100 times over $\mathcal{R}_1(\Lambda, \mathcal{E})$, 26 out of 100 times over $\mathcal{R}_2(\Lambda, \mathcal{E})$, and 55 out of 100 times over $\mathcal{R}_2^+(\Lambda, \mathcal{E})$. These results suggest that our semidefinite relaxations may offer a useful solution framework across the range of applications in which the discrete inverse Sturm-Liouville problem arises.

Induced Subgraph Isomorphism

We present next the utility of our framework in the context of a problem in combinatorial optimization, namely the induced subgraph isomorphism problem. Here we are given two undirected, unweighted graphs $\mathcal{G}$ and $\mathcal{G}'$ on n and n' vertices, respectively, with $n' < n$. The problem is to determine whether $\mathcal{G}'$ is an induced subgraph of $\mathcal{G}$. This problem is NP-complete in general and has

received considerable attention [73].

Suppose $\mathcal{G}'$ is an induced subgraph of $\mathcal{G}$. Letting $A \in \mathbb{S}^n$ and $A' \in \mathbb{S}^{n'}$ be adjacency matrices representing the graphs $\mathcal{G}$ and $\mathcal{G}'$, respectively, such that A' is equal to a principal submatrix of A, there must exist a matrix $M \in \mathbb{S}^n$ that satisfies the following conditions:

$$\operatorname{Tr}(AM) = \sum_{i,j=1}^{n'} (A')_{i,j}\,; \quad (M)_{i,j} = 0 \text{ if } (A)_{i,j} = 0,\ i,j = 1,\ldots,n. \tag{4.10}$$

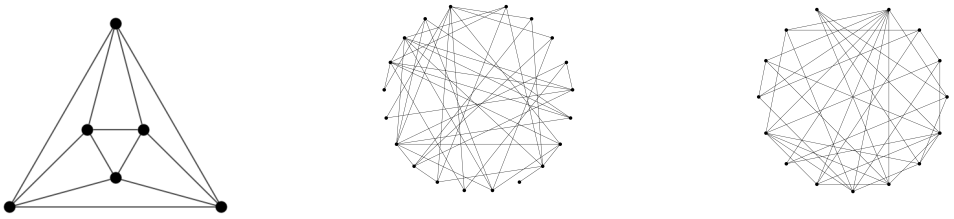
This consequence follows because we may choose M to be equal to A' on the $n' \times n'$ principal submatrix corresponding to corresponding to $\mathcal{G}'$ and zero elsewhere. Thus, a sufficient condition to certify that $\mathcal{G}'$ is not an induced subgraph of $\mathcal{G}$ is to certifying the infeasibility of an affine IEP in which the spectrum is equal to that of A' along with an eigenvalue of zero with multiplicity $n - n'$ and the affine space is given by (4.10).

With this approach, we prove that the octahedral graph with 6 nodes and 12 edges (shown in Figure 4.2a) is not contained as an induced subgraph in either of the larger graphs shown in Figure 4.2b (on 20 nodes with 44 edges) and Figure 4.2c (on 15 nodes with 38 edges). Both of these larger graphs are randomly generated Erdös-Renyi random graphs where any two vertices are independently and randomly connected with probability 0.2 for Figure 4.2b and 0.4 for Figure 4.2c. For the first graph, there exists a decomposition $-1 \in \mathcal{I}_1 + \Sigma$, thus certifying that the octahedral graph is not an induced subgraph. For the second graph, there is no infeasibility certificate of the form $-1 \in \mathcal{I}_1 + \Sigma$ but there is one of the form $-1 \in \mathcal{I}_2^+ + \Sigma$, thus providing a certificate that the octahedral graph is again not an induced subgraph.

Constructing a Real Symmetric Toeplitz Matrix with Desired Spectrum

Finally, we describe how our framework can be utilized for constructing real symmetric Toeplitz matrices with a desired spectrum. As Toeplitz matrices form a subspace, this question is an instance of an affine IEP. Landau showed that there exists a Toeplitz matrix with a desired spectrum, but his proof was non-constructive [77], and numerically constructing such matrices continues to remain a challenge.

In our first experiment, we set $n = 5$ and consider the problem of constructing a symmetric Toeplitz matrix with eigenvalues $\{1, 2, 3, 4, 5\}$. We maximize



(a) (b) (c)

Figure 4.2: From left to right: the octahedral graph, an Erdös-Renyi random graph on 20 nodes with p = 0.2, an Erdös-Renyi random graph on 15 nodes with p = 0.4. Our first convex relaxation certifies that the octahedral graph is not an induced subgraph of the graph shown in Figure 4.2b. A tighter convex relaxation proves the same result for the graph shown in Figure 4.2c.

random linear functionals over the sets $\mathcal{R}_1(\Lambda, \mathcal{E})$, $\mathcal{R}_2(\Lambda, \mathcal{E})$ and $\mathcal{R}_2^+(\Lambda, \mathcal{E})$, and we succeed at identifying a Toeplitz matrix with the desired spectrum 4 out of 100 times with $\mathcal{R}_1(\Lambda, \mathcal{E})$, 12 out of 100 times with $\mathcal{R}_2(\Lambda, \mathcal{E})$, and 41 out of 100 times with $\mathcal{R}_2^+(\Lambda, \mathcal{E})$. In our second experiment we set $n = 8$ and we seek a Toeplitz matrix with eigenvalues -1 (with multiplicity four) and 1 (with multiplicity four). With the same approach as before of maximizing random linear functionals, we identify a Toeplitz matrix with the desired spectrum 17 out of 100 times with $\mathcal{R}_1(\Lambda, \mathcal{E})$, and 84 out of 100 times with both $\mathcal{R}_2(\Lambda, \mathcal{E})$ and $\mathcal{R}_2^+(\Lambda, \mathcal{E})$. In summary, our framework provides a numerical counterpart to Landau's non-constructive existence result.

4.4 Conclusions

In this short note we describe a new framework for the affine IEP by first formulating it as a system of polynomial equations and then employing techniques from the polynomial optimization literature to obtain several semidefinite relaxations. These relaxations offer increasingly tighter approximations at the expense of solving larger semidefinite programming problems. We compare these relaxations both in random problem instances as well as in stylized examples in the context of various applications.

A number of future directions arise from our work. First, it is of interest to identify conditions on a spectrum Λ and an affine space $\mathcal{E}$ so that a particular relaxation such as $\mathcal{R}_1(\Lambda, \mathcal{E})$ is tight, i.e., $\mathcal{R}_1(\Lambda, \mathcal{E}) = \text{conv}(\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}}))$. These would correspond to families of instances of the affine IEP that are exactly solved by a tractable semidefinite program. A related second question is that

in considering a sequence of truncated ideals that are subsets of $\langle \mathcal{S}_{\text{iep}} \rangle$, does there exist a truncated ideal $\mathcal{I}$ of bounded (but possibly large) degree coefficients for which the alternative of the system $-1 \in \mathcal{I} + \Sigma$ is exactly equal to $\text{conv}(\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}}))$? Such a property is sometimes called *finite convergence* in the polynomial optimization literature; it has been shown to be true if $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$ is finite, but more generally, depends on the particular structure of the problem at hand. If this finite convergence property is true for the affine IEP setting considered in this note, then our heuristic for obtaining a solution to the system $\mathcal{S}_{\text{iep}}$ (if it is feasible) based on maximizing random linear functionals over convex outer approximations of $\mathcal{V}_{\mathbb{R}}(\mathcal{S}_{\text{iep}})$ is guaranteed to succeed after finitely many steps.

Chapter 5

CONCLUSIONS

In this section we highlight our main contributions and discuss some future research directions.

5.1 Summary of Contributions

Finding Planted Subgraphs with Few Eigenvalues using the Schur-Horn Relaxation

In Chapter 2 we study the planted subgraph problem, which is a fundamental problem in graph theory whose complexity is NP-hard in general. We introduce a new computationally tractable convex relaxation for it based on majorization inequalities on graph spectrum. We demonstrate that this procedure is particularly effective in finding planted subgraphs if their spectrum comprises of few distinct eigenvalues. Our approach generalizes earlier nuclear norm based convex optimization methods for identifying planted cliques. We demonstrate the utility of our framework via various numerical experiments conducted on various strongly regular graphs.

Convex Graph Invariant Relaxations For Graph Edit Distance

Graph edit distance is a prevalent metric of similarity between two graphs. Calculating the graph edit distance is NP-hard in general due to the underlying combinatorial setting. Building up on our ideas from Chapter 2, we introduce a family of tractable convex relaxations for exactly calculating, or providing lower bounds on the graph edit distance between two graphs. Our relaxations are based on convex graph invariants including the graph spectrum (as in Chapter 2), stability number and maximum-cut value. We present conditions in terms of certain graph parameters under which our relaxation based on the graph spectrum calculates the edit distance between two graphs exactly. We validate the usefulness of our method via numerical experiments on real and synthetic graph edit distance problems.

Sum of Squares Based Convex Relaxations for Inverse Eigenvalue Problems

We study the affine inverse eigenvalue problem, where the goal is to identify a matrix with a given spectrum contained in a given subspace. We express this problem in terms of polynomial equations and investigate convex relaxations for it arising from the earlier levels of the corresponding sum of squares hierarchy. The two convex relaxations we consider vary in their approximation qualities and associated computational costs. We demonstrate that our framework can be utilized for finding solutions for, or certifying infeasibility of the underlying inverse eigenvalue problem. We establish the utility of our framework by conducting numerical experiments on various instances of the affine inverse eigenvalue problem, such as discrete inverse Sturm-Liouville problem.

5.2 Future Directions

Identifying Additional Useful Convex Graph Invariants

In Chapters 2 and 3 we employ convex graph invariants based on graph spectrum, stability number and maximum-cut value for producing effective convex relaxations to two difficult problems arising from graphs. Evidently many other convex graph invariants may be useful in this context. Extending the list of suitable convex graph invariants would increase the applicability and approximation quality of relaxations arising from our framework.

Theoretical analysis of relaxations based on stability number and maximum-cut

In this thesis we extensively study optimality conditions pertaining to optimizing over the Schur-Horn orbitope, which is an invariant convex set based on the graph spectrum. Our analysis is based on Schur-Horn orbitope's geometric aspects, as we closely investigate the structure of its normal cones. However, our analysis for the relaxations based on the stability number and the maximum-cut invariants are mainly expository and limited in depth, since we do not have an as deep understanding of these invariants' geometric/facial structures. Obtaining a better understanding of the geometric properties of these constraint sets may facilitate producing theoretical results regarding the optimality conditions of the corresponding convex relaxations.

Extending our results to nonsymmetric matrices

In Chapters 2 and 3 we investigate problems arising from undirected graphs which can be represented by symmetric adjacency matrices. Similarly, in Chapter 4, our focus is solely limited on inverse eigenvalue problems arising from symmetric matrices. However, in various applications nonsymmetric matrices are of significant importance, and questions similar to the ones we have investigated in this thesis can be raised for their nonsymmetric counterparts. It may be possible to adapt some of the methods described in this thesis to answer such questions. For instance, in order to solve a particular "inverse singular value problem" on nonsymmetric $n \times n$ matrices, one might consider utilizing majorization inequalities on the eigenvalues of the $2n \times 2n$ matrix whose eigenvalues correspond to plus and minus of the desired singular values. Such approaches may enable extending the applicability of our framework from undirected graphs to directed graphs.

BIBLIOGRAPHY

[1] Zeina Abu-Aisheh et al. Graph edit distance contest: Results and future challenges. *Pattern Recognition Letters*, 100:96–103, 2017.

[2] Farid Alizadeh and Donald Goldfarb. Second-order cone programming. *Mathematical programming*, 95(1):3–51, 2003.

[3] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. *Random Structures and Algorithms*, 13(3-4):457–466, 1998.

[4] Dennis Amelunxen, Lotz Martin, Michael McCoy, and Joel Tropp. Living on the edge: Phase transitions in convex programs with random data. *Inf Inference: A Journal of the IMA*, 3(3):224–294, 2014.

[5] Brendan P. W. Ames and Stephen A. Vavasis. Nuclear norm minimization for the planted clique and biclique problems. *Mathematical Programming*, 129(1):69–89, 2011.

[6] Peter J. Artymiuk, Andrew R. Poirrette, Helen M. Grindley, David W. Rice, and Peter Willett. A graph-theoretic approach to the identification of three-dimensional patterns of amino acid side-chains in protein structures. *Journal of Molecular Biology*, 243(2):327–344, 1994.

[7] Zheng-Jian Bai, Delin Chu, and Defeng Sun. A dual optimization approach to inverse quadratic eigenvalue problems with partial eigenstructure. *SIAM Journal on Scientific Computing*, 29(6):2531–2561, 2007.

[8] Afonso S. Bandeira and Ramon van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4):2479–2506, 2016.

[9] Eiichi Bannai and Tatsuro Ito. *Algebraic Combinatorics*. Benjamin/Cummings, 1984.

[10] Victor Barcilon. On the multiplicity of solutions of the inverse problem for a vibrating beam. *SIAM Journal on Applied Mathematics*, 37(3): 605–613, 1979.

[11] Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, volume 2. SIAM, 2001.

[12] Abraham Berman and Robert J. Plemmons. *Nonnegative matrices in the mathematical sciences*, volume 9. Siam, 1994.

[13] Friedrich W. Biegler-König. A newton iteration process for inverse eigenvalue problems. *Numerische Mathematik*, 37(3):349–354, 1981.

[14] Grigoriy Blekherman, Pablo A. Parrilo, and Rekha R. Thomas. *Semidefinite optimization and convex algebraic geometry.* SIAM, 2012.

[15] Daniel Boley and Gene H. Golub. A survey of matrix inverse eigenvalue problems. *Inverse Problems*, 3(4):595, 1987.

[16] Raj C. Bose. Strongly regular graphs, partial geometries and partially balanced designs. *Pacific Journal of Mathematics*, 13(2):389–419, 1963.

[17] Sébastien Bougleux et al. Graph edit distance as a quadratic assignment problem. *Pattern Recognition Letters*, 87:38–46, 2017.

[18] Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge university press, 2004.

[19] W.G. Bridges and R.A. Mena. Multiplicative cones —- a family of three eigenvalue graphs. *Aequationes Mathematicae*, 22(1):208–214, 1981.

[20] Andries E. Brouwer and Willem H. Haemers. *Distance-regular graphs.* Springer, 2012.

[21] Andries E. Brouwer and Jacobus H. van Lint. Strongly regular graphs and partial geometries. *Enumeration and Design*, pages 85–122, 1984.

[22] Christopher I. Byrnes. Pole assignment by output feedback. In *Three decades of Mathematical System Theory*, pages 31–78. Springer, 1989.

[23] Peter J. Cameron. Strongly regular graphs. *Selected Topics in Graph Theory*, 1:pp–337, 1978.

[24] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6): 717–772, 2009.

[25] Emmanuel J Candes and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory*, 52(12):5406–5425, 2006.

[26] Utkan O. Candogan and Venkat Chandrasekaran. Finding planted subgraphs with few eigenvalues using the schur–horn relaxation. *SIAM Journal on Optimization*, 28(1):735–759, 2018.

[27] Utkan O. Candogan and Venkat Chandrasekaran. Convex graph invariant relaxations for graph edit distance. *arXiv preprint arXiv:1904.08934*, 2019.

[28] Utkan O. Candogan, Yong S. Soh, and Venkat Chandrasekaran. Sum of squares based convex relaxations for inverse eigenvalue problems. *arXiv preprint arXiv:1904.08934*, 2019.

[29] Venkat Chandrasekaran, Sujay Sanghavi, Pablo A. Parrilo, and Alan S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

[30] Venkat Chandrasekaran, Pablo A. Parrilo, and Alan S. Willsky. Convex graph invariants. *SIAM Review*, 54(3):513–541, 2012.

[31] Venkat Chandrasekaran, Benjamin Recht, Pablo Parrilo, and Alan Willsky. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6):805–849, 2012.

[32] Yudong Chen, Ali Jalali, Sujay Sanghavi, and Constantine Caramanis. Low-rank matrix recovery from errors and erasures. *IEEE Transactions on Information Theory*, 59(7):4324–4337, 2013.

[33] Moody T. Chu. Inverse eigenvalue problems. *SIAM Review*, 40(1):1–39, 1998.

[34] Moody T. Chu and Gene H. Golub. *Inverse eigenvalue problems: theory, algorithms, and applications*, volume 13. Oxford University Press, 2005.

[35] Donatello Conte, Pasquale Foggia, Carlo Sansone, and Mario Vento. Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(03):265–298, 2004.

[36] Jane Cullum, Wilm E. Donath, and Philip Wolfe. The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices. In *Nondifferentiable Optimization*, pages 35–55. Springer, 1975.

[37] Hua Dai. About an inverse eigenvalue problem arising in vibration analysis. *ESAIM: Mathematical Modelling and Numerical Analysis*, 29(4): 421–434, 1995.

[38] Évariste Daller, Sébastien Bougleux, Benoit Gaüzère, and Luc Brun. Approximate graph edit distance by several local searches in parallel. In *7th International Conference on Pattern Recognition Applications and Methods*, 2018.

[39] Carl De Boor and Gene H. Golub. The numerically stable reconstruction of a jacobi matrix from spectral data. *Linear Algebra and Its Applications*, 21(3):245–260, 1978.

[40] Graciano N. De Oliveira. On the multiplicative inverse eigenvalue problem. *Canadian Mathematical Bulletin*, 15(2):189–194, 1972.

[41] Graciano N. de Oliveria. Matrices with prescribed characteristic polynomial and several prescribed submatrices. *Linear and Multilinear Algebra*, 2(4):357–364, 1975.

[42] Philippe Delsarte and Y. Genin. Spectral properties of finite toeplitz matrices. In *Mathematical Theory of Networks and Systems*, pages 194–213. Springer, 1984.

[43] Yichuan Ding and Henry Wolkowicz. A low-dimensional semidefinite relaxation for the quadratic assignment problem. *Mathematics of Operations Research*, 34(4):1008–1022, 2009.

[44] Radu Dobrin, Qasim K. Beg, Albert-László Barabási, and Zoltán N. Oltvai. Aggregation of topological motifs in the escherichia coli transcriptional regulatory network. *BMC Bioinformatics*, 5(1):10, 2004.

[45] David L. Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell 1$ minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.

[46] Michael Doob. On characterizing certain graphs with four eigenvalues by their spectra. *Linear Algebra and its Applications*, 3(4):461–482, 1970.

[47] Michael Doob. Graphs with a small number of distinct eigenvalues. *Annals of the New York Academy of Sciences*, 175(1):104–110, 1970.

[48] Maryam Fazel. *Matrix rank minimization with applications*. PhD thesis, Stanford University, 2002.

[49] Uriel Feige and Robert Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Structures and Algorithms*, 16(2):195–208, 2000.

[50] Gerd Finke, Rainer E. Burkard, and Franz Rendl. Quadratic assignment problems. *Surveys in combinatorial optimization*, 132:61, 2011.

[51] Andreas Fischer, Ching Y. Suen, Volkmar Frinken, Kaspar Riesen, and Horst Bunke. A fast matching algorithm for graph-based handwriting recognition. In *International Workshop on Graph-Based Representations in Pattern Recognition*, pages 194–203. Springer, 2013.

[52] Roger Fletcher. A nonlinear programming problem in statistics (educational testing). *SIAM Journal on Scientific and Statistical Computing*, 2(3):257–267, 1981.

[53] Shmuel Friedland. Inverse eigenvalue problems. *Linear Algebra and Its Applications*, 17(1):15–51, 1977.

[54] Shmuel Friedland, Jorge Nocedal, and Michael L. Overton. The formulation and analysis of numerical methods for inverse eigenvalue problems. *SIAM Journal on Numerical Analysis*, 24(3):634–667, 1987.

[55] Michael R. Garey and David S. Johnson. *Computers and intractability*, volume 29. W.H. Freeman New York, 2002.

[56] Graham M.L. Gladwell. Inverse problems in vibration. *Applied Mechanics Reviews*, 39(7):1013–1018, 1986.

[57] P.W.M. Glaudemans and P.J. Brussard. *Shell-model applications in nuclear spectroscopy*. North Holland, Amsterdam, 1977.

[58] Chris Godsil. *Algebraic Combinatorics*, volume 6. CRC Press, 1993.

[59] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.

[60] Michael C. Grant and Stephen P. Boyd. CVX: Matlab software for disciplined convex programming, 2008.

[61] Michael C. Grant and Stephen P. Boyd. Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control*, pages 95–110. Springer, 2008.

[62] Willem H. Haemers and Vladimir D. Tonchev. Spreads in strongly regular graphs. *Designs, Codes and Cryptography*, 8(1-2):145–157, 1996.

[63] Ole H. Hald. Discrete inverse Sturm-Liouville problems. *Numerische Mathematik*, 27(2):249–256, 1977.

[64] Harry H. Harman. *Modern factor analysis*. University of Chicago Press, 1976.

[65] Peter E. Hart, Nils J. Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968.

[66] Daniel Hershkowitz. Existence of matrices with prescribed eigenvalues and entries. *Linear and Multilinear Algebra*, 14(4):315–342, 1983.

[67] Rashid Ibragimov, Maximilian Malek, Jiong Guo, and Jan Baumbach. GEDEVO: an evolutionary graph edit distance algorithm for biological network alignment. In *OASIcs-OpenAccess Series in Informatics*, volume 34. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2013.

[68] D. K. Isenor and Safwat G. Zaky. Fingerprint identification using graph matching. *Pattern Recognition*, 19(2):113–122, 1986.

[69] Tao Jiang, Guohui Lin, Bin Ma, and Kaizhong Zhang. A general edit distance between RNA structures. *Journal of Computational Biology*, 9 (2):371–388, 2002.

[70] Min-Jen Jou and Gerard J. Chang. The number of maximum independent sets in graphs. *Taiwanese Journal of Mathematics*, 4(4):685–695, 2000.

[71] Derek Justice and Alfred Hero. A binary linear programming formulation of the graph edit distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1200–1214, 2006.

[72] David Karger, David Karger, Rajeev Motwani, and Madhu Sudan. Approximate graph coloring by semidefinite programming. *Journal of the ACM (JACM)*, 45(2):246–265, 1998.

[73] Richard M. Karp. *Reducibility among combinatorial problems*. Springer, 1972.

[74] Michael Krivelevich and Benny Sudakov. Pseudo-random graphs. In *More Sets, Graphs and Numbers*, pages 199–262. Springer, 2006.

[75] Jean-Louis Krivine. Anneaux préordonnés. *Journal d'analyse Mathématique*, 12(1):307–326, 1964.

[76] Joseph B. Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and Its Applications*, 18(2):95–138, 1977.

[77] Henry J. Landau. The inverse eigenvalue problem for real symmetric toeplitz matrices. *Journal of the American Mathematical Society*, 7(3): 749–767, 1994.

[78] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[79] Jean B. Lasserre. Convex sets with semidefinite representation. *Math. Prog.*, 120(2):457–477, 2009.

[80] Marius Leordeanu, Martial Hebert, and Rahul Sukthankar. An integer projected fixed point method for graph matching and MAP inference. In *Advances in neural information processing systems*, pages 1114–1122, 2009.

[81] Jure Leskovec, Kevin J. Lang, and Michael Mahoney. Empirical comparison of algorithms for network community detection. In *Proceedings of the 19th international conference on World wide web*, pages 631–640. ACM, 2010.

[82] Boris M. Levitan. *Inverse Sturm-Liouville Problems*. Walter de Gruyter GmbH & Co KG, 2018.

[83] Matthew M. Lin, Bo Dong, and Moody T. Chu. Semi-definite programming techniques for structured quadratic inverse eigenvalue problems. *Numerical Algorithms*, 53(4):419–437, 2010.

[84] Zhi-Yong Liu and Hong Qiao. GNCCP-— graduated nonconvexity and concavity procedure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(6):1258–1267, 2014.

[85] Miguel Sousa Lobo, Lieven Vandenberghe, Stephen Boyd, and Hervé Lebret. Applications of second-order cone programming. *Linear algebra and its applications*, 284(1-3):193–228, 1998.

[86] László Lovász and Alexander Schrijver. Cones of matrices and set-functions and 0–1 optimization. *SIAM journal on optimization*, 1(2): 166–190, 1991.

[87] Si Wei Lu, Ying Ren, and Ching Y. Suen. Hierarchical attributed graph representation and recognition of handwritten chinese characters. *Pattern Recognition*, 24(7):617–632, 1991.

[88] Oliver Mason and Mark Verwoerd. Graph theory and networks in biology. *Systems Biology, IET*, 1(2):89–119, 2007.

[89] Vesna Memišević and Nataša Pržulj. C-GRAAL: Common-neighbors-based global GRAph ALignment of biological networks. *Integrative Biology*, 4(7):734–743, 2012.

[90] Nina Mishra, Robert Schreiber, Isabelle Stanton, and Robert E. Tarjan. Clustering social networks. In *Algorithms and Models for the Web-Graph*, pages 56–67. Springer, 2007.

[91] Theodore S. Motzkin and Ernst G. Straus. Maxima for graphs and a new proof of a theorem of turán. *Canad. J. Math*, 17(4):533–540, 1965.

[92] Mikhail Muzychuk and Mikhail Klin. On graphs with three eigenvalues. *Discrete Mathematics*, 189(1):191–207, 1998.

[93] Michel Neuhaus and Horst Bunke. An error-tolerant approximate matching algorithm for attributed planar graphs and its application to fingerprint classification. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 180–189. Springer, 2004.

[94] Michel Neuhaus and Horst Bunke. A graph matching based approach to fingerprint classification using directional variance. In *International Conference on Audio-and Video-Based Biometric Person Authentication*, pages 191–200. Springer, 2005.

[95] Michel Neuhaus and Horst Bunke. Edit distance-based kernel functions for structural pattern classification. *Pattern Recognition*, 39(10):1852–1863, 2006.

[96] Huajiang Ouyang, Dario Richiedei, Alberto Trevisani, and Gabriele Zanardo. Eigenstructure assignment in undamped vibrating systems: a convex-constrained modification method based on receptances. *Mechanical Systems and Signal Processing*, 27:397–409, 2012.

[97] Robert L. Parker. The inverse problem of electromagnetic induction: existence and construction of solutions based on incomplete data. *Journal of Geophysical Research: Solid Earth*, 85(B8):4421–4428, 1980.

[98] Pablo A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization.* PhD thesis, California Institute of Technology, 2000.

[99] Pablo A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96(2):293–320, 2003.

[100] Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2658–2663, 2004.

[101] Yitshak M. Ram and Grahamm M.L. Gladwell. Constructing a finite element model of a vibratory rod from eigendata. *Journal of Sound and Vibration*, 169(2):229–237, 1994.

[102] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

[103] Kaspar Riesen and Horst Bunke. Approximate graph edit distance computation by means of bipartite graph matching. *Image and Vision Computing*, 27(7):950–959, 2009.

[104] Kaspar Riesen, Andreas Fischer, and Horst Bunke. Computing upper and lower bounds of graph edit distance in cubic time. In *IAPR Workshop on Artificial Neural Networks in Pattern Recognition*, pages 129–140. Springer, 2014.

[105] Ralph T. Rockafellar. *Convex Analysis.* Princeton University Press, 2015.

[106] Alberto Sanfeliu and King-Sun Fu. A distance measure between attributed relational graphs for pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(3):353–362, 1983.

[107] Raman Sanyal, Frank Sottile, and Bernd Sturmfels. Orbitopes. *Mathematika*, 57(02):275–314, 2011.

[108] Johan J. Seidel. Strongly regular graphs. *Recent Progress in Combinatorics*, pages 185–198, 1969.

[109] Roded Sharan and Trey Ideker. Modeling cellular machinery through biological network comparison. *Nature Biotechnology*, 24(4):427, 2006.

[110] Roded Sharan et al. Conserved patterns of protein interaction in multiple species. *Proceedings of the National Academy of Sciences*, 102(6):1974–1979, 2005.

[111] Michael J Todd. The many facets of linear programming. *Mathematical Programming*, 91(3):417–436, 2002.

[112] Kim-Chuan Toh, Michael J. Todd, and Reha H. Tütüncü. SDPT3 —- a matlab software package for semidefinite programming, version 1.3. *Optimization Methods and Software*, 11(1-4):545–581, 1999.

[113] S. Toman and Josef M. Pliva. Multiplicity of solutions of the inverse secular problem. *Journal of Molecular Spectroscopy*, 21(1-4):362–371, 1966.

[114] Edwin R. van Dam. Regular graphs with four eigenvalues. *Linear Algebra and Its Applications*, 226:139–162, 1995.

[115] Edwin R. van Dam. *Graphs with few eigenvalues. An interplay between combinatorics and algebra.* PhD thesis, Tilburg University, School of Economics and Management, 1996.

[116] Edwin R. van Dam and Edward Spence. Small regular graphs with four eigenvalues. *Discrete Mathematics*, 189(1):233–257, 1998.

[117] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996.

[118] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing*, chapter 5, pages 210–268. Cambridge University Press, Cambridge, 2012.

[119] John von Neumann. Some matrix inequalities and metrization of matrix space. *Tomsk Univ. Rev*, 1(11):286–300, 1937.

[120] J.Y. Wang and B.S. Garbow. A numerical method for solving inverse real symmetric eigenvalue problems. *SIAM Journal on Scientific and Statistical Computing*, 4(1):45–51, 1983.

[121] Laurenz Wiskott, Jean-Marc Fellous, Norbert Krüger, and Christoph Von Der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 (7):775–779, 1997.

[122] Zhiping Zeng, Anthony K.H. Tung, Jianyong Wang, Jianhua Feng, and Lizhu Zhou. Comparing stars: On approximating graph edit distance. *Proceedings of the VLDB Endowment*, 2(1):25–36, 2009.

[123] Kang Zhao and Guozhu Yao. Application of the alternating direction method for an inverse monic quadratic eigenvalue problem. *Applied Mathematics and Computation*, 244:32–41, 2014.

[124] Fu-Zhao Zhou, Xi-Yan Hu, and Lei Zhang. The solvability conditions for the inverse eigenvalue problems of centro-symmetric matrices. *Linear Algebra and Its Applications*, 364:147–160, 2003.

A p p e n d i x A

APPENDIX FOR CHAPTER 2

In the appendix, we further investigate the connection between our method and Ames and Vavasis' nuclear norm minimization approach [5], for the special case of the planted k-clique problem. In order to identify a k-clique, Ames and Vavasis propose solving the following optimization problem:

$$\begin{aligned} \min_{A \in \mathbb{S}^n} \quad & \|A\|_* \\ \text{s.t.} \quad & \sum_{i,j \in \{1,\ldots,n\}} A_{i,j} \geq k^2, \\ & A_{i,j} = 0 \text{ if } (A_{\mathfrak{G}})_{i,j} = 0, \ i \neq j. \end{aligned} \tag{AV}$$

We establish the claim below:

Proposition 40. *Consider an instance of the planted k-clique problem. The Schur-Horn relaxation (P) succeeds in uniquely identifying the planted k-clique whenever the optimization program (AV) succeeds.*

Proof. We establish the result by presenting a sequence of optimization problems which relate the optimization problem (AV) to the Schur-Horn relaxation (P). Assume that the hidden k-clique (with 1's on the diagonal) is the unique optimal solution of (AV). Consider the first intermediate optimization problem below:

$$\begin{aligned} \max \quad & \sum_{i,j \in \{1,\ldots,n\}} A_{i,j} \\ \text{s.t.} \quad & \sum_{i,j \in \{1,\ldots,n\}} A_{i,j} \geq k^2, \\ & A_{i,j} = 0 \text{ if } (A_{\mathfrak{G}})_{i,j} = 0, \ i \neq j, \\ & \|A\|_* = k. \end{aligned} \tag{$I1$}$$

If the k-clique is the unique optimal solution of (AV), then $(I1)$ has a single feasible point. That is because of the additional constraint on the nuclear norm of the variable, and the fact that the nuclear norm of the k-clique (with 1's on the diagonal) is equal to k. Further, the only feasible point of $(I1)$ is the adjacency matrix of the planted clique.

Given that the optimization problem $(I1)$ is feasible, its first constraint is redundant, as that constraint and the objective function overlap. Moreover, one can replace the objective function with $\mathrm{Tr}(A \cdot A_{\mathfrak{G}})$, since by the planted model, $A_{\mathfrak{G}}$ is equal to 1 on every index where A is equal to 1, i.e., $\sum_{i,j \in \{1,\ldots,n\}} A_{i,j} = \mathrm{Tr}(A \cdot 1_k 1_k^T) = \mathrm{Tr}(A \cdot A_{\mathfrak{G}})$. These modifications lead to the second intermediate optimization problem given below:

$$\begin{aligned} \max \quad & \mathrm{Tr}(A \cdot A_{\mathfrak{G}}) \\ \text{s.t.} \quad & A_{i,j} = 0 \text{ if } (A_{\mathfrak{G}})_{i,j} = 0,\ i \neq j, \\ & \|A\|_* = k. \end{aligned} \tag{$I2$}$$

The optimization problem $(I2)$ is feasible in a potentially bigger set than the optimization problem $(I1)$, but its unique optimal value is still attained by the same matrix – the adjacency matrix of the planted clique.

Now consider adding the constraint $A \succeq 0$ to the constraint set of the optimization problem $(I2)$. Since the adjacency matrix of the k-clique (including the 1's on the diagonal) satisfies this constraint, it is still the unique optimal solution of the resulting problem. Furthermore, under the positive semidefiniteness of A, one can replace the nuclear norm constraint $\|A\|_* = k$ with the trace constraint $\mathrm{Tr}(A) = k$. With this final change, we obtain exactly the Schur-Horn relaxation (P), where the Schur-Horn orbitope is as described in equation (2.3). $\square$

A p p e n d i x B

APPENDIX FOR CHAPTER 3

B.1 Proof of Lemma 29

Proof of Lemma 29. The optimality of E^* follows from a direct application of the KKT conditions. Below, we establish the uniqueness. For this result, we first introduce subspaces S_Ω and $S_\mathcal{T}$:

$$S_\Omega = \{M \mid \mathcal{P}_\Omega(M) = M\},$$
$$S_\mathcal{T} = \{M \mid \mathcal{P}_{ii}(M) = 0, \forall i \in \{1, \ldots, m\}\}.$$

First, we show that if $\xi(\alpha, d, \mathcal{G}) < 1$, then $S_\Omega \cap S_\mathcal{T} = \{0\}$. Let $M \in S_\Omega \cap S_\mathcal{T}$. Then, it is easy to see that $[(I - \sum_{i=1}^m \alpha_i \mathcal{P}_{ii})\mathcal{P}_\Omega](M) = (I - \sum_{i=1}^m \alpha_i \mathcal{P}_{ii})(M) = M$ for any α. However, we have $\xi(\alpha, d, \mathcal{G}) < 1$ for some α. As such, $M = 0$.

Now, let T be a tangent direction from an extreme point X^* of $\mathcal{C}_{\mathcal{SH}(A_1)}$ into $\mathcal{C}_{\mathcal{SH}(A_1)}$, i.e., $X^* + T \in \mathcal{C}_{\mathcal{SH}(A_1)}$. Suppose that $E^* - T$ is another optimal solution of convex program (P). Given that Q is a subgradient of $\|\cdot\|_1$ at E^*, we have:

$$\|E^* - T\|_1 - \|E^*\| \geq \mathrm{Tr}(Q * (-T)). \tag{B.1}$$

Suppose that $\mathcal{P}_{ii}(T) \neq 0$ for some $i \in \{1, \ldots, m\}$. Consider some matrix $B = \mathcal{P}_{ii}(B)$. Since $Q \in \mathrm{relint}(\mathcal{N}_{\mathcal{SH}(A)}(A))$, we have that $Q + \varepsilon B \in \mathcal{N}_{\mathcal{SH}(A)}(A)$. As a result, $\mathrm{Tr}((Q+\varepsilon B)T) \leq 0 \Rightarrow \mathrm{Tr}(QT) < 0$; suggesting that $\|E^* - T\|_1 > \|E^*\|_1$ by equation (B.1). Therefore, we have that $T \in S_\mathcal{T}$, and $\mathrm{Tr}(QT) = 0$.

Now consider Q_E, let $\mathcal{P}_\Omega(Q_E) = E^*$ and $\mathcal{P}_{\Omega^c}(Q_E) = \mathrm{sign}(\mathcal{P}_{\Omega^c}(-T))$. Note that Q_E is a subgradient of $\|\cdot\|_1$ at E^*. Then:

$$\begin{aligned}
\mathrm{Tr}(Q_E(-T)) &= \mathrm{Tr}((\mathcal{P}_{\Omega^c}(Q_E) + Q - \mathcal{P}_{\Omega^c}(Q))(-T)) \\
&= \mathrm{Tr}((\mathcal{P}_{\Omega^c}(Q_E) - \mathcal{P}_{\Omega^c}(Q))(-T)) \\
&= \mathrm{Tr}((\mathcal{P}_{\Omega^c}(Q_E) - \mathcal{P}_{\Omega^c}(Q))\mathcal{P}_{\Omega^c}(-T)) \\
&= \|\mathcal{P}_{\Omega^c}(-T)\|_1 - \mathrm{Tr}(\mathcal{P}_{\Omega^c}(Q)\mathcal{P}_{\Omega^c}(-T)) \\
&\geq \|\mathcal{P}_{\Omega^c}(-T)\|_1 - \|\mathcal{P}_{\Omega^c}(-T)\|_1 \|\mathcal{P}_{\Omega^c}(Q)\|_\infty \\
&= \|\mathcal{P}_{\Omega^c}(-T)\|_1 (1 - \|\mathcal{P}_{\Omega^c}(Q)\|_\infty)
\end{aligned}$$

Since $\|\mathcal{P}_{\Omega^c}(Q)\|_\infty < 1$, we have that $\mathrm{Tr}(Q_E(-T)) > 0$ unless $\mathcal{P}_{\Omega^c}(-T) \neq 0$. However, $\mathrm{Tr}(Q_E(-T)) > 0$ would imply that $\|E^* - T\|_1 > \|E^*\|_1$ by the subgradient condition suggested by Q_E. Hence, $\mathcal{P}_{\Omega^c}(-T) = 0$, i.e., $T \in S_\Omega$. This suggests $T \in S_\Omega \cap S_\mathcal{T} = \{0\}$, indicating that E^* must be the unique optimal solution.

□