Thesis Proposal

On semidefinite program relaxations of quadratically constrained quadratic programs

Alex L. Wang

Computer Science Department
Carnegie Mellon University
alw1@cs.cmu.edu

Thesis Committee

Fatma Kılınç-Karzan (CMU, chair)
Samuel Burer (Univ. Iowa)
Pravesh Kothari (CMU)
Ryan O'Donnell (CMU)
Levent Tunçel (Univ. Waterloo)

This research is supported in part by NSF grant CMMI 1454548
and ONR grant N00014-19-1-2321.

Abstract

Quadratically constrained quadratic programs (QCQPs) are a fundamental class of optimization problems. In a QCQP, we are asked to minimize a (possibly nonconvex) quadratic function subject to a number of (possibly nonconvex) quadratic constraints. Such problems arise naturally in many areas of operations research, computer science, and engineering. Although QCQPs are NP-hard to solve in general, they admit a natural convex relaxation via the standard (Shor) semidefinite program (SDP) relaxation.

The research in this thesis is guided by two fundamental questions related to the SDP relaxation of a general QCQP: (1) What structures within a QCQP ensure that is SDP relaxation is accurate? And, (2) What structures within a QCQP allow its SDP relaxation to be solved efficiently?

In contrast to prior work on SDP relaxations of QCQPs (which has focused largely on *approximation* guarantees), we will interpret Question 1 as asking when *exactness* occurs in the SDP relaxation of a QCQP. In this direction, completed work in this thesis has developed a framework for understanding various forms of exactness: (i) *objective value exactness*—the condition that the optimal value of the QCQP and the optimal value of its SDP relaxation coincide, (ii) *convex hull exactness*—the condition that the convex hull of the QCQP epigraph coincides with the (projected) SDP epigraph, and (iii) the *rank-one generated* (ROG) property—the condition that a particular conic subset of the positive semidefinite matrices related to a given QCQP is generated by its rank-one matrices. Our analysis for objective value exactness and convex hull exactness stems from a geometric treatment of the projected SDP relaxation and crucially considers how the objective function interacts with the constraints. The ROG property complements these results by offering a sufficient condition for both objective value exactness and convex hull exactness which is oblivious to the objective function.

Question 2 seeks to understand when QCQPs (and/or their SDP relaxations) can be solved efficiently. In this direction, this thesis has investigated the generalized trust-region subproblem (GTRS) and structural properties of QCQPs that allow them to be diagonalized via novel transformations. Specifically, while the generalized trust-region subproblem (the class of QCQPs with a single constraint) is known to have an exact SDP relaxation, the relatively large computational complexity of SDP-based algorithms prevent them from being applied out-of-the-box to the GTRS. In this thesis, we develop and analyze a first-order algorithm for solving the GTRS whose running time is linear in the sparsity of the problem and whose dependence on the desired additive error is akin to that of accelerated gradient descent for smooth functions. Additional work in this direction has studied various notions of simultaneous diagonalizability of sets of quadratic forms. These new notions, specifically the almost SDC (ASDC) and d-restricted SDC (d-RSDC) properties, correspond to QCQPs that can be diagonalized after arbitrarily small perturbations of the QCQP data or the introduction of additional variables. Our investigation of this property will result in an explicit characterization of the ASDC property of pairs and nonsingular triples of quadratic forms as well as the conclusion that the 1-RSDC property holds generically for pairs of quadratic forms.

We will conclude by outlining ongoing and future work to be completed in this thesis. Specifically, we will discuss work on exactness in *random QCQPs* as well as first-order algorithms for solving the SDP relaxation of the GTRS and more general QCQPs under an *implicit regularity* assumption.

Contents

Chapter 1

Introduction

Convex optimization has been influential in shaping data science and modern computing. This subfield of optimization has found numerous applications in a variety of domains (e.g., machine learning, statistics, signal processing and engineering). Unfortunately, a growing number of interesting problems encountered by data scientists, engineers, and the scientific community at large are by nature *highly nonconvex*. Simultaneously, the convex optimization community has begun to investigate more "high-powered" machinery (e.g., semidefinite programs or the sum-of-squares hierarchy), much of which is at present considered impractical in large-scale applications.

This thesis attempts to address this divide by answering theoretical questions underpinning the practical application of tools from convex optimization (specifically, semidefinite programs) to interesting structured nonconvex problems (specifically, structured quadratically constrained quadratic programs). The goal of this thesis is to understand when certain nonconvex problems may be solved both *accurately* and *efficiently*.

Quadratically constrained quadratic programs (QCQPs) are a fundamental class of nonconvex optimization problems that naturally arise in operations research, engineering, and computer science; see [101] for additional applications of QCQPs. The ubiquity of this class of optimization problems stems from its expressiveness: any $\{0, 1\}$ integer program or polynomial optimization problem may be recast as a QCQP (see [8, 16, 81] and references therein).

It is well known that QCQPs are NP-hard to solve in general—indeed, the NP-hard combinatorial problem MAX-CUT can be readily recast as a QCQP. On the other hand, the standard (Shor) semidefinite program (SDP) relaxation offers a natural tractable convex relaxation for a general QCQP [89]. This convex relaxation is obtained by first reformulating the QCQP in a lifted space with an additional rank constraint and then dropping the rank constraint.

In passing from the nonconvex QCQP to its convex SDP relaxation, there are two important questions that must be addressed if SDPs are to be of practical importance in this setting:

Question 1. *What structures within a QCQP ensure that its SDP relaxation is* accurate*?*

Question 2. *What structures within a QCQP allow its SDP relaxation to be solved* efficiently*?*

In this thesis, the term "accurate" in Question 1 will be interpreted as "exact." This is in contrast to the extensive work on understanding the *approximation* quality of the SDP relaxation of QCQPs (e.g., [14, 45, 69, 75, 109]), which is highly important in its own right.

The two questions above constitute the two major thrusts of this thesis. In the following two sections, we present a short overview of completed work and research directions related to Questions 1 and 2. We defer more in-depth discussions of related work to the main body of this proposal.

1.1 Overview of completed work related to Question 1

On objective value exactness and convex hull exactness (Chapter 2)

In a series of papers [98, 100, 101], we introduce a framework for deriving sufficient conditions for *objective value exactness* (the condition that the optimal value of the QCQP coincides with the optimal value of its SDP relaxation) and *convex hull exactness* (the condition that the convex hull of the QCQP epigraph coincides with the projected SDP epigraph). Although convex hull exactness is a stronger property than objective value exactness, it is also more widely applicable. Specifically, results showing how to convexify common substructures in nonconvex problems are useful in building strong convex relaxations for more complicated nonconvex problems.

In [98, 101], we offer sufficient conditions for these two notions of exactness in the special setting where a dual object, the *cone of convex Lagrange multipliers*, is polyhedral. As an example, [101] shows that under this polyhedrality assumption, if the dimensions of particular eigenspaces related to a QCQP are large enough, then convex hull exactness holds. While this polyhedrality assumption is quite restrictive, this framework can already be used to recover and extend a number of results in the literature [23, 40, 52, 65, 99]. Follow-up work [100] then shows how to generalize the original framework to additionally handle settings where the cone of convex Lagrange multipliers is not necessarily polyhedral. These results can then be used to recover and generalize a number of known results, for example on quadratic matrix programming [10, 12].

On the rank-one-generated property (Chapter 3)

In many practical applications, the objective function of a QCQP may not be known exactly (for example, it may only be known to come from some class of objective functions or it may be evolving with time). To address this consideration, one may seek to understand QCQP feasible domains for which objective value exactness holds for *every* choice of objective function. This property is naturally related to the *rank-one-generated* (ROG) property of spectrahedra, which has further connections to statistics and sum-of-squares programming. For example, in statistics the ROG property has implications for maximum likelihood estimation in Gaussian graphical models [18] and positive semidefinite (PSD) matrix completion [3, 35, 47].

In [5], we study the facial structure of ROG *cones* and offer sufficient conditions for this property in the context of general conic subsets of the PSD cone. One of our main contributions shows that the sufficient conditions we provide give a complete characterization of the ROG cones defined by two constraints. This extends the only other setting (a single constraint and the well-known S-lemma) for which an explicit characterization of the ROG property is known. In follow-up work [60], we derive additional connections between the ROG property, the problem of minimizing *ratios* of quadratic functions over quadratically constrained domains, and variants of the S-lemma. As an example application of the ROG toolkit, we recover a known result regarding the SDP relaxation of the regularized total least squares problem (a common problem in signal recovery).

1.2 Overview of completed work related to Question 2

While semidefinite programs are extremely powerful and have been shown to address difficult problems, they have yet to see widespread adoption in practice. Notably, the task of solving an SDP is still considered *impractical*, especially in modern large-data settings. In this direction, this thesis has investigated properties of QCQPs that allow their SDP relaxations to be solved *efficiently* via specialized algorithms.

On the generalized trust-region subproblem (Chapter 4)

The generalized trust-region subproblem (GTRS) is the subclass of QCQPs containing exactly one quadratic constraint. This class has myriad applications in robust optimization, quadratic integer programming, and compressed sensing (see [99] and references therein). When the single constraint is additionally convex, the resulting problem is known as the trust-region subproblem (TRS) and is foundational in the area of nonlinear programming. Indeed, iterative algorithms based on the TRS (known sometimes as trust-region methods) are among the most empirically successful techniques for general nonlinear programs.

Recent work on the GTRS [2, 57, 58] has resulted in specialized algorithms for solving the SDP relaxation of the GTRS, which is known to be exact [107]. Building on this line of work, in [99], we derive a convex reformulation of the GTRS in the *original space* of variables (this contrasts with the SDP relaxation, which lives in a lifted space) and show how to efficiently solve this convex reformulation using accelerated gradient descent. The running time of the algorithm presented in [99] is comparable to the running time of a sparse minimum eigenvalue call and matches the running time of other specialized algorithms developed for the trust-region subproblem. For comparison, note that the minimum eigenvalue problem is an instance of the GTRS.

On diagonalizing quadratic forms (Chapter 5)

In contrast to the general setting, the SDP relaxation of a diagonal QCQP, i.e., a QCQP where the quadratic forms involved are diagonal is much more computationally tractable. Specifically, in this setting, the semidefinite constraint may be replaced by a product of second-order cone constraints. Furthermore, in relation to Question 1, the exactness of the SDP relaxation of a diagonal QCQP is better understood than that of a general QCQP [100]. A natural question from a computational perspective, then, is whether a given QCQP may be rewritten (possibly after a small perturbation or a lifting) as a diagonal QCQP.

Prior work in this direction [24, 56, 64, 77] has investigated sets of quadratic forms that are *simultaneously diagonalizable via congruence* (SDC). Given a QCQP for which the set of corresponding quadratic forms is SDC, one may perform a change of variables to get a diagonal QCQP. In [97], we extend this line of work by investigating the almost SDC (ASDC) sets and the d-restricted SDC (d-RSDC) sets. We say that a set of quadratic forms is ASDC if it is the limit of SDC sets and d-RSDC if it is the restriction of an SDC set in $n+d$ dimensions to an n-dimensional subspace. These definitions allow us to diagonalize a wider family of QCQPs: Given a QCQP for which the corresponding set of quadratic forms is ASDC (resp. d-RSDC), we may first perturb the QCQP by arbitrarily small amounts (resp. introduce d additional variables), then perform a change of variables to arrive at a diagonal QCQP. In this direction, we give a complete characterization of the ASDC pairs and the nonsingular ASDC triples. As an important corollary, our characterization of the ASDC pairs implies that *almost every* pair of quadratic forms is 1-RSDC. In particular, it is always possible to diagonalize the GTRS with additional linear constraints (a problem with applications, for example, in portfolio deleveraging [68]) after perturbing the constraints by arbitrarily small amounts and introducing a single additional variable. Preliminary numerical experiments suggest that our constructions significantly reduce the time needed for off-the-shelf optimization packages to solve such problems.

1.2.1 Notation

For nonnegative integers $m \leq n$, define $[n] \coloneqq \{1, \dots, n\}$ and $[m, n] \coloneqq \{m, m+1, \dots, n-1, n\}$. Let $\mathbb{R}_+$ denote the nonnegative reals. For $i \in [n]$, let $e_i \in \mathbb{R}^n$ denote the ith standard basis vector. Let $\mathbf{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ denote the $n-1$ sphere and let $\mathbf{B}(\bar{x}, r) = \{x \in \mathbb{R}^n : \|x - \bar{x}\| \leq r\}$ denote the n-ball with radius r and center $\bar{x}$. Let $\mathbb{S}^n$ denote the set of real symmetric $n \times n$ matrices and $\mathbb{S}^n_+$ the cone of positive semidefinite (PSD) matrices. We write $A \succeq 0$ (respectively, $A \succ 0$) if A is positive semidefinite (respectively, positive definite). Let $\lambda_{\min}(A)$ denote the minimum eigenvalue of A. Given $M \in \mathbb{R}^{m \times n}$, let $\operatorname{range}(M)$ and $\ker(M)$ denote the range and kernel of M respectively. When $m = n$, let $\operatorname{tr}(M)$ denote the trace of M. Let 0_n, $I_n \in \mathbb{S}^n$ denote the $n \times n$ zero matrix and identity matrix respectively; we will simply write 0 or I when the dimension is clear. We will also let $0_n \in \mathbb{R}^n$ denote the zero vector; whether 0 or 0_n is a scalar, vector, or matrix will be clear from context. We endow $\mathbb{S}^n$ with the inner product $\langle A, B \rangle \coloneqq \operatorname{tr}(A^\top B)$. Given W a subspace of $\mathbb{R}^n$ with dimension k, a surjective map $U : \mathbb{R}^k \to W$ and $A \in \mathbb{S}^n$, let A_W denote the restriction of A to W, i.e., $A_W = U^\top A U$. When U is inconsequential, we will omit specifying it. Given $u \in W$ and $v \in W^\perp$, let $u \oplus v$ denote their direct sum. For $A \in \mathbb{S}^n$ and $B \in \mathbb{S}^m$, let $A \oplus B \in \mathbb{S}^{n+m}$ and $A \otimes B \in \mathbb{S}^{nm}$ denote the direct sum and Kronecker product of A and B respectively. For a subset $\mathcal{D}$ of some Euclidean space (e.g., $\mathbb{R}^n$ or $\mathbb{S}^n$) let $\mathcal{D}^\circ$, $\operatorname{int}(\mathcal{D})$, $\operatorname{extr}(\mathcal{D})$, $\operatorname{cl}(\mathcal{D})$, $\operatorname{conv}(\mathcal{D})$, $\operatorname{clconv}(\mathcal{D})$, $\operatorname{cone}(\mathcal{D})$, $\operatorname{clcone}(\mathcal{D})$, $\operatorname{span}(\mathcal{D})$, $\operatorname{aff}(\mathcal{D})$, $\dim(\mathcal{D})$, $\operatorname{aff}\dim(\mathcal{D})$ and $\mathcal{D}^\perp$ denote the polar, interior, extreme points, closure, convex hull, closed convex hull, conic hull, closed conic hull, lineal hull, affine hull, dimension, affine dimension, and orthogonal complement of $\mathcal{D}$, respectively.

Chapter 2

Objective value and convex hull exactness

This chapter is based on joint work [98, 100, 101] with Fatma Kılınç-Karzan.

This chapter considers SDP relaxations of QCQPs and derives sufficient conditions on the QCQP data for various notions of exactness.

Additional notation. Given W a subspace of $\mathbb{R}^n$, let Π_W denote projection onto W. For a cone K, we will write $\mathcal{F} \trianglelefteq K$ to denote the fact that $\mathcal{F}$ is a face of K.

2.1 Related work

While there has been much work on understanding the *approximation* quality of the SDP relaxation of specific concrete classes of QCQPs (e.g., MAX-CUT [45]) and even for more *abstract* classes of QCQPs [14, 69, 75, 109], not much is known about when the SDP relaxation of a given QCQP is *exact* beyond simple settings (e.g., the case where the QCQP has a single constraint).

In this direction, Laurent and Poljak [63] showed that it is NP-hard to determine whether the SDP relaxation of a given QCQP has *objective value exactness*, i.e., whether the optimum objective value of the QCQP matches the optimum objective value of its SDP relaxation. Nevertheless, there has been recent interest on *sufficient* conditions for this property. One line of prior work has focused on the case where there are only a few (usually one or two) nonconvex quadratic functions in the QCQP. This research vein can be traced back to Yakubovich's S-procedure [40, 107] (also known as the S-lemma) and the work of Sturm and Zhang [93]. In particular, the classical trust region subproblem (TRS)—the problem of minimizing a nonconvex quadratic function over an ellipsoid—and its variants have attracted significant attention and cases under which an exact SDP reformulation is possible have been investigated; see the excellent survey by Burer [19] and references therein. For example, Jeyakumar and Li [55] showed that the standard SDP relaxation of the TRS with additional linear inequalities has objective value exactness under a condition regarding the dimension of the minimum generalized eigenspace.

A more recent line of research has focused on sufficient conditions for objective value exactness which do not make explicit assumptions on the number of nonconvex quadratic functions. As an example, Burer and Ye [23], Locatelli [67] establish sufficient conditions under which diagonal QCQPs (those QCQPs with diagonal quadratic forms) have this property. Objective value exactness has also been established for a special class of QCQPs known as quadratic matrix programs (QMPs) [10, 12]. A QMP is an optimization problem of the form

$$\inf_{X\in\mathbb{R}^{n\times k}} \left\{ \operatorname{tr}(X^\top \mathcal{A}_{\text{obj}} X) + 2\operatorname{tr}(B_{\text{obj}}^\top X) + c_{\text{obj}} : \begin{array}{r} \operatorname{tr}(X^\top \mathcal{A}_i X) + 2\operatorname{tr}(B_i^\top X) + c_i \leq 0, \\ \forall i \in [m] \end{array} \right\},$$

where $\mathcal{A}_i \in \mathbb{S}^n$, $B_i \in \mathbb{R}^{n\times k}$, and $c_i \in \mathbb{R}$, and arises often in robust least squares or as a result of Burer-Monteiro reformulations for rank-constrained SDPs [10, 21]. Note that by replacing the matrix variable $X \in \mathbb{R}^{n\times k}$ by the vector variable $x \in \mathbb{R}^{nk}$, we may reformulate a QMP as a QCQP with special structure. In this direction, Beck [10] showed that objective value exactness holds for this vectorized reformulation of a QMP as long as the number of total constraints, m, is at most the width of the matrix variable, k.

2.2 Preliminaries

Problem setup

In this proposal, we will restrict ourselves to considering *epigraphs* of QCQPs with only inequality constraints. See [100] for similar results on general quadratically constrained sets. For $n \in \mathbb{N}$, let $\mathbb{S}^n$ denote the set of $n \times n$ symmetric matrices. We will identify $\mathbb{S}^{n+1}$ with the space of quadratic functions on $\mathbb{R}^n$, i.e., given $M = \left(\begin{smallmatrix} A & b \\ b^\top & c\end{smallmatrix}\right) \in \mathbb{S}^{n+1}$ for $A \in \mathbb{S}^n$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$, define

$$q_M(x) \coloneqq \begin{pmatrix} x \\ 1\end{pmatrix}^\top \begin{pmatrix} A & b \\ b^\top & c\end{pmatrix} \begin{pmatrix} x \\ 1\end{pmatrix} = x^\top A x + 2b^\top x + c.$$

Given a finite subset $\mathcal{M} \coloneqq \{M_1, \dots, M_m\} \subseteq \mathbb{S}^{n+1}$, define $q_i(x) \coloneqq q_{M_i}(x)$ and set

$$\mathcal{X} \coloneqq \{x \in \mathbb{R}^n : \, q_i(x) \le 0,\, \forall i \in [m]\}.$$

Given an additional $M_{\mathrm{obj}} \in \mathbb{S}^{n+1}$, define $q_{\mathrm{obj}}(x) \coloneqq q_{M_{\mathrm{obj}}}(x)$ and consider the QCQP

$$\mathrm{Opt} \coloneqq \inf_{x\in\mathbb{R}^n} \{q_{\mathrm{obj}}(x) : \, x \in \mathcal{X}\} = \inf_{x\in\mathbb{R}^n} \left\{ \left\langle \begin{pmatrix} x \\ 1\end{pmatrix}\begin{pmatrix} x \\ 1\end{pmatrix}^\top, M_{\mathrm{obj}} \right\rangle : \; \left\langle \begin{pmatrix} x \\ 1\end{pmatrix}\begin{pmatrix} x \\ 1\end{pmatrix}^\top, M_i \right\rangle \le 0,\, \forall i \in [m] \right\}. \tag{2.1}$$

Let $\mathcal{D}$ denote the epigraph of this QCQP, i.e.,

$$\mathcal{D} \coloneqq \left\{ (x,t) \in \mathbb{R}^n \times \mathbb{R} : \; \begin{array}{l} q_{\mathrm{obj}}(x) \le t \\ x \in \mathcal{X} \end{array} \right\} = \left\{ (x,t) \in \mathbb{R}^n \times \mathbb{R} : \; \begin{array}{l} \left\langle \begin{pmatrix} x \\ 1\end{pmatrix}\begin{pmatrix} x \\ 1\end{pmatrix}^\top, M_{\mathrm{obj}} \right\rangle \le t \\ \left\langle \begin{pmatrix} x \\ 1\end{pmatrix}\begin{pmatrix} x \\ 1\end{pmatrix}^\top, M_i \right\rangle \le 0,\, \forall i \in [m] \end{array} \right\}. \tag{2.2}$$

In the following, it will be useful to also give names to the quadratic forms, linear forms, and constants in the quadratic functions $q_{\mathrm{obj}}(x)$ and $q_i(x)$. Let $A_{\mathrm{obj}} \in \mathbb{S}^n$, $b_{\mathrm{obj}} \in \mathbb{R}^n$ and $c_{\mathrm{obj}} \in \mathbb{R}$ such that $M_{\mathrm{obj}} = \left(\begin{smallmatrix} A_{\mathrm{obj}} & b_{\mathrm{obj}} \\ b_{\mathrm{obj}}^\top & c_{\mathrm{obj}}\end{smallmatrix}\right)$. Similarly define $A_i \in \mathbb{S}^n$, $b_i \in \mathbb{R}^n$, and $c_i \in \mathbb{R}$.

Our framework in this chapter revolves around the Lagrangian dual and aggregation. We thus introduce the following notation. Let $q : \mathbb{R}^n \to \mathbb{R}^m$ denote the vector-valued function with $q(x)_i = q_i(x)$. Define $A(\gamma) \coloneqq \sum_{i=1}^m \gamma_i A_i$. Similarly define $b(\gamma)$ and $c(\gamma)$. Finally, we will also introduce the following "projective" version of these aggregated quantities: Let $A[\gamma] \coloneqq A_{\mathrm{obj}} + A(\gamma)$ and similarly define $b[\gamma]$, $c[\gamma]$. Let $q[\gamma, x] \coloneqq q_{\mathrm{obj}}(x) + \langle \gamma, q(x)\rangle$.

Note that with these definitions

$$q[\gamma, x] = x^\top A[\gamma] x + 2b[\gamma]^\top x + c[\gamma] = q_{\mathrm{obj}}(x) + \langle \gamma, q(x)\rangle = q_{\mathrm{obj}}(x) + x^\top A(\gamma) x + 2b(\gamma)^\top x + c(\gamma).$$

The projected semidefinite programming relaxation

A natural convex relaxation of $\mathcal{D}$ is given by the standard (Shor) semidefinite programming (SDP) relaxation. To simplify our notation, given an arbitrary set $\mathcal{M} \subseteq \mathbb{S}^{n+1}$, we define

$$\mathcal{S}(\mathcal{M}) \coloneqq \left\{ Z \in \mathbb{S}^{n+1} : \; \begin{array}{l} \langle M, Z\rangle \le 0,\, \forall M \in \mathcal{M} \\ Z \succeq 0 \end{array} \right\}. \tag{2.3}$$

Note that $\mathcal{S}(\mathcal{M})$ is a closed convex cone. We will revisit this set and its properties in Chapter 3.

Definition 1. The *projected SDP relaxation of the QCQP* (2.1) is

$$\mathrm{Opt}_{\mathrm{SDP}} \coloneqq \inf_{x\in\mathbb{R}^n} \left\{ \langle M_{\mathrm{obj}}, Z\rangle : \begin{array}{l} \exists Z = \begin{pmatrix} X & x \\ x^\top & 1 \end{pmatrix} \in \mathbb{S}^{n+1} : \\ Z \in \mathcal{S}(\mathcal{M}) \end{array} \right\}. \tag{2.4}$$

The *projected SDP relaxation of* $\mathcal{D}$ is

$$\mathcal{D}_{\mathrm{SDP}} \coloneqq \left\{ (x,t) \in \mathbb{R}^n \times \mathbb{R} : \begin{array}{l} \exists Z = \begin{pmatrix} X & x \\ x^\top & 1 \end{pmatrix} \in \mathbb{S}^{n+1} : \\ \langle M_{\mathrm{obj}}, Z\rangle \le t \\ Z \in \mathcal{S}(\mathcal{M}) \end{array} \right\}. \tag{2.5}$$

□

By taking $X = xx^\top$, it is clear that $\mathcal{D} \subseteq \mathcal{D}_{\mathrm{SDP}}$. Furthermore, as $\mathcal{D}_{\mathrm{SDP}}$ is the projection of a convex set, it is itself convex. In particular, $\mathrm{conv}(\mathcal{D}) \subseteq \mathcal{D}_{\mathrm{SDP}}$ and $\mathrm{Opt} \ge \mathrm{Opt}_{\mathrm{SDP}}$.

Preliminaries on aggregation

We next present an alternative description of $\mathcal{D}_{\mathrm{SDP}}$ which will highlight the role played by the set of *convex Lagrange multipliers*.

Definition 2. The cone of *convex Lagrange multipliers*, Γ, associated with $\mathcal{D}$ is

$$\Gamma \coloneqq \left\{ (\gamma_{\mathrm{obj}}, \gamma) \in \mathbb{R}_+ \times \mathbb{R}^m_+ : \ \gamma_{\mathrm{obj}} A_{\mathrm{obj}} + A(\gamma) \succeq 0 \ \right\}.$$

It will also be useful to define the following slice of Γ,

$$\Gamma_P \coloneqq \left\{ \gamma \in \mathbb{R}^m_+ : (1,\gamma) \in \Gamma \right\} = \left\{ \gamma \in \mathbb{R}^m_+ : A[\gamma] \succeq 0 \right\}.$$

□

Note that Γ is a convex cone and that Γ_P is a closed convex set. These sets and their variants are known to be important in studying SDP relaxations of QCQPs; see for example [105, Chapter 13.4.2] where a lifted version of Γ is used to describe the SDP relaxation of a QCQP (cf. Lemma 1).

We will make the following definiteness assumption which can be interpreted as requiring the dual of (2.4) be strictly feasible. This is a standard assumption in the literature.

Assumption 1. There exists $(\gamma^*_{\mathrm{obj}}, \gamma^*) \in \Gamma$ such that $\gamma^*_{\mathrm{obj}} A_{\mathrm{obj}} + A(\gamma^*) \succ 0$. □

A version of the following description specifically for the SDP objective value was first recorded by Fujie and Kojima [42].

Lemma 1 ([101, Lemma 1])**.** *Suppose Assumption 1 holds. Then,*

$$\begin{aligned} \mathcal{D}_{\mathrm{SDP}} &= \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : \gamma_{\mathrm{obj}}(q_{\mathrm{obj}}(x) - t) + \langle \gamma, q(x)\rangle \le 0,\ \forall (\gamma_{\mathrm{obj}}, \gamma) \in \Gamma\} \\ &= \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : q[\gamma, x] \le t,\ \forall \gamma \in \Gamma_P\}. \end{aligned}$$

In particular, $\mathcal{D}_{\mathrm{SDP}} = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : (q_{\mathrm{obj}}(x) - t, q(x)) \in \Gamma^\circ\}$ *where* Γ° *is the polar cone of* Γ.

In other words, Lemma 1 states that $\mathcal{D}_{\mathrm{SDP}}$ is given by imposing the convex quadratic constraints on $(x,t) \in \mathbb{R}^n \times \mathbb{R}$ that can be obtained via Lagrange aggregation. We illustrate this in greater detail in the following example.

Example 1. Consider the following QCQP

$$\inf_{x\in\mathbb{R}^2} \left\{ x^\top A_{\mathrm{obj}} x : \begin{array}{l} x^\top A_1 x + 1 \le 0 \\ x^\top A_2 x + 1 \le 0 \end{array} \right\},$$

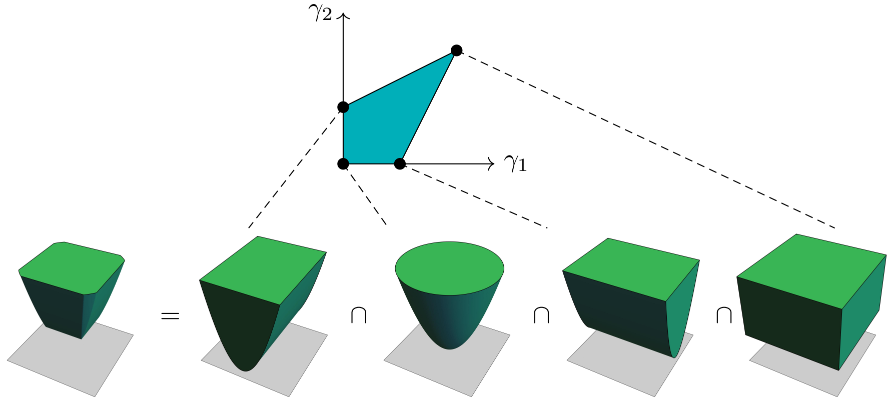

Figure 2.1: The blue region (first row) is a plot of Γ_P in Example 1. Lemma 1 then states that $\mathcal{D}_{\mathrm{SDP}}$ (the leftmost set on the second row) is equal to the intersection of the sets $\{(x,t)\in\mathbb{R}^2\times\mathbb{R} :\, q[\gamma,x]\leq t\}$ (the remaining sets on the bottom row) over the extreme points of this blue region.

where

$$A_{\mathrm{obj}} = \begin{pmatrix}1 & \\ & 1\end{pmatrix},\quad A_1 = \begin{pmatrix}-2 & \\ & 1\end{pmatrix},\quad A_2 = \begin{pmatrix}1 & \\ & -2\end{pmatrix}.$$

In this case, we may compute Γ and Γ_P explicitly. For example:

$$\begin{aligned}\Gamma_P &= \left\{\gamma\in\mathbb{R}^2_+ :\, A[\gamma]\succeq 0\right\}\\ &= \left\{\gamma\in\mathbb{R}^2_+ :\, \begin{array}{l}1-2\gamma_1+\gamma_2\geq 0\\ 1+\gamma_1-2\gamma_2\geq 0\end{array}\right\}\\ &= \operatorname{conv}\left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1/2\\0\end{pmatrix},\begin{pmatrix}0\\1/2\end{pmatrix},\begin{pmatrix}1\\1\end{pmatrix}\right\}.\end{aligned}$$

As Assumption 1 holds (e.g., take $(\gamma^*_{\mathrm{obj}},\gamma^*) = (1,0,0)\in\Gamma$), Lemma 1 implies that

$$\begin{aligned}\mathcal{D}_{\mathrm{SDP}} &= \left\{(x,t)\in\mathbb{R}^2\times\mathbb{R} :\; q[\gamma,x]\leq 0,\ \forall\gamma\in\Gamma_P\;\right\}\\ &= \left\{(x,t)\in\mathbb{R}^n\times\mathbb{R} :\; \begin{array}{l}x_1^2+x_2^2\leq t\\ (3/2)x_2^2+1/2\leq t\\ (3/2)x_1^2+1/2\leq t\\ 2\leq t\end{array}\right\}.\end{aligned}$$

Here, the second line follows as $\gamma\mapsto q[\gamma,x]$ is linear and Γ_P is the convex hull of its extreme points so that it suffices to impose the constraint in the first line on the extreme points of Γ_P. See Figure 2.1 for a visual depiction of the set $\mathcal{D}_{\mathrm{SDP}}$.

A key idea from this example is that when Γ and Γ_P are “simple”, Lemma 1 gives us a powerful tool to explicitly describe both $\mathcal{D}_{\mathrm{SDP}}$ and $\mathrm{Opt}_{\mathrm{SDP}}$. □

Remark 1. Note that under Assumption 1, Lemma 1 implies that $\mathcal{D}_{\mathrm{SDP}}$ is closed. □

2.3 Objective value exactness and polyhedral Γ

In this section, we take a first look at our framework which is predicated on understanding how the quantities $A(\gamma)$ and $b(\gamma)$ interact on faces of Γ. In this section, we will assume that Γ is polyhedral and present some conditions under which *objective value exactness*, $\mathrm{Opt}_{\mathrm{SDP}} = \mathrm{Opt}$, holds.

Assumption 2. The set of convex Lagrange multipliers, Γ, is polyhedral. □

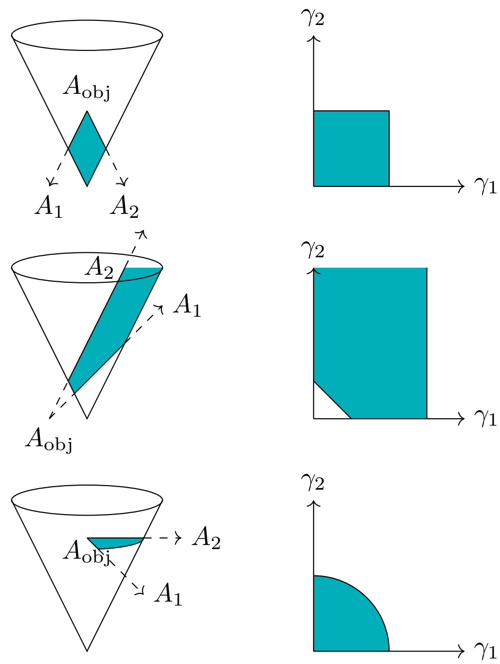

Figure 2.2: In each row above, we illustrate first the set $\left\{A[\gamma] \in \mathbb{S}^2 : \gamma \in \mathbb{R}^2_+\right\}$ on the left and the set $\Gamma_P = \left\{\gamma \in \mathbb{R}^2_+ : A[\gamma] \in \mathbb{S}^2_+\right\}$ on the right. It is not hard to show that, under Assumption 1, Γ is polyhedral if and only if Γ_P is polyhedral.

Although Assumption 2 is rather restrictive, it is general enough to cover the case where the set of quadratic forms $\{A_{\mathrm{obj}}, A_1, \dots, A_m\}$ is diagonal or simultaneously diagonalizable—a class of QCQPs which have been studied extensively in the literature [13, 66]. See also [85] for a characterization of polyhedral spectrahedra and a reduction showing that deciding whether a given spectrahedron is polyhedral is coNP-hard in general. See Figure 2.2 for an illustration of examples and nonexamples of Assumption 2.

Remark 2. We note that under Assumption 2 the set $\mathcal{D}_{\mathrm{SDP}}$ is in fact SOC-representable. Specifically, under this assumption, we may pick a finite subset $\left\{(\gamma^{(j)}_{\mathrm{obj}}, \gamma^{(j)})\right\} \subseteq \Gamma$ that generates Γ, i.e.,

$$\Gamma = \mathrm{cone}\left(\left\{(\gamma^{(j)}_{\mathrm{obj}}, \gamma^{(j)})\right\}\right).$$

Then, as in Example 1, $\mathcal{D}_{\mathrm{SDP}}$ is defined by finitely many convex quadratic constraints,

$$\mathcal{D}_{\mathrm{SDP}} = \left\{(x,t) \in \mathbb{R}^n \times \mathbb{R} : \gamma^{(j)}_{\mathrm{obj}}\, q_{\mathrm{obj}}(x) + \left\langle \gamma^{(j)}, q(x)\right\rangle \leq \gamma^{(j)}_{\mathrm{obj}}\, t,\ \forall j\right\},$$

whence it is SOC-representable.

A number of authors [13, 66] have noted that when the set of quadratic forms $\{A_{\mathrm{obj}}, A_1, \dots, A_m\}$ is simultaneously diagonalizable, that $\mathcal{D}_{\mathrm{SDP}}$ is SOC-representable. It is not hard to show that simultaneous diagonalizability implies Assumption 2 so that we also recover an SOC-representability result under simultaneous diagonalizability. In contrast, our SOC representation is in the original space but may involve exponentially many constraints (one for each extreme ray of Γ), while the SOC representation of $\mathcal{D}_{\mathrm{SDP}}$ given in [13, 66] uses n additional variables but only linearly many convex quadratic constraints. □

Definition 3. Let $\mathcal{F}$ be a face of Γ, denoted $\mathcal{F} \trianglelefteq \Gamma$. We say that $\mathcal{F}$ is a *definite face* if there exists $(\gamma_{\mathrm{obj}}, \gamma) \in \mathcal{F}$ such that $\gamma_{\mathrm{obj}} A_{\mathrm{obj}} + A(\gamma) \succ 0$. Otherwise, we say that $\mathcal{F}$ is a *semidefinite face.* □

We highlight that based on this definition, a face of Γ is *either* a definite face or a semidefinite face.

Definition 4. Given a subset $\mathcal{F} \subseteq \mathbb{R}^{m+1}$, define

$$\mathcal{V}(\mathcal{F}) \coloneqq \left\{ v \in \mathbb{R}^n : \, v^\top (\gamma_{\text{obj}} A_{\text{obj}} + A(\gamma)) v = 0, \, \forall (\gamma_{\text{obj}}, \gamma) \in \mathcal{F} \right\}. \qquad \square$$

The set $\mathcal{V}(\mathcal{F})$ plays an important role in our analysis. The main property we use of this set is the following: Suppose $(\hat{x}, \hat{t}) \in \mathcal{D}_{\text{SDP}}$ and $\mathcal{F}$ is the face of Γ exposed by $(q_{\text{obj}}(\hat{x}) - \hat{t}, q(\hat{x}))$. Then, for $x' \in \mathcal{V}(\mathcal{F})$ and $t' \in \mathbb{R}$, we have that the quadratic expressions associated with $(\gamma_{\text{obj}}, \gamma) \in \mathcal{F}$, i.e., $\gamma_{\text{obj}}(q_{\text{obj}}(x) - t) + \langle \gamma, q(x) \rangle \leq 0$, behave *linearly* when we perturb $(\hat{x}, \hat{t})$ in the direction $(x', 2t')$.[1] More formally, for any $(\gamma_{\text{obj}}, \gamma) \in \mathcal{F}$, $x' \in \mathcal{V}(\mathcal{F})$ and $t' \in \mathbb{R}$, the function $\epsilon \mapsto \gamma_{\text{obj}}(q_{\text{obj}}(\hat{x} + \epsilon x') - (\hat{t} + 2\epsilon t')) + \langle \gamma, q(\hat{x} + \epsilon x') \rangle$ is linear.

Remark 3. Note that if $\mathcal{F}$ is a subset of Γ, then

$$\mathcal{V}(\mathcal{F}) = \{ v \in \mathbb{R}^n : \, (\gamma_{\text{obj}} A_{\text{obj}} + A(\gamma)) v = 0, \, \forall (\gamma_{\text{obj}}, \gamma) \in \mathcal{F} \}$$

is the shared zero eigenspace of the matrices corresponding to $\mathcal{F}$. In particular $\mathcal{V}(\mathcal{F})$ is a linear subspace. It is not hard to show that when $\mathcal{F}$ is a semidefinite face of Γ, that $\dim(\mathcal{V}(\mathcal{F})) \geq 1$ (see [101, Lemma 2]). $\square$

We are now ready to state our first sufficient condition for objective value exactness. The following result comes from [101, Theorem 3].

Theorem 1. *Suppose Assumptions 1 and 2 hold. If for every semidefinite face $\mathcal{F}$ of Γ we have*

$$0 \notin \Pi_{\mathcal{V}(\mathcal{F})} \{ b[\gamma] : \, (1, \gamma) \in \mathcal{F} \}, \tag{2.6}$$

then any optimizer $(x^*, t^*) \in \arg\min_{(x,t) \in \mathcal{D}_{\text{SDP}}} t$ satisfies $(x^*, t^*) \in \mathcal{D}$. In particular, $\text{Opt} = \text{Opt}_{\text{SDP}}$.

We remark that this theorem proves something stronger than objective value exactness. Specifically, Theorem 1 states that the *only* optimizers of the projected SDP relaxation are the original optimizers.

We give a high-level proof sketch of this statement; see [101, Theorem 3] for a complete proof. We emphasize that the structure of this proof will serve as the foundation for many of the proofs in this section and that we will routinely return to this discussion.

Proof sketch of Theorem 1. Let $(\hat{x}, \hat{t}) \in \mathcal{D}_{\text{SDP}}$ and let $\mathcal{F}$ denote the face of Γ maximizing the linear function

$$(\gamma_{\text{obj}}, \gamma) \mapsto \gamma_{\text{obj}}(q_{\text{obj}}(\hat{x}) - \hat{t}) + \langle \gamma, q(\hat{x}) \rangle .$$

Equivalently, $\mathcal{F}$ is the face of Γ exposed by $(q_{\text{obj}}(\hat{x}) - \hat{t}, q(\hat{x}))$.

If $\mathcal{F} = \Gamma$, then by the fact that Γ is full-dimensional (Assumption 1), it must be the case that $q_{\text{obj}}(\hat{x}) = \hat{t}$ and $q_i(\hat{x}) = 0$ for all $i \in [m]$. More generally, it is possible to show that if $\mathcal{F}$ is a definite face, then $(\hat{x}, \hat{t}) \in \mathcal{D}$ (see [101, Lemma 3]).

To prove Theorem 1, we will suppose $(\hat{x}, \hat{t}) \in \arg\min_{(x,t) \in \mathcal{D}_{\text{SDP}}} t$ corresponds to a semidefinite face of Γ and construct a new point $(\hat{x} + \epsilon x', \hat{t} + 2\epsilon t') \in \mathcal{D}_{\text{SDP}}$ such that $\hat{t} + 2\epsilon t' < \hat{t}$, contradicting the assumption that $(\hat{x}, \hat{t}) \in \arg\min_{(x,t) \in \mathcal{D}_{\text{SDP}}} t$.

We will use (2.6) to construct this new point. Note that (2.6) is equivalent the existence of $x' \in \mathcal{V}(\mathcal{F})$ and $t' < 0$ such that

$$\langle b[\gamma], x' \rangle \leq t', \, \forall (1, \gamma) \in \mathcal{F}. \tag{2.7}$$

This holds vacuously if $\{\gamma : \, (1, \gamma) \in \mathcal{F}\} = \varnothing$ as we may take $x' = 0$ and $t' = -1$. If $\{\gamma : \, (1, \gamma) \in \mathcal{F}\}$ is nonempty, then we may apply the hyperplane separation theorem to strictly separate 0 from the polyhedral set $\Pi_{\mathcal{V}(\mathcal{F})} \{ b_{\text{obj}} + b(\gamma) : \, (1, \gamma) \in \mathcal{F} \}$ (recall Assumption 2) and take (x', t') according to the separating hyperplane.

We claim that in either case, for all $(\gamma_{\text{obj}}, \gamma) \in \mathcal{F}$ and $\epsilon > 0$, we have

$$\gamma_{\text{obj}} \left(q(\hat{x} + \epsilon x') - (\hat{t} + 2\epsilon t') \right) + \langle \gamma, q(\hat{x} + \epsilon x') \rangle \leq 0.$$

[1] The factor of 2 here is not important and is only included to unify notation with future sections.

This clearly holds in the first case as $x' = 0$ and $\gamma_{\text{obj}} = 0$ for all $(\gamma_{\text{obj}}, \gamma) \in \mathcal{F}$. In the second case, note that $\mathcal{F} = \text{clcone}(\{(1, \gamma) : (1, \gamma) \in \mathcal{F}\})$ so that it suffices to verify this inequality for points of the form $(1, \gamma) \in \mathcal{F}$. The claim then follows from the observation that $x' \in \mathcal{V}(\mathcal{F})$ and $\langle b[\gamma], x' \rangle \le t' < 0$ for all $(1, \gamma) \in \mathcal{F}$.

On the other hand, for $(\gamma_{\text{obj}}, \gamma) \in \Gamma \setminus \mathcal{F}$,

$$\gamma_{\text{obj}} \left(q(\hat{x} + \epsilon x') - (\hat{t} + 2\epsilon t') \right) + \langle \gamma, q(\hat{x} + \epsilon x') \rangle \tag{2.8}$$

is a quadratic function in ϵ which is negative at $\epsilon = 0$.

Then, using the fact that Γ is polyhedral (so that it suffices to ensure that (2.8) is nonpositive for only finitely many choices of $(\gamma_{\text{obj}}, \gamma)$), we deduce that there exists $\epsilon > 0$ small enough such that $(\hat{x} + \epsilon x', \hat{t} + 2\epsilon t') \in \mathcal{D}_{\text{SDP}}$. This contradicts the assumption that $(\hat{x}, \hat{t}) \in \arg\min_{(x,t) \in \mathcal{D}_{\text{SDP}}} t$ and concludes the proof sketch. ■

Remark 4. We highlight some of the key steps in the above proof with additional intuition and motivation. Recall that Γ is assumed to be polyhedral so that $\mathcal{D}_{\text{SDP}}$ is defined by finitely many convex quadratic constraints. We will for the sake of simplicity assume that Γ_P is bounded so that it is the convex hull of its extreme points $\Gamma_P = \text{conv}\left(\left\{\gamma^{(j)}\right\}\right)$. The above proof sketch starts with $(\hat{x}, \hat{t}) \in \mathcal{D}_{\text{SDP}}$ such that the face $\mathcal{F}$ of Γ exposed by $(q_{\text{obj}}(\hat{x}) - \hat{t}, q(\hat{x}))$ is semidefinite. Note that if j satisfies $(1, \gamma^{(j)}) \in \mathcal{F}$, then the corresponding inequality is tight at $(\hat{x}, \hat{t})$, i.e., $q[\gamma^{(j)}, \hat{x}] = \hat{t}$. For all other j, the corresponding quadratic constraints are strictly satisfied at $(\hat{x}, \hat{t})$ so that they remain satisfied under small enough perturbations of $(\hat{x}, \hat{t})$. In particular, given a perturbation direction $(x', 2t')$ we have $(x + \epsilon x', t + 2\epsilon t') \in \mathcal{D}_{\text{SDP}}$ for all $\epsilon > 0$ small enough if and only if each of the tight constraints continues to hold for all $\epsilon > 0$ small enough, i.e.,

$$q[\gamma^{(j)}, \hat{x} + \epsilon x'] \le \hat{t} + 2\epsilon t'$$

for every j such that $(1, \gamma^{(j)}) \in \mathcal{F}$ and $\epsilon > 0$ small enough. For a given x', it is possible to show that there exists $t' < 0$ satisfying this requirement if and only if each of the tight quadratic constraints is strictly decreasing at $\hat{x}$ in the direction x'. That is to say, we would like to find a direction x', such that each of the convex quadratic functions $q[\gamma^{(j)}, x]$ for $(1, \gamma^{(j)}) \in \mathcal{F}$ has a negative first derivative at $\hat{x}$ in the direction x'. Expanding, we would like to find x' such that $\left\langle A[\gamma^{(j)}]\hat{x} + b[\gamma^{(j)}], x' \right\rangle < 0$ for all $(1, \gamma^{(j)}) \in \mathcal{F}$. Finally, restricting our search of x' to $\mathcal{V}(\mathcal{F})$, i.e., directions in which all of the tight quadratic constraints are only linear, we then have that such an $(x', 2t')$ exists if and only if there exists $x' \in \mathcal{V}(\mathcal{F})$ such that

$$\left\langle b[\gamma^{(j)}], x' \right\rangle < 0, \; \forall (1, \gamma^{(j)}) \in \mathcal{F}.$$ □

A similar proof strategy allows us to relax the conditions of Theorem 1 if only objective value exactness is required.

Theorem 2. *Suppose Assumptions 1 and 2 hold. If for every semidefinite face $\mathcal{F}$ of Γ we have*

$$\text{cone}\left(\Pi_{\mathcal{V}(\mathcal{F})} \left\{ b[\gamma] : (1, \gamma) \in \mathcal{F} \right\}\right) \ne \mathcal{V}(\mathcal{F}), \tag{2.9}$$

then $\text{Opt} = \text{Opt}_{\text{SDP}}$.

Proof sketch of Theorem 2. This proof follows a similar structure to the proof of Theorem 1. Again, given $(\hat{x}, \hat{t}) \in \mathcal{D}_{\text{SDP}}$, we will look at $\mathcal{F}$, the face of Γ exposed by $(q_{\text{obj}}(\hat{x}) - \hat{t}, q(\hat{x}))$. If $\mathcal{F}$ is definite, then we may conclude that in fact $(\hat{x}, \hat{t}) \in \mathcal{D}$. On the other hand, if $\mathcal{F}$ is semidefinite, we will use (2.9) to find a direction (x', t') and $\alpha \in \mathbb{R}$ such that $(\hat{x} + \alpha x', \hat{t} + \alpha t') \in \mathcal{D}_{\text{SDP}}$ with $\hat{t} + \alpha t' \le \hat{t}$. To complete this proof, one then needs to show that by picking α appropriately and iterating this procedure a finite number of times, we eventually end up with a point in $\mathcal{D}$. ■

Remark 5. Note that the condition in Theorem 2 can be rewritten as the existence, for each semidefinite face $\mathcal{F}$ of Γ, of a nonzero $x' \in \mathcal{V}(\mathcal{F})$ such that

$$\langle b[\gamma], x' \rangle \le 0, \; \forall (1, \gamma) \in \mathcal{F}. \tag{2.10}$$

Indeed, this equivalence, i.e., (2.9) $\iff$ (2.10), holds vacuously if $\{\gamma : (1, \gamma) \in \mathcal{F}\} = \varnothing$. Else, $\{\gamma : (1, \gamma) \in \mathcal{F}\}$ is nonempty and the equivalence follows from the general result for a nonempty convex subset C of a Euclidean vector space $\mathbb{E}$, that $\operatorname{cone}(C) = \mathbb{E}$ if and only if $\operatorname{clcone}(C) = \mathbb{E}$ if and only if $C^\circ = \{0\}$.

In this form, it is also easy to see that the sufficient condition in Theorem 1 implies the sufficient condition in Theorem 2. Specifically we have (i) (2.6) $\iff$ (2.7) (with $t' < 0$), (ii) (2.7) (with $t' < 0$) $\implies$ (2.10), and (iii) (2.10) $\iff$ (2.9). Here, the first equivalence was sketched in the proof of Theorem 1. □

Remark 6. Burer and Ye [23] consider the standard SDP relaxation of diagonal QCQPs and present sufficient conditions on the input data that guarantee objective value exactness. Specifically, they show that if for all $j \in [n]$, it holds that

$$(1, \gamma) \in \Gamma,\ A[\gamma]_{j,j} = 0 \implies b[\gamma]_j \neq 0, \tag{2.11}$$

then objective value exactness holds. It is not hard to see that (2.11) implies the assumptions of Theorem 1 so that Theorem 1 recovers [23, Theorem 1] as a special case (see [101, Proposition 5]).

Locatelli [66] considers the standard SDP relaxation of the TRS with additional linear constraints

$$\inf_{x \in \mathbb{R}^n} \left\{ q_{\text{obj}}(x) : \begin{array}{l} 2b_i^\top x + c_i \leq 0,\ \forall i \in [m-1] \\ x^\top x - 1 \leq 0 \end{array} \right\}.$$

For this setup, Locatelli [66] shows that the SDP relaxation for this problem is exact if for all $\epsilon > 0$, there exists $\|h_\epsilon\| \leq \epsilon$ such that

$$0 \notin \Pi_{\mathcal{V}} \left\{ b[\gamma] + h_\epsilon : \gamma \in \mathbb{R}^m_+ \right\}, \tag{2.12}$$

where $\mathcal{V}$ is the subspace of $\mathbb{R}^n$ corresponding to the minimum eigenvalue (assumed to be negative) of A_{obj}. [101, Proposition 4] establishes that (2.12) implies the assumptions of Theorem 2 so that Theorem 2 recovers [66, Theorem 3.1] as a special case. □

We next discuss a few applications of these results.

Example 2 (Convex QCQPs)**.** As a first example, let us consider the setting where $\mathcal{D}$ corresponds to a QCQP with a strongly convex objective and convex constraints. Specifically, suppose $A_{\text{obj}} \succ 0$, $A_i \succeq 0$ for all $i \in [m]$. Then,

$$\Gamma \coloneqq \left\{ (\gamma_{\text{obj}}, \gamma) \in \mathbb{R}_+ \times \mathbb{R}^m_+ : A_{\text{obj}} + A(\gamma) \succeq 0 \right\} = \mathbb{R}_+ \times \mathbb{R}^m_+$$

is polyhedral. Furthermore, as $A_{\text{obj}} \succ 0$, no semidefinite face $\mathcal{F}$ of Γ can contain a point of the form $(1, \gamma)$. Then, we deduce by Theorem 1, that $\text{Opt} = \text{Opt}_{\text{SDP}}$ for convex QCQPs. Here, the requirement that $A_{\text{obj}} \succ 0$ can be removed by a standard perturbation argument; see [23, Proposition 4]. □

Example 3 (Diagonal QCQPs with sign-definite linear terms)**.** The following condition is presented in [90, Corollary 1] (see also [23]). Consider the following setup: Suppose $A_{\text{obj}}, A_1, \ldots, A_m$ are all diagonal. Furthermore, suppose Assumption 1 holds and that for each $j \in [n]$, the set of coefficients

$$\{(b_{\text{obj}})_j\} \cup \{(b_i)_j : i \in [m]\}$$

are either all nonnegative or nonpositive. For all $j \in [n]$, let $\sigma_j \in \{\pm 1\}$ be such that $(b_{\text{obj}})_j \sigma_j \leq 0$ and $(b_i)_j \sigma_j \leq 0$ for all $i \in [m]$. We can apply Theorem 2 in this setting. Indeed, suppose $\mathcal{F}$ is a semidefinite face of Γ. As A_{obj} and $A_1, \ldots, A_m$ are diagonal, there exists a nonempty subset of indices $\mathcal{J} \subseteq [n]$ such that

$$\mathcal{F} = \left\{ (\gamma_{\text{obj}}, \gamma) \in \Gamma : \operatorname{diag}\left(\gamma_{\text{obj}} A_{\text{obj}} + A(\gamma)\right)_j = 0,\ \forall j \in \mathcal{J} \right\}.$$

Without loss of generality, we may assume $\mathcal{J} = [\ell]$ with $\ell \geq 1$ so that $\sigma_1 e_1 \in \mathcal{V}(\mathcal{F})$. Then, for all $(1, \gamma) \in \mathcal{F}$,

$$\langle b[\gamma], \sigma_1 e_1 \rangle = b[\gamma]_1 \sigma_1 = (b_{\text{obj}})_1 \sigma_1 + \sum_{i=1}^m \gamma_i (b_i)_1 \sigma_1 \leq 0,$$

where the last inequality follows from $\gamma \in \mathbb{R}^m_+$ (as $(1,\gamma) \in \mathcal{F}$) and that $(b_{\mathrm{obj}})_j\sigma_j \le 0$ and $(b_i)_j\sigma_j \le 0$ for all $i \in [m]$. Then, we deduce by Theorem 2 and Remark 5 that $\mathrm{Opt} = \mathrm{Opt}_{\mathrm{SDP}}$ for diagonal QCQPs with sign-definite linear terms. ☐

Example 4 (QCQPs with centered constraints and polyhedral Γ)**.** Suppose $b_i = 0$ for all $i \in [m]$ and that Assumptions 1 and 2 hold. Then, for any semidefinite face $\mathcal{F}$ of Γ, we have

$$\mathrm{cone}\left(\Pi_{\mathcal{V}(\mathcal{F})}\left\{b[\gamma] :\ (1,\gamma) \in \mathcal{F}\right\}\right) \subseteq \mathrm{cone}\left(\Pi_{\mathcal{V}(\mathcal{F})} b_{\mathrm{obj}}\right)$$

is a cone generated by a single point and cannot be $\mathcal{V}(\mathcal{F})$ (recall that $\mathcal{V}(\mathcal{F})$ has dimension at least one as $\mathcal{F}$ is semidefinite). Then, we deduce by Theorem 2 that $\mathrm{Opt} = \mathrm{Opt}_{\mathrm{SDP}}$ for QCQPs with centered constraints and polyhedral Γ. Further specializing this example, we may consider the QCQP

$$\min_{x\in\mathbb{R}^n}\left\{-x^\top x :\ \begin{array}{l} x^\top x \le 1 \\ x^\top A_i x = 0,\ \forall i \in [1,m] \end{array}\right\},$$

which takes the value -1 when the quadratic forms $\{A_i\}$ have a nontrivial joint zero and 0 otherwise. Barvinok [9] shows that a specialized algebraic algorithm (not based on the SDP relaxation) can be used to decide the value of this program in polynomial time for any constant m. On the other hand, Theorem 2 shows that we can decide the value of this program by solving a semidefinite program (for any m) whenever Assumption 2 holds. See also [101, Remark 9]. ☐

Remark 7. The conditions presented in Theorems 1 and 2 may be strengthened by taking into account the values of c_i. For example, it suffices to impose the constraints of Theorems 1 and 2 on only the semidefinite faces exposed by vectors of the form $(q_{\mathrm{obj}}(x) - t, q(x))$ for some $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$. Ideas related to this observation have been studied in further detail in [67, 100]. See also Section 2.5. ☐

2.4 Convex hull exactness and polyhedral Γ

In this section, we will continue to assume that Γ is polyhedral (Assumption 2) and present sufficient conditions under this assumption for *convex hull exactness*, $\mathrm{conv}(\mathcal{D}) = \mathcal{D}_{\mathrm{SDP}}$. As in the previous section, the sufficient conditions stem from understanding how $A(\gamma)$ and $b(\gamma)$ interact on faces of Γ.

The following theorem provides a sufficient condition for convex hull exactness.

Theorem 3 ([101, Theorem 1])**.** *Suppose Assumptions 1 and 2 hold. If for every semidefinite face $\mathcal{F}$ of Γ we have*

$$\mathrm{aff}\left(\Pi_{\mathcal{V}(\mathcal{F})}\left\{b[\gamma] :\ (1,\gamma) \in \mathcal{F}\right\}\right) \ne \mathcal{V}(\mathcal{F}), \tag{2.13}$$

then $\mathrm{conv}(\mathcal{D}) = \mathcal{D}_{\mathrm{SDP}}$*.*

Proof sketch of Theorem 3. The proof of this statement follows a similar structure to the proofs of Theorems 1 and 2: Recalling that $\mathrm{conv}(\mathcal{D}) \subseteq \mathcal{D}_{\mathrm{SDP}}$, it suffices to show that we may write any $(\hat{x},\hat{t}) \in \mathcal{D}_{\mathrm{SDP}}$ as a convex combination of points in $\mathcal{D}$. Given $(\hat{x},\hat{t}) \in \mathcal{D}_{\mathrm{SDP}}$, we will examine $\mathcal{F}$, the face of Γ exposed by $(q_{\mathrm{obj}}(\hat{x})-t, q(\hat{x}))$. If $\mathcal{F}$ is definite, then we may conclude that in fact $(\hat{x},\hat{t}) \in \mathcal{D}$. On the other hand, if $\mathcal{F}$ is semidefinite, we will use (2.13) to find a direction (x',t') and $\alpha_1 < 0 < \alpha_2$ such that $(\hat{x}+\alpha x', \hat{t}+\alpha t') \in \mathcal{D}_{\mathrm{SDP}}$ for both α_1 and α_2. To complete this proof, one then needs to show that by picking α_1, α_2 appropriately and iterating this procedure for both choices of α, we eventually end up with a convex decomposition of $(\hat{x},\hat{t})$ as points in $\mathcal{D}$. See [101, Section 4] for details. ■

Remark 8. We compare the assumptions of Theorems 2 and 3. Note that we may rewrite the condition in Theorem 3 as the existence, for each semidefinite face $\mathcal{F}$ of Γ, of a nonzero $v \in \mathcal{V}(\mathcal{F})$ and $r \in \mathbb{R}$ such that

$$\langle b[\gamma], v\rangle = r,\ \forall (1,\gamma) \in \mathcal{F}. \tag{2.14}$$

This equivalence, i.e., (2.13) $\iff$ (2.14), holds vacuously if $\{\gamma : (1,\gamma) \in \mathcal{F}\} = \varnothing$. Else, $\{\gamma : (1,\gamma) \in \mathcal{F}\}$ is nonempty and this equivalence follows from the observation that an affine subspace is not the entirety of a vector space if and only if it is contained in an affine hyperplane.

In this form, it is clear that the assumptions of Theorem 3 imply the assumptions of Theorem 2 (see Remark 5). We remark that replacing the inequality in (2.10) by an equality in (2.14) is natural when moving from objective value exactness to convex hull exactness. Specifically, in contrast to the proof of Theorem 2, which needs only move in a single direction (i.e., in a direction in which the objective value is nonincreasing), the proof of Theorem 3 must move in both directions (i.e., for both $\alpha_1 < 0$ and $0 < \alpha_2$). $\square$

Remark 9. Recall that when Γ is polyhedral we have that $\mathcal{D}_{\mathrm{SDP}}$ is SOC-representable (see Remark 2). In particular, under the assumptions of Theorem 3, we have that $\operatorname{conv}(\mathcal{D})$ is also SOC-representable.

A number of results in the literature give second-order cone representations of convex hulls of quadratically constrained sets with a small number of constraints. For example, Ho-Nguyen and Kılınç-Karzan [52] show that the epigraph of the TRS is given by the intersection of two convex quadratic regions. Wang and Kılınç-Karzan [99] extend these results to the GTRS. Yıldıran [110] shows that the convex hull of the intersection of two strict quadratic inequalities is given by (open) second-order cone constraints. Follow-up work by Modaresi and Vielma [71] show that it is possible to take the closure under an additional technical assumption. Burer and Kılınç-Karzan [20] examine the convex hull of the intersection of the second-order cone with a nonconvex quadratic constraint established that it is SOC-representable under certain conditions. More recently, Santana and Dey [87] showed that the convex hull of the intersection of a polytope and a single nonconvex quadratic constraints is SOC-representable. $\square$

We close this section with example applications of our results.

Example 5 (GTRS)**.** The problem of minimizing a (possibly nonconvex) quadratic function subject to a (possibly nonconvex) quadratic inequality[2] constraint is known as the generalized trust-region subproblem (GTRS). The special case where the quadratic constraint is convex is known as the trust-region subproblem (TRS) and is fundamental in the area of nonlinear programming. Specifically, trust-region methods (iterative methods for solving nonlinear optimization problems that solve a TRS instance at each iteration) see strong theoretical guarantees as well as empirical performance [30].

Supposing that Assumption 1 holds, we may apply Theorem 3 to deduce that $\operatorname{conv}(\mathcal{D}) = \mathcal{D}_{\mathrm{SDP}}$ for the GTRS. Specifically, in this setting we can write Γ as

$$\Gamma = \left\{(\gamma_{\mathrm{obj}}, \gamma_1) \in \mathbb{R}^2_+ : \gamma_{\mathrm{obj}} A_{\mathrm{obj}} + \gamma_1 A_1 \succeq 0\right\}.$$

As Γ is a conic subset of $\mathbb{R}^2$, it is immediately polyhedral. It is not hard to verify that if $\mathcal{F}$ is a semidefinite face of Γ, then $\{\gamma \in \mathbb{R} : (1,\gamma) \in \mathcal{F}\}$ is either empty or a single point. We deduce by Theorem 3 and Remark 8 that $\operatorname{conv}(\mathcal{D}) = \mathcal{D}_{\mathrm{SDP}}$ for the GTRS. This recovers [99, Theorem 1] as a special case. $\square$

Example 6 (Swiss cheese)**.** Suppose $A_{\mathrm{obj}}, A_i \in \{I, 0, -I\}$ for all $i \in [m]$. This setting captures, for example, the problem of finding the minimum norm point on some domain defined by "inside ball" constraints, "outside ball" constraints, and halfspaces. In this setting, for any semidefinite face $\mathcal{F}$ of Γ, we have $\mathcal{V}(\mathcal{F}) = \mathbb{R}^n$ so that

$$\begin{aligned}\operatorname{aff\,dim}\left(\Pi_{\mathcal{V}(\mathcal{F})}\left\{b[\gamma] : (1,\gamma) \in \mathcal{F}\right\}\right) &= \operatorname{aff\,dim}\left\{b[\gamma] : (1,\gamma) \in \mathcal{F}\right\}\\ &\le \operatorname{aff\,dim}\left\{\gamma : (1,\gamma) \in \mathcal{F}\right\} \le m-1.\end{aligned}$$

Here, the last inequality follows as $\mathcal{F}$ is a semidefinite face of Γ (note that if $\operatorname{aff\,dim}\{\gamma : (1,\gamma) \in \mathcal{F}\} = m$, then $\mathcal{F}$ has affine dimension $m+1$, which contradicts that $\mathcal{F}$ is a semidefinite face of Γ).

We deduce by Theorem 3 that if $m \le n$, then $\operatorname{conv}(\mathcal{D}) = \mathcal{D}_{\mathrm{SDP}}$.

Similar setups have been considered in the literature. For example, Bienstock and Michalka [17] devise an enumerative algorithm for minimizing an *arbitrary* quadratic function over a feasible domain defined by *a constant number* of "inside ball," "outside ball," and halfspace constraints. In contrast, our results

[2] Similar statements can also be derived for the GTRS with an equality constraint.

(via Theorem 3) deal only with objective functions of a particular form, but work with the standard SDP relaxation and do not make any assumption on the number of constraints.

Yang et al. [108] consider QCQPs with additional "hollow" constraints. Formally, they show that if a QCQP with bounded domain $\mathcal{X}$ satisfies objective value exactness, then so too does the QCQP with domain $\mathcal{X} \setminus \bigcup_\alpha \operatorname{int}(E_\alpha)$ where $\{E_\alpha\}$ is a finite set of *non-intersecting* ellipsoids completely contained within $\mathcal{X}$. Taking $\mathcal{X}$ to be the unit ball, this result then says that semidefinite programs can correctly minimize an arbitrary quadratic function over the sphere missing a finite number of nonintersecting ellipsoids inside the ball. In contrast, our results (via Theorem 3) deal only with spherical objective functions and constraints (as opposed to general objective functions and ellipsoidal constraints), but do not make any assumption on how the constraints intersect. See also [101, Remark 10]. The follow-up recent work [59] shows that a similar result holds even if $\mathcal{X}$ is unbounded and the "hollow" constraints are not necessarily ellipsoidal. □

Example 7 (QCQPs with large amounts of symmetry)**.** The following setup is considered in [101, Section 3]. Consider a general QCQP of the form (2.1) and let $1 \leq k \leq n$ denote the largest positive integer such that each of the quadratic forms $A_{\text{obj}}, A_1, \dots, A_m$ in the QCQP can be written in the form

$$A_{\text{obj}} = I_k \otimes \mathbb{A}_{\text{obj}}, \quad A_i = I_k \otimes \mathbb{A}_i$$

for some $\mathbb{A}_{\text{obj}}, \mathbb{A}_i \in \mathbb{S}^{n/k}$. This quantity is referred to as the *quadratic eigenvalue multiplicity* of the underlying QCQP and can be thought of as a measure of the amount of symmetry in the QCQP. In this example, we will assume that Assumptions 1 and 2 hold and that k is large.

Such structure arises naturally when considering the vectorized reformulation of quadratic matrix programs (QMPs) [10, 12]. Specifically, a QMP is an optimization problem of the form,

$$\inf_{Y \in \mathbb{R}^{s\times k}} \left\{ \operatorname{tr}(Y^\top \mathbb{A}_{\text{obj}} Y) + 2\operatorname{tr}(B_{\text{obj}}^\top Y) + c_{\text{obj}} : \quad \operatorname{tr}(Y^\top \mathbb{A}_i Y) + 2\operatorname{tr}(B_i^\top Y) + c_i \leq 0,\ \forall i \in [m] \ \right\},$$

where $\mathbb{A}_i \in \mathbb{S}^s$, $B_i \in \mathbb{R}^{s\times k}$, and $c_i \in \mathbb{R}$. Then, letting $x \in \mathbb{R}^{sk}$ (resp. $b \in \mathbb{R}^{sk}$) denote the vector obtained by stacking the columns of $Y \in \mathbb{R}^{s\times k}$ (resp. $B \in \mathbb{R}^{s\times k}$) on top of each other, we have

$$\operatorname{tr}(Y^\top \mathbb{A} Y) + 2\operatorname{tr}(B^\top Y) + c = x^\top (I_k \otimes A)x + 2b^\top x + c.$$

The quadratic eigenvalue multiplicity can also be viewed as an example of a group symmetry in $\{A_{\text{obj}}, A_1, \dots, A_m\}$. Group symmetries have been studied in more generality with the goal of reducing the size of large SDPs [33, 43] and have enabled the efficient solution of numerous large-scale problems; see for example [32]. Specifically, the *Wedderburn decomposition* of the matrix $\mathbb{C}^*$-algebra generated by $\{A_{\text{obj}}, A_1, \dots, A_m\}$ plays a prominent role in the analysis of such symmetries (see [34, 44] for background on the Wedderburn decomposition and related numerical questions). From this point of view, one may compare the quadratic eigenvalue multiplicity as defined above with the "block multiplicity" of a basic algebra in the Wedderburn decomposition of the corresponding $\mathbb{C}^*$ algebra. See also [101, Remark 5] and references therein.

It is not hard to show for any semidefinite face $\mathcal{F}$ of Γ, that $\dim(\mathcal{V}(\mathcal{F})) \geq k$. Indeed, there exists a nonzero $y \in \mathbb{R}^{n/k}$ such that for all $(\gamma_{\text{obj}}, \gamma) \in \mathcal{F}$, we have $(\gamma_{\text{obj}}\mathbb{A}_{\text{obj}} + \mathbb{A}(\gamma))\, y = 0$. We thus deduce for all $(\gamma_{\text{obj}}, \gamma) \in \mathcal{F}$ and $w \in \mathbb{R}^k$ that

$$(\gamma_{\text{obj}} A_{\text{obj}} + A(\gamma))\,(w \otimes y) = 0,$$

whence $\dim(\mathcal{V}(\mathcal{F})) \geq k$. See also [101, Lemma 6].

On the other hand, for any semidefinite face $\mathcal{F}$ of Γ, we may upper bound

$$\operatorname{aff\,dim}\left(\Pi_{\mathcal{V}(\mathcal{F})}\left\{b[\gamma] :\ (1,\gamma) \in \mathcal{F}\right\}\right) \leq \min\left(\left|\{i \in [m] :\ b_i \neq 0\}\right|,\ m-1\right).$$

We deduce by (2.13) that $\operatorname{conv}(\mathcal{D}) = \mathcal{D}_{\text{SDP}}$ as long as

$$k \geq \min\left(\left|\{i \in [m] :\ b_i \neq 0\}\right| + 1,\ m\right). \tag{2.15}$$

Note that this statement immediately subsumes Examples 4 to 6. See [101, Section 4.3] for constructions showing that our bounds on the value of k guaranteeing both objective value and convex hull exactness are sharp. □

2.5 Removing the polyhedrality assumption

While the results in [98, 101] are general enough to cover a number of interesting settings, the assumption that the cone of convex Lagrange multipliers is polyhedral (Assumption 2) is nonetheless quite restrictive. In this section, we show how to extend the results in Sections 2.3 and 2.4 to non-polyhedral Γ. This generalization widely broadens the applicability of our framework and recovers convex hull results for quadratic matrix programs without an explicit polyhedrality assumption [10, 12] as well as a basic mixed-binary set related to the "perspective reformulation/relaxation trick" [41, 48]. We remark that the "perspective reformulation/relaxation trick" is well known in the literature and has been useful in deriving convex hull exactness for a variety of sets arising in sparsity-constrained optimization [6, 27, 38, 41, 48, 102, 103].

In contrast to the results thus far, which have been stated in terms of the faces of Γ, in this section we will focus on the faces of Γ°, the polar cone of Γ. The following definitions and results are from [100].

Definition 5. For $(\hat{x},\hat{t}) \in \mathcal{D}_{\rm SDP}$, let $\mathcal{G}(\hat{x},\hat{t})$ denote the minimal face of Γ° containing $(q_{\rm obj}(\hat{x}) - \hat{t}, q(\hat{x}))$. $\square$

Remark 10. Let $\mathcal{G}$ be a face of Γ°. We have by definition of $\mathcal{V}$ (see Definition 4)

$$\mathcal{V}(\mathcal{G}^\perp) = \left\{v \in \mathbb{R}^n :\, v^\top(\gamma_{\rm obj}A_{\rm obj} + A(\gamma))v = 0,\, \forall(\gamma_{\rm obj},\gamma) \in \mathcal{G}^\perp\right\}.$$

Note that in particular, the definition of $\mathcal{V}(\mathcal{G}^\perp)$ contains nonconvex quadratic constraints. $\square$

In the general setting where Γ may not be polyhedral, much of our analysis will be done with the object $\mathcal{G}^\perp$. This object will replace the face of Γ that we used extensively in Sections 2.3 and 2.4. The following example compares $\mathcal{G}^\perp$ with the face of Γ that is natural to consider in this setting. Specifically, we will compare $\mathcal{G}^\perp$ with $\mathcal{G}^\triangle$, the conjugate face of $\mathcal{G}$ in Γ. Recall that given a face $\mathcal{G}$ of Γ°, the conjugate face $\mathcal{G}^\triangle$ is the face of Γ given by $\mathcal{G}^\triangle \coloneqq \Gamma \cap \mathcal{G}^\perp$.

Example 8. Consider the setting where Γ and Γ° are both the standard second-order cone in $\mathbb{R}^{n+1}$. Let $\mathcal{G}$ be a one-dimensional face of Γ° and let $\mathcal{G}^\triangle$ be the face of Γ conjugate to $\mathcal{G}$. Then, $\mathcal{G}^\triangle$ is a one-dimensional face of Γ so that $\operatorname{span}(\mathcal{G}^\triangle)$ is a one-dimensional subspace. On the other hand, $\mathcal{G}^\perp$ is an n-dimensional subspace. See Figure 2.3 for an illustration of the relevant sets for $n = 2$.

In general, $\operatorname{span}(\mathcal{G}^\triangle) \subseteq \mathcal{G}^\perp$ where equality may not necessarily hold. On the other hand, it is possible to show that equality holds whenever Γ and Γ° are polyhedral (see [100, Theorem 4]). $\square$

The following theorem gives a generalization of Theorem 3 by establishing a sufficient condition for convex hull exactness without relying on a polyhedrality assumption on Γ.

Theorem 4. *Suppose Assumption 1 holds. If for every $(\hat{x},\hat{t}) \in \mathcal{D}_{\rm SDP} \setminus \mathcal{D}$, there exists $x' \in \mathcal{V}(\mathcal{G}(\hat{x},\hat{t})^\perp)$ and $t' \in \mathbb{R}$ such that $(x',t') \neq (0,0)$ and*

$$\langle A[\gamma]\hat{x} + b[\gamma], x'\rangle = t',\, \forall(1,\gamma) \in \mathcal{G}(\hat{x},\hat{t})^\perp, \tag{2.16}$$

then $\operatorname{conv}(\mathcal{D}) = \mathcal{D}_{\rm SDP}$.

Remark 11. It is possible to show (see the proof of [100, Lemma 2]) that when Γ° is facially exposed (as is the case when Γ is polyhedral), for every face $\mathcal{G}$ of Γ°, we have

$$\mathcal{V}(\mathcal{G}^\perp) = \mathcal{V}(\mathcal{G}^\triangle).$$

In particular, defining $\mathcal{F}(\hat{x},\hat{t}) \coloneqq \mathcal{G}(\hat{x},\hat{t})^\triangle$, we may rewrite the condition of Theorem 4 as the assumption that for every $(\hat{x},\hat{t}) \in \mathcal{D}_{\rm SDP} \setminus \mathcal{D}$,

$$\operatorname{aff}\left(\Pi_{\mathcal{V}(\mathcal{F}(\hat{x},\hat{t}))}\left\{(A_{\rm obj} + A(\gamma))\hat{x} + (b_{\rm obj} + b(\gamma)) :\, (1,\gamma) \in \mathcal{G}^\perp\right\}\right) \neq \mathcal{V}(\mathcal{F}(\hat{x},\hat{t})).$$

One may compare this condition with the condition in Theorem 3. We additionally conjecture that $\mathcal{V}(\mathcal{G}^\perp) = \mathcal{V}(\mathcal{G}^\triangle)$ even without the facially exposed assumption. $\square$

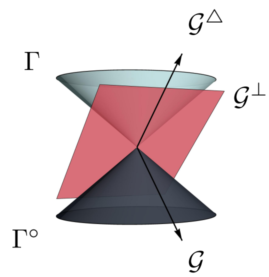

Figure 2.3: This figure plots Γ, Γ°, $\mathcal{G}$, $\mathcal{G}^{\triangle}$, and $\mathcal{G}^{\perp}$ from Example 8. Note that $\mathcal{G}^{\triangle} \subseteq \operatorname{span}(\mathcal{G}^{\triangle}) \subsetneq \mathcal{G}^{\perp}$. Specifically, $\mathcal{G}^{\perp}$ contains additional directions tangent to Γ at $\mathcal{G}^{\triangle}$.

Remark 12. The main results of [100] show that the sufficient condition presented in Theorem 4 is in fact also *necessary* in the setting where Γ° is *facially exposed* (see [100, Theorems 1 and 2]). The condition that Γ° is facially exposed holds, for example, when Γ° is a (slice of the) nonnegative cone, second-order cone, or the positive semidefinite cone. See [79] for a longer discussion of this assumption and its connections to the *nice cones*. In general, all nice cones are facially exposed. □

We close this section with two examples that illustrate the use of Theorem 4.

Example 9 (A mixed-binary set)**.** Consider the following toy example. Define

$$\mathcal{D} := \left\{ (x, y, t) : \begin{array}{l} x^2 \leq t \\ x(1-y) = 0 \\ y(1-y) = 0 \end{array} \right\}.$$

In words, y is a binary variable and x is constrained to be zero whenever y is "off." Furthermore, $t \geq x^2$. The convex hull of this set is well known to be given by a perspective reformulation trick [27, 41, 48]. On the other hand, it is also possible to show that $\operatorname{conv}(\mathcal{D}) = \mathcal{D}_{\mathrm{SDP}}$ using Theorem 4 (see [100, Section 4.3] for details). Specifically, for this example the set Γ (whence also Γ°) is given by a rotated second-order cone. Verifying that the conditions of Theorem 4 hold is then a slightly tedious but ultimately straightforward task. □

Example 10 (QMP)**.** Recall the quadratic matrix programming framework previously considered in Example 7. Specifically, suppose that for each $i \in [m]$, we can write

$$A_i = I_k \otimes \mathbb{A}_i$$

for some $\mathbb{A}_i \in \mathbb{S}^{n/k}$. Similarly suppose $A_{\mathrm{obj}} = I_k \otimes \mathbb{A}_{\mathrm{obj}}$. Previously, we saw that $\operatorname{conv}(\mathcal{D}) = \mathcal{D}_{\mathrm{SDP}}$ if Γ is polyhedral and k satisfies the lower bound in (2.15). In the case of *arbitrary* Γ, using Theorem 4, [100, Section 4.1] shows that that $\operatorname{conv}(\mathcal{D}) = \mathcal{D}_{\mathrm{SDP}}$ whenever $k \geq m$; see also [101]. The QMP setting was previously considered by Beck [10] who proved that $\mathrm{Opt} = \mathrm{Opt}_{\mathrm{SDP}}$ whenever $k \geq m$. See [12] for further results on exactness in quadratic matrix programming. □

Chapter 3

Rank-one-generated cones

This chapter is based on joint work [5] with C.J. Argue and Fatma Kılınç-Karzan.

We say that a closed convex conic subset (in the following, simply a *conic subset*) of the positive semidefinite (PSD) cone is ROG if it is equal to the convex hull of its rank-one matrices. As an immediate example, the PSD cone itself is ROG. One may compare the ROG property of a conic subset of the positive semidefinite cone with the integrality property of a polytope. In both cases, the property states that the convex set in question is the convex hull of a nonconvex set of interest.

In this chapter, we present work from [5, 60] examining the question of what the ROG property of a conic subset of the positive semidefinite cone corresponds to in terms of its defining linear matrix inequalities (LMIs). We will present connections between the ROG property and various forms of exactness (e.g., objective value exactness and convex hull exactness), a number of new sufficient conditions for the ROG property, and a complete characterization of the ROG property for conic subsets of the PSD cone defined by two LMIs.

While we will not formally investigate Theorem 6 until Section 3.7, we will nonetheless mention some of its corollaries before Section 3.7. The interested reader is encouraged to take a look at the statement of Theorem 6 ahead of Remark 16 and Example 13.

3.1 Related work

In contrast to well-known sufficient conditions, e.g., total unimodularity or total dual integrality, for the integrality property of polyhedra (see [31] and references therein), the research on the ROG property is much more recent and limited. Indeed, this property until recently had only been studied *incidentally* to other research questions. A series of works in the matrix completion literature [3, 47, 80] show that the set of positive semidefinite matrices with a fixed chordal support is ROG. The celebrated S-lemma [107] (see also [93]) can be interpreted as saying that the intersection of the positive semidefinite cone with a single linear matrix inequality (LMI) is ROG. A closely related line of work gives explicit descriptions of the ROG cones related to quadratic programs over triangles, tetrahedra, and quadrilaterals [4], and ellipsoids with missing caps [22]; see also the excellent survey paper [19]. More recently, Blekherman et al. [18], Hildebrand [50] study algebraic properties of ROG cones obtained by adding linear matrix equalities (LMEs) to the positive semidefinite cone.

3.2 Preliminaries

Let $\mathcal{M} \subseteq \mathbb{S}^n$ and define

$$\mathcal{S}(\mathcal{M}) \coloneqq \left\{ Z \in \mathbb{S}^n_+ : \langle M, Z\rangle \leq 0, \forall M \in \mathcal{M} \right\}.$$

In contrast to Chapter 2, we will no longer assume that $\mathcal{M}$ is finite.

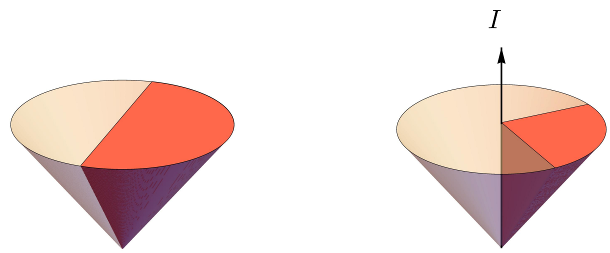

Figure 3.1: Every point in the interior of $\mathbb{S}^2_+$ has rank two and every point on the boundary of $\mathbb{S}^2_+$ has rank at most one. The set on the left, $\mathcal{S}(\{M_1\})$, is ROG as it is equal to the convex hull of its rank-one matrices. The set on the right, $\mathcal{S}(\mathcal{M})$ as defined in the third example of Example 11, is not ROG. Specifically, the identity matrix is not in the convex hull of the rank-one matrices belonging to $\mathcal{S}(\mathcal{M})$ in Example 11(3).

Definition 6. A closed convex cone $\mathcal{S} \subseteq \mathbb{S}^n_+$ is *rank-one generated* (ROG) if

$$\mathcal{S} = \operatorname{conv}(\mathcal{S} \cap \left\{zz^\top : z \in \mathbb{R}^n\right\}). \qquad \square$$

Remark 13. Note that for a closed convex cone $\mathcal{S} \subseteq \mathbb{S}^n_+$, we have $\operatorname{conv}(\mathcal{S} \cap \left\{zz^\top : z \in \mathbb{R}^n\right\}) = \operatorname{clconv}(\mathcal{S} \cap \left\{zz^\top : z \in \mathbb{R}^n\right\})$. $\square$

Example 11. The material in this chapter is best understood with a few examples and nonexamples of ROG cones in mind:

1. By the spectral theorem, the PSD cone $\mathbb{S}^n_+$ is ROG. In particular $\mathcal{S}(\varnothing)$ is ROG.
2. A well-known result says that $\mathcal{S}(\mathcal{M})$ is ROG whenever $|\mathcal{M}| = 1$ (see [93] and Lemma 10). We illustrate this fact in $\mathbb{S}^2_+$ in Figure 3.1.
3. There exist $\mathcal{M}$ with $|\mathcal{M}| = 2$ such that $\mathcal{S}(\mathcal{M})$ is not ROG. Specifically, consider $M_1 = \begin{pmatrix}1 & \\ & -1\end{pmatrix}$ and $M_2 = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$. It is clear that $I \in \mathcal{S}(\mathcal{M})$. We claim that $I \notin \operatorname{conv}\left(\mathcal{S}(\mathcal{M}) \cap \left\{zz^\top : z \in \mathbb{R}^2\right\}\right)$. Indeed, supposing otherwise, we deduce from $\langle M_1, I\rangle = 0$ and $\langle M_2, I\rangle = 0$ that I is a convex combination of rank-one matrices $zz^\top$ satisfying $0 = \left\langle M_1, zz^\top\right\rangle = z_1^2 - z_2^2$ and $0 = \left\langle M_2, zz^\top\right\rangle = 2z_1z_2$. This is a contradiction as the only vector $z \in \mathbb{R}^2$ satisfying $z_1^2 - z_2^2 = 2z_1z_2 = 0$ is the zero vector. We illustrate this fact in Figure 3.1. $\square$

We will observe in Section 3.3 that the ROG property can be used to derive both objective value and convex hull exactness results. For example, we will see that if the ROG property holds for the cone $\mathcal{S}(\mathcal{M})$ corresponding to a set of quadratic constraints $\mathcal{M} \subseteq \mathbb{S}^{n+1}$, then objective value exactness holds for *every* choice of objective function $M_{\mathrm{obj}} \in \mathbb{S}^{n+1}$ for which the SDP is finite.

Recall the following definition.

Definition 7. For $Z \in \mathbb{S}^n$ nonzero, let $\mathbb{R}_+ Z \coloneqq \{\alpha Z : \alpha \geq 0\}$ denote the *ray spanned by* Z. We say that $\mathbb{R}_+ Z$ is an *extreme ray* of $\mathcal{S}$ if for any $X, Y \in \mathcal{S}$ such that $Z = (X+Y)/2$, we have $X, Y \in \mathbb{R}_+ Z$. $\square$

The following lemma gives an alternate characterization of ROG cones in terms of its extreme rays.

Lemma 2 ([5, Lemma 1])**.** *Let $\mathcal{S} \subseteq \mathbb{S}^n_+$ be a closed convex cone. Then, $\mathcal{S}$ is ROG if and only if for each extreme ray $\mathbb{R}_+ Z$ of $\mathcal{S}$, we have* $\operatorname{rank}(Z) = 1$.

In contrast to the above characterization, which relates the ROG property of a cone to primal properties (e.g., the rank of its extreme rays), the remainder of this chapter will be concerned with understanding the ROG property of a cone $\mathcal{S}(\mathcal{M})$ in terms of its defining inequalities $\mathcal{M}$.

Remark 14. The ROG property is also relevant in the context of sum-of-squares (SOS) programming. Let

$$V \coloneqq \left\{x \in \mathbb{R}^n : x^\top A_i x = 0,\ \forall i \in [m]\right\}.$$

It is possible to show that "every quadratic form $A \in \mathbb{S}^n$ that is nonnegative on V is immediately nonnegative"[1] if and only if

$$\left\{X \in \mathbb{S}^n_+ : \langle A_i, X\rangle = 0, \forall i \in [m]\right\}$$

is ROG. See [18, Section 6] for additional applications and connections of the ROG property in real algebraic geometry and statistics. ☐

3.3 Connections to objective value exactness and convex hull exactness

The ROG property of a cone $\mathcal{S}(\mathcal{M})$ is intimately related to exactness results for both homogeneous and inhomogeneous QCQPs and their relaxations.

Objective value exactness and the ROG property

We begin with objective value exactness results based on the ROG property. To this end, the following lemma (see [5, Lemma 19]) says that a cone $\mathcal{S} \subseteq \mathbb{S}^n_+$ is ROG if and only if the SDP relaxation of the corresponding homogeneous QCQP is exact for all choices of objective function.

Lemma 3. *Let $\mathcal{A} \subseteq \mathbb{S}^n$. Then $\mathcal{S}(\mathcal{A})$ is ROG if and only if for every $A_{\text{obj}} \in \mathbb{S}^n$,*

$$\inf_{x\in\mathbb{R}^n} \left\{\left\langle A_{\text{obj}}, xx^\top\right\rangle : xx^\top \in \mathcal{S}(\mathcal{A})\right\} = \inf_{X\in\mathbb{S}^n} \left\{\langle A_{\text{obj}}, X\rangle : X \in \mathcal{S}(\mathcal{A})\right\}. \tag{3.1}$$

Lemma 3 is closely related to S-lemma type results. Recall that the S-lemma, which can be traced back to [36, 107], states the following: Let $A_1 \in \mathbb{S}^n$ such that for some $x \in \mathbb{R}^n$, we have $x^\top A_1 x < 0$. Then, for all $A_{\text{obj}} \in \mathbb{S}^n$,

$$x^\top A_1 x \le 0 \implies x^\top A_{\text{obj}} x \ge 0$$

if and only if there exists $\gamma_1 \ge 0$ such that

$$A_{\text{obj}} + \gamma_1 A_1 \succeq 0.$$

In words, the S-lemma gives conditions under which any homogeneous quadratic consequence inequality has a PSD certificate. See also the survey article [82] on the S-lemma and its variants.

Lemma 3 allows us to write a variant of the S-lemma in the setting of ROG cones. Specifically, let $\mathcal{A} \coloneqq \{A_1, \dots, A_m\} \subseteq \mathbb{S}^n$ and suppose there exists $x \in \mathbb{R}^n$ such that $x^\top A_i x < 0$ for all $i \in [m]$. This assumption, a Slater condition, ensures that strong duality holds between the following SDP and its dual:

$$\inf_{X\in\mathbb{S}^n} \left\{\langle A_{\text{obj}}, X\rangle : X \in \mathcal{S}(\mathcal{A})\right\} = \sup_{\gamma\in\mathbb{R}^m_+} \left\{0 : A_{\text{obj}} + \sum_{i=1}^m \gamma_i A_i \succeq 0\right\}. \tag{3.2}$$

In particular, (3.2) takes the value 0 if and only if there exists $\gamma \in \mathbb{R}^m_+$ such that $A_{\text{obj}} + \sum_{i=1}^m A_i \succeq 0$. On the other hand,

$$\inf_{x\in\mathbb{R}^n} \left\{x^\top A_{\text{obj}} x : xx^\top \in \mathcal{S}(\mathcal{A})\right\}$$

takes the value 0 if and only if $x^\top A_{\text{obj}} x \ge 0$ for every $x \in \mathbb{R}^n$ such that $x^\top A_i x \le 0$ for all $i \in [m]$.

The following statement then follows from the above observations and Lemma 3.

Corollary 1. *Let $\mathcal{A} \coloneqq \{A_1, \dots, A_m\} \subseteq \mathbb{S}^n$ and suppose there exists $\bar{x} \in \mathbb{R}^n$ such that $\bar{x}^\top A_i \bar{x} < 0$ for all $i \in [m]$. Then, the following are equivalent*

[1] Formally, this is the property that $x^\top A x \ge 0$ for all $x \in V \implies A \in \mathbb{S}^n_+ + \text{span}(\{A_i\})$.

- *For all* $A_{\text{obj}} \in \mathbb{S}^n$*,*

$$x^\top A_i x \le 0, \forall i \in [m] \quad \implies \quad x^\top A_{\text{obj}} x \ge 0$$

if and only if there exists $\gamma \in \mathbb{R}^m_+$ *such that*

$$A_{\text{obj}} + \sum_{i=1}^m \gamma_i A_i \succeq 0.$$

- $\mathcal{S}(\mathcal{A})$ *is ROG.*

In words, under a Slater condition, the S-lemma variant for $\mathcal{A}$ and an arbitrary A_{obj} holds if and only if $\mathcal{S}(\mathcal{A})$ is ROG.

Let us now examine the relation in the case of a general inequality-constrained QCQP and its SDP relaxation. Note that by introducing a homogenizing variable, a general QCQP (2.1) and its SDP relaxation (2.4) can be rewritten as

$$\inf_{z\in\mathbb{R}^{n+1}} \left\{ \left\langle M_{\text{obj}}, zz^\top \right\rangle : \begin{array}{l} zz^\top \in \mathcal{S}(\mathcal{M}) \\ \left\langle e_{n+1}e_{n+1}^\top, zz^\top \right\rangle = 1 \end{array} \right\} \ge \inf_{Z\in\mathbb{S}^n} \left\{ \left\langle M_{\text{obj}}, Z \right\rangle : \begin{array}{l} Z \in \mathcal{S}(\mathcal{M}) \\ \left\langle e_{n+1}e_{n+1}^\top, Z \right\rangle = 1 \end{array} \right\}.$$

In particular, we may always rewrite our QCQPs to contain exactly one inhomogeneous equality constraint. The following result (see [5, Lemma 20]) relates the ROG property of a cone to SDP exactness results for its corresponding inhomogeneous QCQPs.

Lemma 4. *Let* $\mathcal{M} \subseteq \mathbb{S}^{n+1}$ *and* $B \in \mathbb{S}^{n+1}$*. If* $\mathcal{S}(\mathcal{M})$ *is ROG, then*

$$\inf_{z\in\mathbb{R}^{n+1}} \left\{ \left\langle M_{\text{obj}}, zz^\top \right\rangle : \begin{array}{l} zz^\top \in \mathcal{S}(\mathcal{M}), \\ \left\langle B, zz^\top \right\rangle = 1 \end{array} \right\} = \inf_{Z\in\mathbb{S}^{n+1}} \left\{ \left\langle M_{\text{obj}}, Z \right\rangle : \begin{array}{l} Z \in \mathcal{S}(\mathcal{M}), \\ \left\langle B, Z \right\rangle = 1 \end{array} \right\} \tag{3.3}$$

for all $M_{\text{obj}} \in \mathbb{S}^{n+1}$ *for which the optimum SDP objective value is bounded from below. In particular, this equality holds whenever the SDP feasible domain is bounded.*

By taking $B = e_{n+1}e_{n+1}^\top$ in Lemma 4, we have that objective value exactness holds whenever $\mathcal{S}(\mathcal{M})$ is ROG and the SDP optimum value is bounded from below.

The freedom to pick $B \ne e_{n+1}e_{n+1}^\top$ in Lemma 4 will be useful in Section 3.8 where we will use it to analyze the problem of minimizing a ratio of quadratic functions over a quadratically constrained domain.

Remark 15. We make a few additional observations about Lemma 4. First, Lemma 4 extends [50, Lemma 1.2], which shows that the same statement holds in the case of finitely many linear matrix equalities. The proof in [5] also differs from the proof in [50] as it immediately shows how to construct a QCQP feasible solution achieving the SDP value (or a sequence approaching the SDP value). Next, we highlight that the reverse implication in Lemma 4 is not true in general. Specifically, there exist cones $\mathcal{S}(\mathcal{M})$ for which equality holds in (3.3) for all M_{obj} for which the SDP objective value is bounded below that are *not* ROG; see [5, Example 4]. Finally, it is possible to show that the boundedness assumption in Lemma 4 cannot be dropped even in the case where B is specialized to $B = e_{n+1}e_{n+1}^\top$; see [5, Example 5]. □

As in Corollary 1, we may use Lemma 4 to derive the following ROG-based variant of the inhomogeneous S-lemma.

Corollary 2. *Let* $\mathcal{M} := \{M_1, \dots, M_m\} \subseteq \mathbb{S}^{n+1}$ *and* $M_{\text{obj}} \in \mathbb{S}^{n+1}$ *be such that*

- $\mathcal{S}(\mathcal{M})$ *is ROG,*
- *there exists* $\bar{x} \in \mathbb{R}^n$ *such that* $q_i(\bar{x}) < 0$ *for all* $i \in [m]$*, and*
- *there exists* $\bar{\gamma} \in \mathbb{R}^m_+$ *and* $\bar{\alpha} \in \mathbb{R}$ *such that* $M_{\text{obj}} + \sum_{i=1}^m \bar{\gamma}_i M_i - \bar{\alpha} e_{n+1}e_{n+1}^\top \in \mathbb{S}^{n+1}_+$*.*

Then,

$$q_i(x) \leq 0,\ \forall i \in [m] \quad \Longrightarrow \quad q_{\text{obj}}(x) \geq \alpha$$

if and only if there exists $\gamma \in \mathbb{R}^m_+$ *such that*

$$q_{\text{obj}}(x) + \sum_{i=1}^m \gamma_i q_i(x) \geq \alpha.$$

In words, under a Slater condition and dual feasibility assumption, the inhomogeneous S-lemma variant for $\mathcal{M}$ and an arbitrary M_{obj} holds whenever $\mathcal{S}(\mathcal{M})$ is ROG.

Convex hulls of quadratically constrained sets

The following proposition (see [5, Proposition 6]) states that the ROG property of $\mathcal{S}(\mathcal{M})$ guarantees that the closed convex hull of the epigraph of a QCQP with constraints defined by $\mathcal{M}$ is given by its SDP relaxation. As before, we will make use of a definiteness assumption.

Proposition 1. *Let* $M_{\text{obj}} \in \mathbb{S}^{n+1}$ *and* $\mathcal{M} \subseteq \mathbb{S}^{n+1}$. *Suppose there exists* $M^* = \begin{pmatrix} A^* & b^* \\ b^{*\top} & c^* \end{pmatrix} \in \operatorname{clcone}(\{M_{\text{obj}}\} \cup \mathcal{M})$ *such that* $A^* \succ 0$. *If* $\mathcal{S}(\mathcal{M})$ *is ROG, then* $\operatorname{clconv}(\mathcal{D}) = \mathcal{D}_{\text{SDP}}$.

Note that when $\mathcal{M}$ is finite, the assumption that M^* exists is equivalent to Assumption 1. We can further relax this assumption by applying a perturbation argument to arrive at the following result (see [5, Corollary 6]).

Corollary 3. *Let* $M_{\text{obj}} \in \mathbb{S}^{n+1}$ *and* $\mathcal{M} \subseteq \mathbb{S}^{n+1}$. *Suppose there exists* $M^* = \begin{pmatrix} A^* & b^* \\ b^{*\top} & c^* \end{pmatrix} \in \operatorname{clcone}(\{M_{\text{obj}}\} \cup \mathcal{M})$ *such that* $A^* \succeq 0$. *If* $\mathcal{S}(\mathcal{M})$ *is ROG, then* $\operatorname{clconv}(\mathcal{D}) = \operatorname{cl}(\mathcal{D}_{\text{SDP}})$.

The following example recovers [99, Theorem 4] (see also Example 5) as an immediate application of Corollary 3.

Example 12 (GTRS)**.** Recall that the GTRS asks us to minimize a quadratic objective function $q_{\text{obj}}(x)$ subject to a single quadratic inequality constraint $q_1(x) \leq 0$. Here, we assume that q_{obj}, q_1 are such that there exists $\gamma^* \geq 0$ such that $A_{\text{obj}} + \gamma^* A_1 \succeq 0$. It is well known (see for example [93]) that the cone $\mathcal{S}(\{M_1\})$ is ROG. We then deduce by Corollary 3 that

$$\operatorname{clconv}\left\{(x,t) \in \mathbb{R}^n \times \mathbb{R} : \begin{array}{l} q_{\text{obj}}(x) \leq t \\ q_1(x) \leq 0 \end{array}\right\} = \operatorname{cl}\left\{(x,t) \in \mathbb{R}^n \times \mathbb{R} : \begin{array}{l} \exists X \succeq xx^\top : \\ \langle A_{\text{obj}}, X\rangle + 2\langle b_{\text{obj}}, x\rangle + c_{\text{obj}} \leq t \\ \langle A_1, X\rangle + 2\langle b_1, x\rangle + c_1 \leq 0 \end{array}\right\}. \square$$

We close this subsection by demonstrating that the ROG property is (unsurprisingly) strictly stronger than convex hull exactness.

Remark 16. Consider the following QCQP

$$\inf_{x \in \mathbb{R}^2} \left\{ \|x\|^2 : \begin{array}{l} x^\top \begin{pmatrix} -1 & \\ & 1 \end{pmatrix} x - 1 \leq 0 \\ x^\top \begin{pmatrix} 2 & \\ & -1 \end{pmatrix} x - 1 \leq 0 \end{array} \right\}.$$

The corresponding set $\mathcal{M}$ for this example is $\mathcal{M} = \{\operatorname{Diag}(-1,1,-1), \operatorname{Diag}(2,-1,-1)\}$. We will soon see in Theorem 6 that $\mathcal{S}(\mathcal{M})$ is not ROG so that Proposition 1 cannot be applied to this example. On the other hand, the set of convex Lagrange multipliers Γ for this QCQP is polyhedral so that we may apply the analysis of Example 7 to deduce that convex hull exactness holds. $\square$

$$\boxed{\mathcal{M} \text{ is finite and } \forall \mathcal{M}' \subseteq \mathcal{M},\ \mathcal{T}(\mathcal{M}') \text{ ROG}} \implies \boxed{\mathcal{S}(\mathcal{M}) \text{ ROG}} \implies \boxed{\mathcal{T}(\mathcal{M}) \text{ ROG}}$$

Figure 3.2: A summary of Lemma 7 and Corollary 4

3.4 Relating LMIs to LMEs

In this section, we will present a series of lemmas relating the ROG property of $\mathcal{S}(\mathcal{M})$ to that of

$$\mathcal{T}(\mathcal{M}') := \left\{ Z \in \mathbb{S}^{n+1}_+ : \langle M, Z \rangle = 0,\ \forall M \in \mathcal{M}' \right\}$$

for $\mathcal{M}' \subseteq \mathcal{M}$. Note that in contrast to $\mathcal{S}(\mathcal{M})$, the definition of $\mathcal{T}(\mathcal{M}')$ uses linear matrix *equalities* (LMEs). These results will be particularly useful for analyzing spectrahedral sets defined by finitely many LMIs/LMEs.

Remark 17. For any $\mathcal{M} \subseteq \mathbb{S}^{n+1}$, we have $\mathcal{S}(\mathcal{M}) = \mathcal{S}(\operatorname{clcone}(\mathcal{M}))$ and $\mathcal{T}(\mathcal{M}) = \mathcal{T}(\operatorname{span}(\mathcal{M}))$. Consequently, we may without loss of generality assume that $\mathcal{M}$ is finite when analyzing sets of the form $\mathcal{T}(\mathcal{M})$—simply replace $\mathcal{M}$ with a finite basis of $\operatorname{span}(\mathcal{M})$. On the other hand, $\operatorname{clcone}(\mathcal{M})$ is not necessarily finitely generated. ☐

The following result from [5, Lemma 3] relates the facial structure of $\mathcal{S}(\mathcal{M})$ to the ROG property. This result is analogous to the statement that a polytope is integral if and only if each of its faces is integral.

Lemma 5. *For any set $\mathcal{M} \subseteq \mathbb{S}^{n+1}$, the following are equivalent:*

1. *$\mathcal{S}(\mathcal{M})$ is ROG.*
2. *Every face of $\mathcal{S}(\mathcal{M})$ is ROG.*
3. *$\mathcal{S}(\mathcal{M}) \cap \mathcal{T}(\mathcal{M}')$ is ROG for every $\mathcal{M}' \subseteq \mathcal{M}$.*

We have the following immediate corollary of Lemma 5; see [5, Corollary 1].

Corollary 4. *For any set $\mathcal{M} \subseteq \mathbb{S}^{n+1}$, if $\mathcal{S}(\mathcal{M})$ is ROG then $\mathcal{T}(\mathcal{M})$ is ROG.*

We can strengthen Lemma 5 in a few ways (see [5, Lemmas 5 and 6]).

Lemma 6. *Let $\mathcal{M} \subseteq \mathbb{S}^{n+1}$ be compact. Then, $\mathcal{S}(\mathcal{M})$ is ROG if and only if $\mathcal{S}(\mathcal{M}) \cap \mathcal{T}(\mathcal{M}')$ is ROG for every $\varnothing \neq \mathcal{M}' \subseteq \mathcal{M}$.*

Lemma 7. *Let $\mathcal{M} \subseteq \mathbb{S}^{n+1}$ be finite. If $\mathcal{T}(\mathcal{M}')$ is ROG for every $\mathcal{M}' \subseteq \mathcal{M}$, then $\mathcal{S}(\mathcal{M})$ is ROG.*

We next note ([5, Lemma 7]) that the ROG property of $\mathcal{T}(\mathcal{M})$ is equivalent to the ROG property of $\mathcal{T}(\overline{\mathcal{M}})$ where $\overline{\mathcal{M}}$ is the restriction of $\mathcal{M}$ onto the joint range of the matrices $M \in \mathcal{M}$.

Lemma 8. *Let $W := \operatorname{span}\left(\bigcup_{M\in\mathcal{M}} \operatorname{range}(M)\right)$. For $M \in \mathcal{M}$, let $\overline{M} = M_W$ denote the restriction of M to W. Let $\overline{\mathcal{M}} = \left\{\overline{M} : M \in \mathcal{M}\right\}$. Then, $\mathcal{T}(\mathcal{M})$ is ROG if and only if $\mathcal{T}(\overline{\mathcal{M}})$ is ROG.*

Considering the definitions of $\mathcal{S}(\mathcal{M})$ and $\mathcal{T}(\mathcal{M})$, one may ask whether it is possible to "lift" the inequalities in $\mathcal{S}(\mathcal{M})$ to equalities while preserving the ROG property. More formally, given a finite set $\mathcal{M} = \{M_1, \dots, M_m\}$ such that $\mathcal{S}(\mathcal{M})$ is ROG, is the set $\mathcal{T}(\overline{\mathcal{M}}) \subseteq \mathbb{S}^{n+m}$ ROG? Here,

$$\overline{\mathcal{M}} = \left\{\overline{M}_1, \dots, \overline{M}_m\right\} \quad \text{and} \quad \overline{M}_i = \begin{pmatrix} M_i & \\ & -e_i e_i^\top \end{pmatrix}.$$

If this were possible, then there would be no need to study the ROG property of sets of the form $\mathcal{S}(\mathcal{M})$. The following example ([5, Example 3]) shows that this is not possible in general.

Example 13. Consider the set

$$\mathcal{S} := \left\{ Z \in \mathbb{S}^3_+ : \begin{array}{l} Z_{1,2} = 0 \\ Z_{1,3} \leq 0 \end{array} \right\}.$$

$$\begin{array}{ccc}
\boxed{\mathcal{S}(\mathcal{M})\text{ ROG}} & \Longleftrightarrow & \boxed{\forall Z \in \mathcal{S}(\mathcal{M}) \setminus \{0\}\,,\ \mathrm{range}(Z) \cap \mathcal{E}(Z, \mathcal{M}) \neq \{0\}} \\
\Downarrow & & \Downarrow \\
\boxed{\mathcal{T}(\mathcal{M})\text{ ROG}} & \Longleftrightarrow & \boxed{\forall Z \in \mathcal{T}(\mathcal{M}) \setminus \{0\}\,,\ \mathrm{range}(Z) \cap \mathcal{N}(\mathcal{M}) \neq \{0\}}
\end{array}$$

Figure 3.3: A summary of Lemma 9 and Corollary 5.

We will see soon (see Theorem 6) that this set is ROG. We can replace the LMIs defining $\mathcal{S}$ with LMEs in a lifted space as follows: Let $\Pi : \mathbb{S}^4 \to \mathbb{S}^3$ denote the projection of a 4×4 matrix onto its top-left 3×3 principal submatrix. Then,

$$\mathcal{S} = \Pi\left(\left\{ Z \in \mathbb{S}^4 : \begin{array}{l} Z_{1,2} = 0 \\ Z_{1,3} + Z_{4,4} = 0 \end{array} \right\}\right) = \Pi\left(\mathcal{T}(\{\overline{M}_1, \overline{M}_2\})\right),$$

where

$$\overline{M}_1 := \begin{pmatrix} 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \overline{M}_2 := \begin{pmatrix} 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Define $\overline{\mathcal{M}} := \{\overline{M}_1, \overline{M}_2\}$. Theorem 6 implies that $\mathcal{T}(\overline{\mathcal{M}})$ is not ROG. We conclude that the obvious lifting of LMIs into LMEs can take ROG sets $\mathcal{S}(\mathcal{M})$ to non-ROG sets $\mathcal{T}(\overline{\mathcal{M}})$. □

3.5 The ROG property and solutions of quadratic systems

The ROG property of a set is naturally connected to the existence of nonzero solutions of underlying systems of quadratic inequality or equality constraints. We examine this connection next. To this end, for $\mathcal{M} \subseteq \mathbb{S}^{n+1}$ and $Z \in \mathcal{S}(\mathcal{M})$, define

$$\mathcal{E}(Z, \mathcal{M}) := \left\{ z \in \mathbb{R}^{n+1} : \langle M, Z \rangle \leq z^\top M z \leq 0,\ \forall M \in \mathcal{M} \right\}.$$

Based on this, we have the following characterization of the ROG property [5, Lemma 11].

Lemma 9. *$\mathcal{S}(\mathcal{M})$ is ROG if and only if for every nonzero $Z \in \mathcal{S}(\mathcal{M})$ we have* $\mathrm{range}(Z) \cap \mathcal{E}(Z, \mathcal{M}) \neq \{0\}$.

Note that by definition, $\mathcal{T}(\mathcal{M}) = \mathcal{S}(\mathcal{M}')$ where $\mathcal{M}' := \{M : M \in \mathcal{M}\} \bigcup \{-M : M \in \mathcal{M}\}$. Therefore, in the case of $\mathcal{T}(\mathcal{M})$, the set $\mathcal{E}(Z, \mathcal{M})$ in Lemma 9 is replaced with a simpler set corresponding to solutions to a homogeneous system of quadratic equations. In particular, for $\mathcal{M} \subseteq \mathbb{S}^{n+1}$, define

$$\mathcal{N}(\mathcal{M}) := \left\{ z \in \mathbb{R}^{n+1} : z^\top M z = 0,\ \forall M \in \mathcal{M} \right\}.$$

Remark 18. It is easy to see that for every $\mathcal{M} \subseteq \mathbb{S}^{n+1}$ and $Z \in \mathcal{S}(\mathcal{M})$, we have $\mathcal{N}(\mathcal{M}) \subseteq \mathcal{E}(Z, \mathcal{M})$. □

Therefore, we arrive at the following corollary; see [5, Corollary 2].

Corollary 5. *$\mathcal{T}(\mathcal{M})$ is ROG if and only if for every nonzero $Z \in \mathcal{T}(\mathcal{M})$ we have* $\mathrm{range}(Z) \cap \mathcal{N}(\mathcal{M}) \neq \{0\}$.

Note that for any rank-one matrix $Z = zz^\top$ in $\mathcal{S}(\mathcal{M})$, we always have $z \in \mathrm{range}(Z) \cap \mathcal{E}(Z, \mathcal{M})$. Therefore, when applying Lemma 9, it suffices to check the right hand side only for matrices Z with rank at least two. The same is true for Corollary 5.

Our tool set allows us to recover a number of known results from the literature. The following result regarding spectrahedral cones defined by a single LMI is from Sturm and Zhang [93].

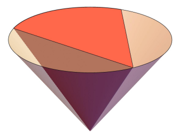 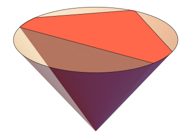

Figure 3.4: Illustrations of Lemma 11 (left picture) and Proposition 2 (right picture) in the setting $\mathbb{S}^{n+1}_+ = \mathbb{S}^2_+$. Recall that the matrices on the boundary of $\mathbb{S}^2_+$ all have rank at most one.

Lemma 10. *Consider any $M \in \mathbb{S}^{n+1}$, and let $\mathcal{M} = \{M\}$. Then $\mathcal{S}(\mathcal{M})$ is ROG.*

Using our tool set we see that Lemma 10 follows from Lemma 6 and Corollary 5. In the case of ROG sets defined by two LMIs, Lemmas 7 and 10 and Corollary 4 lead to the following characterization (see [5, Corollary 3]).

Corollary 6. *Suppose $|\mathcal{M}| = 2$, then $\mathcal{S}(\mathcal{M})$ is ROG if and only if $\mathcal{T}(\mathcal{M})$ is ROG.*

In Section 3.7 we will present necessary and sufficient conditions for a $\mathcal{T}(\{M_1, M_2\})$ to be ROG; see Theorem 6. Combined with Corollary 6, this then gives a complete characterization of ROG sets of the form $\mathcal{S}(\{M_1, M_2\})$.

3.6 Sufficient conditions

In this section, we present a number of sufficient conditions for the ROG property. We begin with a result related to the S-lemma [40] and a convexity theorem due to Dines [36].

Lemma 11. *Let $\mathcal{M} = \{M_1, M_2\}$ and suppose there exists $(\alpha_1, \alpha_2) \neq (0, 0)$ such that $\alpha_1 M_1 + \alpha_2 M_2 \in \mathbb{S}^{n+1}_+$. Then, $\mathcal{S}(\mathcal{M})$ is ROG.*

The proof of this lemma is straightforward and follows from Corollary 6.

Remark 19. The condition in Lemma 11 has a simple geometric interpretation. Specifically, this condition guarantees that the two LMEs defining $\mathcal{T}(\{M_1, M_2\})$ only interact with each other on a single (possibly trivial) face of the positive semidefinite cone. Furthermore, on this face, the two LMEs impose the same (possibly trivial) constraint. See Figure 3.4. □

The following result from [5, Proposition 1] generalizes Lemma 11.

Proposition 2. *Let $\mathcal{M} \subseteq \mathbb{S}^{n+1}$ be finite. Suppose that for every distinct pair $M, M' \in \mathcal{M}$, there exists $(\alpha, \beta) \neq (0, 0)$ such that $\alpha M + \beta M'$ is positive semidefinite. Then, $\mathcal{S}(\mathcal{M})$ is ROG.*

Intuitively, the conditions in this proposition have a similar geometric interpretation to the conditions in Lemma 11 (see Remark 19). Specifically, it is possible to show that for every $\mathcal{M}' \subseteq \mathcal{M}$, the set $\mathcal{T}(\mathcal{M}')$ is contained in some face of the positive semidefinite cone on which every constraint defining $\mathcal{T}(\mathcal{M}')$ imposes the same constraint. See Figure 3.4.

Lemma 9 and Remark 18 lead to the following sufficient condition for the ROG property [5, Theorem 1].

Theorem 5. *Suppose $\mathcal{M} = \{\operatorname{Sym}(ab^\top) : b \in \mathcal{B}\}$ for some $a \in \mathbb{R}^{n+1}$ and $\mathcal{B} \subseteq \mathbb{R}^{n+1}$. Then, for every positive semidefinite Z of rank at least two, we have $\operatorname{range}(Z) \cap \mathcal{N}(\mathcal{M}) \neq \{0\}$. In particular, $\mathcal{S}(\mathcal{M})$ is ROG.*

Theorem 5 has a few immediate corollaries. The first of these corollaries allows us to handle conic constraints added to the positive semidefinite cone. Let K be a closed convex cone and define $\mathcal{M} := \{\operatorname{Sym}(-cb^\top) : b \in K^*\}$ where K^* is the dual cone of K. Then, $\{Z \in \mathbb{S}^{n+1}_+ : Zc \in K\} = \mathcal{S}(\mathcal{M})$. We arrive at the following corollaries; see [5, Corollaries 4 and 5].

Corollary 7. *Let $K \subseteq \mathbb{R}^{n+1}$ be any closed convex cone and consider an arbitrary vector $c \in \mathbb{R}^{n+1}$. Then, the set $\{Z \in \mathbb{S}^{n+1}_+ : Zc \in K\}$ is ROG.*

Corollary 8. *Let $a, b, c \in \mathbb{R}^{n+1}$. Then, the set $\{Z \in \mathbb{S}^{n+1}_+ : a^\top Zc \geq 0,\ b^\top Zc \geq 0\}$ is ROG.*

As an immediate application of Corollary 8 and Proposition 1, we recover the result presented in Example 9 on the perspective reformulation trick.

Remark 20. Defining $L \coloneqq \mathrm{Diag}(1, \ldots, 1, -1) \in \mathbb{S}^{n+1}$, we can write $\mathbb{L}^{n+1} = \left\{z \in \mathbb{R}^{n+1} : z^\top L z \leq 0,\ z_{n+1} \geq 0\right\}$. Sturm and Zhang [93] (see also [19, Section 6.1]) established that the set

$$\mathcal{S} \coloneqq \left\{ Z \in \mathbb{S}^{n+1}_+ : \begin{array}{c} Zc \in \mathbb{L}^{n+1} \\ \langle L, Z\rangle \leq 0 \end{array} \right\},$$

where $c \in \mathbb{R}^{n+1}$, is ROG (cf. Corollary 7). This result can also be recovered from a straightforward application of Lemma 10 and Corollary 7; see [5, Lemma 12]. □

3.7 Necessary conditions

In this section, we discuss the complete characterization of ROG cones defined by two LMIs or LMEs given in [5, Theorem 3].

Theorem 6. *Let $\mathcal{M} = \{M_1, M_2\}$. Then, $\mathcal{T}(\mathcal{M})$ (and thus $\mathcal{S}(\mathcal{M})$) is ROG if and only if one of the following holds:*

(i) there exists $(\alpha_1, \alpha_2) \neq (0, 0)$ such that $\alpha_1 M_1 + \alpha_2 M_2 \in \mathbb{S}^{n+1}_+$, or

(ii) there exists $a, b, c \in \mathbb{R}^{n+1}$ such that $M_1 = \mathrm{Sym}(ac^\top)$ and $M_2 = \mathrm{Sym}(bc^\top)$.

Note that the *if* direction of Theorem 6 is a direct consequence of the sufficient conditions identified in Proposition 2 and Corollary 8. Furthermore, recall from Corollary 6 that when $|\mathcal{M}| = 2$, the set $\mathcal{S}(\mathcal{M})$ is ROG if and only if $\mathcal{T}(\mathcal{M})$ is ROG. Thus it suffices to show that if $\mathcal{T}(\mathcal{M})$ is ROG then one of the conditions (i) or (ii) must hold.

Remark 21. The conic Gordan–Stiemke Theorem (see Equation 2.3 in [92] and its surrounding comments) implies that for any subspace $W \subseteq \mathbb{S}^{n+1}$,

$$W \cap \mathbb{S}^{n+1}_+ = \{0\} \iff W^\perp \cap \mathbb{S}^{n+1}_{++} \neq \varnothing.$$

In particular, applying the conic Gordan–Stiemke Theorem in the context of Theorem 6 we deduce that if M_1, M_2 are linearly independent, then condition (i) in Theorem 6 fails if and only if $\mathcal{T}(\{M_1, M_2\})$ contains a positive definite matrix. □

Conditions (i) and (ii) in Theorem 6 have simple geometric interpretations. See Remark 19 for a geometric interpretation of (i). Condition (ii) covers the important case when the two LMEs interact in a nontrivial manner inside $\mathbb{S}^{n+1}_+$. Suppose for the sake of presentation that $a = e_1$, $b = e_2$, $c = e_{n+1}$. Then, Corollary 8 implies that

$$\begin{aligned}
\mathcal{T}(\mathcal{M}) &= \mathrm{conv}(\left\{zz^\top : z_1 z_{n+1} = 0,\ z_2 z_{n+1} = 0\right\}) \\
&= \mathrm{conv}\left(\mathrm{conv}\left\{zz^\top : z_1 = z_2 = 0\right\} \cup \mathrm{conv}\left\{zz^\top : z_{n+1} = 0\right\}\right) \\
&= \mathrm{conv}\left((0_2 \oplus \mathbb{S}^{n-1}_+) \cup (\mathbb{S}^n_+ \oplus 0_1)\right).
\end{aligned}$$

In other words, condition (ii) covers the case where $\mathcal{T}(\mathcal{M})$ is the convex hull of the union of two faces of the positive semidefinite cone with a particular intersection structure. Theorem 6 states that these are the only ways for $\mathcal{T}(\mathcal{M})$ to be ROG when $|\mathcal{M}| = 2$.

Remark 22. Both directions of Theorem 6 admit small certificates; see [5, Remark 11]. Let $\mathcal{M} = \{M_1, M_2\}$.

- Suppose $\mathcal{S}(\mathcal{M})$ is ROG. Then Theorem 6 implies that there exists either aggregation weights $(\alpha_1, \alpha_2) \neq (0, 0)$ for which $\alpha_1 M_1 + \alpha_2 M_2 \in \mathbb{S}^{n+1}_+$ or vectors $a, b, c \in \mathbb{R}^{n+1}$ for which $M_1 = \mathrm{Sym}(ac^\top)$ and $M_2 = \mathrm{Sym}(bc^\top)$.

- Suppose $\mathcal{S}(\mathcal{M})$ is not ROG. Then, based on Theorem 6, it suffices to certify that neither conditions (i) nor (ii) hold. Note that M_1 and M_2 are linearly independent since otherwise we would have $\mathcal{S}(\mathcal{M})$ is ROG. Then, from the Gordan–Stiemke Theorem (see Remark 21) we deduce that condition (i) fails if and only if there exists a positive definite matrix Z in $\mathcal{T}(\mathcal{M})$. That is, presenting a positive definite matrix in $\mathcal{T}(\mathcal{M})$ will certify that condition (i) fails. If either $\operatorname{rank}(M_1) \geq 3$ or $\operatorname{rank}(M_2) \geq 3$, then the spectral decomposition of the corresponding M_i certifies that condition (ii) does not hold. Else, M_1 and M_2 are both indefinite rank-two matrices and we can write $M_1 = \eta_1 \operatorname{Sym}(ab^\top)$ and $M_2 = \eta_2 \operatorname{Sym}(cd^\top)$ where $\eta_i \in \mathbb{R}$ and $a, b, c, d \in \mathbb{R}^{n+1}$ are unit vectors. This decomposition is unique up to renaming a and b or c and d. Then, condition (ii) does not hold if and only if a, b, c, d are distinct. In particular, this decomposition certifies that condition (ii) does not hold. $\square$

The proof of Theorem 6 is nontrivial and requires several arguments. We give a proof outline below. See [5, Section 4] for the detailed proof.

Proof outline for Theorem 6. We may assume that

$$\operatorname{span}\left(\operatorname{range}(M_1) \cup \operatorname{range}(M_2)\right) = \mathbb{R}^{n+1}$$

without loss of generality. Indeed, when the set $\{M_1, M_2\}$ does not satisfy this assumption, we may consider the set of restricted matrices $\{(M_1)_W, (M_2)_W\}$ where $(M_i)_W$ is the restriction of M_i onto the minimal subspace, W, containing $\operatorname{range}(M_1) \cup \operatorname{range}(M_2)$. It is not hard to show that the ROG property as well as conditions (i) and (ii) are invariant under this operation (see [5, Lemma 13]).

By Lemma 11 and Corollary 8, it suffices to show that if $\mathcal{T}(\mathcal{M})$ is ROG, then either condition (i) or condition (ii) holds. We will split the proof of Theorem 6 into a number of cases depending on the dimension $n+1$.

- The case $n + 1 = 1$ holds vacuously as we can set (α_1, α_2) to either $(1, 0)$ or $(-1, 0)$ to satisfy (i).
- For $n + 1 = 2$, we can show that condition (i) necessarily holds. Indeed, supposing otherwise, we can explicitly construct a rank-two extreme ray of $\mathcal{T}(\mathcal{M})$ using the Gordan–Stiemke Theorem. The construction crucially uses the geometry of $\mathbb{R}^2$ (and $\mathbb{S}^2$); see [5, Proposition 2].
- For $n + 1 = 3$, when neither conditions (i) nor (ii) are satisfied, we can explicitly construct extreme rays of $\mathcal{T}(\mathcal{M})$ with rank two. The construction is based on understanding what the corresponding $\mathcal{N}(\mathcal{M})$ set looks like. This construction crucially uses the geometry of $\mathbb{R}^3$. In particular, it establishes that in this case where neither conditions (i) nor (ii) are satisfied, $\mathcal{N}(\mathcal{M})$ is the union of at most four one-dimensional subspaces of $\mathbb{R}^3$ (see [5, Lemma 17]).[2] Following this, we may then apply Dine's Theorem [36] and [50, Lemma 3.13] to construct our desired rank-two extreme ray of $\mathcal{T}(\mathcal{M})$ (see [5, Proposition 3]).
- Finally, we will reduce the case where $n+1 \geq 4$ to the case where $n+1 = 3$. Specifically, supposing that $\mathcal{T}(\mathcal{M})$ is a ROG cone with $n + 1 \geq 4$ for which condition (i) does not hold, it is possible to construct a three-dimensional subspace W of $\mathbb{R}^{n+1}$ such that the restriction of $\mathcal{M}$ to W, denoted $\mathcal{M}_W$, satisfies: $\mathcal{T}(\mathcal{M}_W)$ is ROG, neither conditions (i) nor (ii) hold for $\mathcal{M}_W$. This gives us our desired contradiction. See [5, Proposition 4]. ■

Remark 23. The above proof outline in fact shows something stronger than Theorem 6. Specifically, in the cases where

$$\dim\left(\operatorname{span}\left(\operatorname{range}(M_1) \cup \operatorname{range}(M_2)\right)\right) \neq 3,$$

we were able to derive contradictions by simply assuming that condition (i) did not hold. In other words, condition (i) itself completely characterizes the ROG property of a cone defined by two LMIs whenever the dimension of their joint span is not three-dimensional. $\square$

We close by illustrating the proof of Theorem 6 on a prototypical example where the joint span of M_1 and M_2 is three-dimensional; see [5, Example 2].

[2] For readers familiar with algebraic geometry, this may be viewed as a consequence of Bézout's theorem.

Example 14. Suppose $\mathcal{M} = \{M_1, M_2\}$ where $M_1 = \mathrm{Diag}(1, -1, 0)$ and $M_2 = \mathrm{Diag}(0, 1, -1)$ so that

$$\mathcal{T}(\mathcal{M}) = \left\{Z \in \mathbb{S}^3_+ : Z_{1,1} = Z_{2,2} = Z_{3,3}\right\}.$$

Note that $\alpha_1 M_1 + \alpha_2 M_2 = \mathrm{Diag}(\alpha_1, \alpha_2 - \alpha_1, -\alpha_2)$ is positive semidefinite if and only if $(\alpha_1, \alpha_2) = (0, 0)$ so that condition (i) of Theorem 6 is violated. Next, we claim that condition (ii) of Theorem 6 does not hold. Indeed, assuming condition (ii), we have that $\alpha_1 M_1 + \alpha_2 M_2 = \mathrm{Sym}((\alpha_1 a + \alpha_2 b)c^\top)$ has rank at most two. Observing that $2M_1 + M_2 = \mathrm{Diag}(2, -1, -1)$ has rank three, we deduce that condition (ii) of Theorem 6 cannot hold. We conclude that $\mathcal{T}(\mathcal{M})$ is not ROG.

Below, we construct a rank-two extreme ray of $\mathcal{T}(\mathcal{M})$.

Note that $\mathcal{N}(\mathcal{M}) = \left\{z \in \mathbb{R}^3 : z_1^2 = z_2^2 = z_3^2\right\}$ is a union of the four lines generated by $(1, 1, 1)$, $(1, 1, -1)$, $(1, -1, 1)$, and $(1, -1, -1)$. Then,

$$\mathcal{R} := \bigcup_{x,y \in \mathcal{N}(\mathcal{M})} \mathrm{span}(x, y)$$

consists of all the vectors in $\mathbb{R}^3$ with at most two different magnitudes.

As $\mathcal{R}$ is a union of finitely many planes in $\mathbb{R}^3$, there exists a vector $w \notin \mathcal{R}$. For example, we may pick $w = (0, 1, \sqrt{2})$. Dine's Theorem [36] states that as condition (i) does not hold, there exists some $u \in \mathbb{R}^3$ such that

$$\begin{pmatrix} u^\top M_1 u \\ u^\top M_2 u \end{pmatrix} = -\begin{pmatrix} w^\top M_1 w \\ w^\top M_2 w \end{pmatrix}.$$

Indeed, $u = (\sqrt{2}, 1, 0)$ is such a vector. Then, $Z := ww^\top + uu^\top$ is a rank-two matrix contained in $\mathcal{T}(\mathcal{M})$. By Corollary 5, it suffices to show that $\mathrm{range}(Z) \cap \mathcal{N}(\mathcal{M}) = \{0\}$. We will write a general element from $\mathrm{range}(Z)$ as $\left(\sqrt{2}\alpha, \alpha + \beta, \sqrt{2}\beta\right)$. Then

$$\mathrm{range}(Z) \cap \mathcal{N}(\mathcal{M}) = \left\{ \begin{pmatrix} \sqrt{2}\alpha \\ \alpha + \beta \\ \sqrt{2}\beta \end{pmatrix} : \quad 2\alpha^2 = (\alpha + \beta)^2 = 2(\beta)^2 \right\}.$$

Note that $2\alpha^2 + 2\beta^2 = 2(\alpha + \beta)^2$ implies that $\alpha\beta = 0$ and $2\alpha^2 = 2\beta^2$ implies that $|\alpha| = |\beta|$. We conclude $\mathrm{range}(Z) \cap \mathcal{N}(\mathcal{M}) = \{0\}$ and that $\mathcal{T}(\mathcal{M})$ is not ROG. ☐

3.8 Minimizing ratios of quadratic functions over ROG domains

In this section, we show how a "re-homegenization" trick can be combined with our toolset (specifically Lemma 4) to minimize the ratio of two quadratic functions over a ROG domain. Let $M_{\mathrm{obj}}, B \in \mathbb{S}^{n+1}$ and let $\mathcal{M} \subseteq \mathbb{S}^{n+1}$. We will consider the following optimization problem:

$$\inf_{\tilde{z} \in \mathbb{R}^{n+1}} \left\{ \frac{\tilde{z}^\top M_{\mathrm{obj}} \tilde{z}}{\tilde{z}^\top B \tilde{z}} : \begin{array}{l} \tilde{z}\tilde{z}^\top \in \mathcal{S}(\mathcal{M}) \\ \tilde{z}^\top B \tilde{z} > 0 \\ \tilde{z}_{n+1}^2 = 1 \end{array} \right\}. \tag{3.4}$$

Remark 24. Note that the variant of (3.4) where the constraint $\tilde{z}^\top B \tilde{z} > 0$ is replaced with $\tilde{z}^\top B \tilde{z} \neq 0$ can be decomposed as two instances of (3.4) based on the sign of $\tilde{z}^\top B \tilde{z}$ (and negating both M_{obj} and B on the portion of the domain where $\tilde{z}^\top B \tilde{z}$ is negative). ☐

We derive an SDP relaxation to (3.4) as follows:

$$\inf_{\tilde{z}\in\mathbb{R}^{n+1}} \left\{ \frac{\tilde{z}^\top M_{\mathrm{obj}}\tilde{z}}{\tilde{z}^\top B\tilde{z}} : \begin{array}{l} \tilde{z}\tilde{z}^\top \in \mathcal{S}(\mathcal{M}) \\ \tilde{z}^\top B\tilde{z} > 0 \\ \tilde{z}_{n+1}^2 = 1 \end{array} \right\} = \inf_{z\in\mathbb{R}^{n+1}} \left\{ z^\top M_{\mathrm{obj}} z : \begin{array}{l} zz^\top \in \mathcal{S}(\mathcal{M}) \\ z^\top Bz = 1 \\ z_{n+1}^2 > 0 \end{array} \right\} \tag{3.5}$$

$$\geq \inf_{z\in\mathbb{R}^{n+1}} \left\{ z^\top M_{\mathrm{obj}} z : \begin{array}{l} zz^\top \in \mathcal{S}(\mathcal{M}) \\ z^\top Bz = 1 \end{array} \right\} \tag{3.6}$$

$$\geq \inf_{Z\in\mathbb{S}^{n+1}} \left\{ \langle M_{\mathrm{obj}}, Z\rangle : \begin{array}{l} Z \in \mathcal{S}(\mathcal{M}) \\ \langle B, Z\rangle = 1 \end{array} \right\}, \tag{3.7}$$

where (3.5) follows from a simple change of variables to have the desired scaling relations, (3.6) is obtained by dropping the constraint $z_{n+1}^2 > 0$, and we dropped the rank-1 requirement on the matrix in the final relaxation step of (3.7).

Lemma 4 implies that the inequality between (3.6) and (3.7) holds with equality whenever $\mathcal{S}(\mathcal{M})$ is ROG and (3.7) is bounded below. This boundedness holds under relatively minor assumptions. Similarly, a variety of different assumptions may be used to guarantee that the inequality relation between (3.5) and (3.6) holds with equality. The following lemma demonstrates one such pair of sufficient conditions.

Lemma 12. *Let $M_{\mathrm{obj}}, B \in \mathbb{S}^{n+1}$ and $\mathcal{M} \subseteq \mathbb{S}^{n+1}$. Suppose $\mathcal{S}(\mathcal{M})$ is ROG, there exists $M^* \in \operatorname{clcone}(\mathcal{M})$ and $\lambda \in \mathbb{R}$ such that $M_{\mathrm{obj}} + M^* + \lambda B \succeq 0$, and*

$$\operatorname{cl}\left\{ z \in \mathbb{R}^{n+1} : \begin{array}{l} zz^\top \in \mathcal{S}(\mathcal{M}) \\ z_{n+1}^2 > 0 \end{array} \right\} = \left\{ z \in \mathbb{R}^{n+1} : zz^\top \in \mathcal{S}(\mathcal{M}) \right\}. \tag{3.8}$$

Then, equality holds throughout (3.5) to (3.7).

Example 15 (Regularized total least squares)**.** The total least squares problem (TLS) adapts least squares regression to the setting where both the independent and dependent variables may be corrupted by noise [46]. A variant of the TLS, known as the regularized total least squares problem (RTLS), introduces an additional regularization constraint that protects against poorly behaved solutions which arise when the data matrix has small singular values. This regularization is well studied from both theoretical and practical points of view (see [11, 106] and references therein).

By eliminating variables, the RTLS can be rewritten as minimizing the ratio of a nonnegative quadratic function and a positive quadratic function over a nonempty ellipsoid (see for example [46]). In particular, the RTLS can be written in the form of (3.4) where $M_{\mathrm{obj}}, B \in \mathbb{S}^{n+1}_+$ and $|\mathcal{M}| = 1$. It is then straightforward to verify that the assumptions of (12) are satisfied so that the RTLS admits an exact SDP relaxation in the sense of objective value exactness. □

Chapter 4

Efficient algorithms for the GTRS

This chapter is based on joint work [99] with Fatma Kılınç-Karzan.

This chapter derives an algorithm for solving the generalized trust-region subproblem (GTRS) and is based on work in [99].

The GTRS, introduced and first studied by Moré and Sorensen [73], Stern and Wolkowicz [91], seeks to minimize a general quadratic objective function subject to a general quadratic inequality constraint. Although the GTRS, as stated, is nonlinear and nonconvex, it is well-known that objective value exactness holds for its SDP relaxation under a Slater condition [40, 82]. Thus, while QCQPs are NP-hard in general, there are polynomial-time SDP-based algorithms for solving the GTRS. Nevertheless, the relatively large computational complexity of SDP-based algorithms has motivated and spurred the development of custom approaches for solving this problem. This chapter summarizes work in [99], which presents an explicit description of the convex hull of the GTRS epigraph and derives a first-order algorithm for solving the corresponding convex reformulation.

4.1 Related work

One line of work has explored the connections between the GTRS and generalized eigenvalues of the matrix pencil $A_{\text{obj}} + \gamma A_1$. Here, A_{obj} and A_1 are the quadratic forms associated to q_{obj} and q_1 respectively. Pong and Wolkowicz [84] study the optimality structure of the GTRS and propose a generalized-eigenvalue-based algorithm which exploits this structure. Adachi and Nakatsukasa [2] present an iterative algorithm for the GTRS based on the generalized eigenvalues of an auxiliary pair of matrices. Ignoring issues of exact computation, the runtime of this algorithm is $O(n^3)$ in the dense setting. Jiang and Li [57] show how to reformulate the GTRS as a convex quadratic program using generalized eigenvalues. They establish that a saddle-point-based first-order algorithm can be used to solve the reformulation within an ϵ additive error in $O(1/\epsilon)$ time.

In this line of work, it is often assumed that the generalized eigenvalues are given or can be computed exactly. A notable exception is [58], which presents an algorithm for solving the GTRS up to an ϵ additive error using only approximate eigenvalue calls. This algorithm relies on machinery developed by [49] for solving the TRS and runs in time

$$\tilde{O}\left(\frac{N}{\sqrt{\epsilon}}\log\left(\frac{n}{p}\right)\log\left(\frac{1}{\epsilon}\right)^2\right), \tag{4.1}$$

where N is the number of nonzero entries in A_{obj} and A_1, ϵ is the additive error, n is the dimension, and p is the failure probability. The algorithm proposed in [58] solves the dual SDP via a binary search scheme and makes an eigenvalue call at each step. This contrasts with the reformulation-based algorithm we will present in this chapter, which will "morally" compute only two generalized eigenvalues.[1] This parallels recent developments on the TRS where the dual SDP-based approach in [49] (which computes an eigenvalue in each iteration) was later replaced by a reformulation-based approach in [52] (which computes a single generalized eigenvalue). Following the lead of [56], we similarly perform a careful analysis of the impact of inexact (generalized) eigenvalue calls in all of our algorithms. This is an important contribution as in practice we cannot hope to compute eigenvalues exactly. For earlier work in this domain see [39, 72, 91] and references therein.

A second line of related work has studied the convex hulls of sets defined by a small number of quadratic constraints. Within this line of work, Yıldıran [110] studied the convex hull of the intersection of two *strict* quadratic inequalities (note that the resulting set is open) under the assumption that there exists $\gamma \geq 0$ such that $A_{\text{obj}} + \gamma A_1$ is positive semidefinite, and Modaresi and Vielma [71] analyze conditions under which one can safely take the closure of the sets in Yıldıran [110] and still obtain the desired closed convex hull results. In contrast to [71, 110], our analysis leverages the additional structure present in an epigraph set to give a more direct proof of the convex hull result. Furthermore, our analysis immediately suggests a rounding procedure (given a solution to the convex reformulation, we show how to find a solution to the original GTRS). This contrasts the analysis in Yıldıran [110], where such a rounding procedure is not obvious.

4.2 Preliminaries

Let $q_{\text{obj}}, q_1 : \mathbb{R}^n \to \mathbb{R}$ be quadratic functions. Write $q_{\text{obj}}(x) = x^\top A_{\text{obj}} x + 2b_{\text{obj}}^\top x + c_{\text{obj}}$ for $A_{\text{obj}} \in \mathbb{S}^n$, $b_{\text{obj}} \in \mathbb{R}^n$ and $c_{\text{obj}} \in \mathbb{R}$. Similarly define A_1, b_1, c_1. The GTRS is the following optimization problem:

$$\text{Opt} \coloneqq \inf_{x \in \mathbb{R}^n} \left\{ q_{\text{obj}}(x) : \, q_1(x) \leq 0 \right\}.$$

We will adopt the notation of Chapter 2: Denote $A[\gamma] \coloneqq A_{\text{obj}} + \gamma A_1$ and similarly define $b[\gamma]$, $c[\gamma]$. Let $q[\gamma, x] \coloneqq q_{\text{obj}}(x) + \gamma q_1(x)$. We will assume throughout this chapter that there exists a $\gamma^* \geq 0$ for which $A[\gamma^*] \succ 0$ (see [99, Theorem 2] for a weaker assumption). Further assume that q_{obj} and q_1 are both nonconvex.

Note that $\Gamma_P \coloneqq \{\gamma \in \mathbb{R}_+ : \, A[\gamma] \succeq 0\}$ is a compact interval. Let $[\gamma_-, \gamma_+]$ denote the endpoints of this interval and define $q_-(x) = q[\gamma_-, x]$. Similarly define $q_+(x)$. Then, our sufficient condition for objective value exactness (Section 2.4) immediately implies that

$$\text{Opt} = \min_{x \in \mathbb{R}^n} \max \left(q_-(x), q_+(x) \right). \tag{4.2}$$

We next give a proof sketch of objective value exactness in this special setting. The proof will show how to *round* a solution to the convex reformulation on the right hand side of (4.2) to a solution for the original nonconvex GTRS and will be useful in our later developments.

Proof of (4.2). ($\geq$) Let $x \in \mathbb{R}^n$ such that $q_1(x) \leq 0$. Then, $q_-(x) = q_{\text{obj}}(x) + \gamma_- q_1(x) \leq q_{\text{obj}}(x)$. Similarly $q_+(x) \leq q_{\text{obj}}(x)$. We deduce that $\text{Opt} \geq \inf_{x \in \mathbb{R}^n} \max(q_-(x), q_+(x))$.

($\leq$) Let $x \in \mathbb{R}^n$ and suppose $q_1(x) > 0$. Let v be a unit vector in $\ker(A[\gamma_+])$. By negating v if necessary, we may assume that $\langle b[\gamma_+], v \rangle \leq 0$. Note that $\alpha \mapsto q_-(x+\alpha v) - q_+(x+\alpha v)$ is a strongly convex quadratic function that is negative at $\alpha = 0$. Set $\hat{\alpha} > 0$ such that this expression evaluates to zero. Then, by construction, $q_1(x + \hat{\alpha} v) = 0$ and $q_{\text{obj}}(x + \hat{\alpha} v) = q_+(x + \hat{\alpha} v) \leq q_+(x)$. A similar construction can be carried out using a vector in $\ker(A[\gamma_-])$ if $q_1(x) < 0$. ■

4.3 Constructing and solving the convex reformulation

In this section, we show how to translate (the proof of) (4.2) into an algorithm for numerically solving the GTRS. In order to establish an explicit running time of an algorithm based on the above idea, we

[1] In truth, the algorithm in this chapter will simulate two generalized eigenvalue calls via a sequence of regular eigenvalue calls.

must carefully handle a number of numerical issues. In practice, we cannot expect to compute $\gamma_\pm$ exactly. Instead, we will show how to compute estimates $\tilde{\gamma}_\pm$ up to some accuracy δ, solve the corresponding convex reformulation, round the resulting point to a feasible solution in the original QCQP, and derive bounds on the errors introduced in this process. Our approach is summarized in Algorithm 1.

Throughout this section, we will make the following regularity assumptions. These assumptions are common in the literature on the GTRS; see Jiang and Li [58, Assumption 2.3] and the discussion following it.

Assumption 3. We have algorithmic access to $\hat{\gamma}, \xi, \zeta > 0$ such that $A[\hat{\gamma}] \succeq \xi I$ and $\gamma_+ \leq \zeta$. Further assume that $\xi \leq 1$, $\zeta \geq 1$, $\|A_{\text{obj}}\|, \|b_{\text{obj}}\|, |c_{\text{obj}}|, \|A_1\|, \|b_1\|, |c_1| \leq 1$. □

For convenience define $\kappa \coloneqq \zeta/\xi$ and $\delta \coloneqq \frac{\epsilon}{72\kappa^2}$. Here, δ is the accuracy to which we will compute $\tilde{\gamma}_\pm$.

Algorithm 1 ApproxConvex$(q_{\text{obj}}, q_1, \xi, \zeta, \hat{\gamma}, \epsilon, p)$

Given q_{obj}, q_1, $(\xi, \zeta, \hat{\gamma})$ satisfying Assumption 3, error parameter $0 < \epsilon \leq \kappa^2 \xi$, and failure probability $p > 0$

1. Find $\tilde{\gamma}_-$ and $\tilde{\gamma}_+$ such that

$$\tilde{\gamma}_- \in [\gamma_-, \gamma_- + \delta], \qquad \tilde{\gamma}_+ \in [\gamma_+ - \delta, \gamma_+], \qquad \lambda_{\min}(A[\tilde{\gamma}_\pm]) \leq \delta/\kappa, \tag{4.3}$$

   with failure probability of at most p. Define $\tilde{q}_-(x) \coloneqq q[\tilde{\gamma}_-, x]$ and $\tilde{q}_+(x) \coloneqq q[\tilde{\gamma}_+, x]$.
2. Define $\widetilde{\text{Opt}} \coloneqq \min_{x \in \mathbb{R}^n} \max(\tilde{q}_-(x), \tilde{q}_+(x))$ and solve $\widetilde{\text{Opt}}$ up to accuracy $\epsilon/2$.
3. Output $\tilde{\gamma}_-$, $\tilde{\gamma}_+$, and the approximate optimizer $\tilde{x}$.

The following lemma bounds $\text{Opt} - \widetilde{\text{Opt}}$. This bound follows from an estimate on the norms of the corresponding optimizers (see [99, Section 4.1]).

Lemma 13. *Suppose Assumption 3 holds, $\tilde{\gamma}_\pm$ satisfy (4.3) and $\widetilde{\text{Opt}} \coloneqq \min_{x \in \mathbb{R}^n} \max(\tilde{q}_-(x), \tilde{q}_+(x))$. Then,* $\text{Opt} - \frac{\epsilon}{2} \leq \widetilde{\text{Opt}} \leq \text{Opt}$.

[99, Algorithm 2] illustrates a simple binary search scheme for approximating $\gamma_\pm$ to the desired accuracies. In each step of this algorithm, we compute a single minimum eigenvalue using the Lanczos method. The following lemma gives a bound on the running time of [99, Algorithm 2].

Lemma 14. *Suppose Assumption 3 holds. Then, [99, Algorithm 2] outputs $\tilde{\gamma}_-$ and $\tilde{\gamma}_+$ satisfying*

$$\tilde{\gamma}_- \in [\gamma_-, \gamma_- + \delta], \qquad \tilde{\gamma}_+ \in [\gamma_+ - \delta, \gamma_+], \qquad \lambda_{\min}(A[\tilde{\gamma}_\pm]) \leq \delta/\kappa,$$

with probability $1 - p$. This algorithm runs in time

$$\tilde{O}\left(\frac{N\sqrt{\kappa\zeta}}{\sqrt{\delta}} \log\left(\frac{n}{p}\right) \log\left(\frac{\kappa}{\delta}\right)\right).$$

Remark 25. Similar algorithms for approximating $\gamma_\pm$ given $\hat{\gamma}$ have been proposed in the literature [2, 57, 72, 84]. However to our knowledge, Lemma 14 is the first to establish an explicit convergence rate; see the discussion after Remark 2.11 in [57] on this issue. □

Finally, we will solve $\widetilde{\text{Opt}}$ up to an accuracy of $\epsilon/2$ using Nesterov's accelerated gradient descent scheme for minimax problems [76]. As the final piece of this puzzle, the following lemma states that the necessary "gradient mapping" step in each iteration of Nesterov's scheme for minimax problems can be solved exactly in $O(N)$ time.

Lemma 15. *For any $y \in \mathbb{R}^n$, the quantity*

$$\arg\min_x \max\left(\tilde{q}_-(y) + \langle \nabla \tilde{q}_-(y), x - y\rangle + L\|x - y\|^2, \tilde{q}_+(y) + \langle \nabla \tilde{q}_+(y), x - y\rangle + L\|x - y\|^2\right)$$

can be computed in time $O(N)$.

Together, Lemmas 13 to 15 imply the following theorem.

Theorem 7. *Given q_{obj}, q_1 and $(\xi, \zeta, \hat{\gamma})$ satisfying Assumption 3, error parameter $0 < \epsilon \leq \kappa^2\xi$, and failure probability $p > 0$, ApproxConvex (Algorithm 1) outputs $\tilde{\gamma}_-$, $\tilde{\gamma}_+$, and $\tilde{x} \in \mathbb{R}^n$ such that*

$$\mathrm{Opt} \leq \max\left(\tilde{q}_-(\tilde{x}), \tilde{q}_+(\tilde{x})\right) \leq \widetilde{\mathrm{Opt}} + \epsilon/2 \leq \mathrm{Opt} + \epsilon$$

with probability $1 - p$. This algorithm runs in time

$$\tilde{O}\left(\frac{N\kappa^{3/2}\sqrt{\zeta}}{\sqrt{\epsilon}} \log\left(\frac{n}{p}\right) \log\left(\frac{\kappa}{\epsilon}\right)\right).$$

Finally, one may additionally bound the error introduced in rounding an approximate minimizer $\tilde{x}$ of the convex reformulation to an approximate minimizer of the original GTRS (i.e., the proof of (4.2)). This bound, which follows from estimates of the norm of $\tilde{x}$ and our bounds on $\lambda_{\min}(A[\tilde{\gamma}_\pm])$, results in the following theorem.

Theorem 8. *Given q_{obj}, q_1 and $(\xi, \zeta, \hat{\gamma})$ satisfying Assumption 3, error parameter $0 < \epsilon \leq \kappa^3\xi$, and failure probability $p > 0$, [99, Algorithm 4] outputs $\bar{x}$ such that*

$$\begin{aligned} q_{\mathrm{obj}}(\bar{x}) &\leq \mathrm{Opt} + \epsilon \\ q_1(\bar{x}) &= 0 \end{aligned}$$

with probability $1 - p$. This algorithm runs in time

$$\tilde{O}\left(\frac{N\kappa^{2}\sqrt{\zeta}}{\sqrt{\epsilon}} \log\left(\frac{n}{p}\right) \log\left(\frac{\kappa}{\epsilon}\right)\right).$$

4.4 Additional convex hull results

Again, following [101], it is possible to show that the convex hull of the GTRS epigraph is given by the projected SDP epigraph.

Theorem 9. *Suppose A_{obj} and A_1 both have at least one negative eigenvalue and there exists $\gamma \geq 0$ such that $A[\gamma] \succ 0$. Then*

$$\mathrm{conv}\left(\left\{(x,t) : \begin{array}{l} q_{\mathrm{obj}}(x) \leq t \\ q_1(x) \leq 0 \end{array}\right\}\right) = \mathrm{conv}\left(\left\{(x,t) : \begin{array}{l} q_{\mathrm{obj}}(x) \leq t \\ q_1(x) = 0 \end{array}\right\}\right) = \left\{(x,t) : \begin{array}{l} q_-(x) \leq t \\ q_+(x) \leq t \end{array}\right\}.$$

Recall, here, that $q_- = q_{\mathrm{obj}} + \gamma_- q_1$ and $q_+ = q_{\mathrm{obj}} + \gamma_+ q_1$ are convex quadratic functions.

Example 16. Define the homogeneous quadratic functions q_{obj} and q_1 by

$$A_{\mathrm{obj}} \coloneqq \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \qquad A_1 \coloneqq \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}.$$

As $\det(A_{\mathrm{obj}}) = -3$ and $\det(A_1) = -1$, the matrices A_{obj} and A_1 must both have negative eigenvalues. Furthermore,

$$A[2] = A_{\mathrm{obj}} + 2A_1 = I \succ 0.$$

Thus, the assumption of Theorem 9 is satisfied.

We now compute γ_- and γ_+. Note that as $A[\gamma]$ is a 2×2 matrix, $A(\gamma) \succeq 0$ if and only if $\mathrm{tr}(A[\gamma]) \geq 0$ and $\det(A[\gamma]) \geq 0$. Note that $\mathrm{tr}(A[\gamma]) = 2 \geq 0$ is satisfied for all γ. We compute $\det(A[\gamma]) = 1 - (2-\gamma)^2$ and deduce that $\gamma_- = 1$ and $\gamma_+ = 3$. Theorem 9 then implies

$$\mathrm{conv}\left(\left\{(x,t) \in \mathbb{R}^3 : \begin{array}{l} x_1^2 + 4x_1x_2 + x_2^2 \leq t \\ -2x_1x_2 \leq 0 \end{array}\right\}\right) = \left\{(x,t) \in \mathbb{R}^3 : \begin{array}{l} (x_1+x_2)^2 \leq t \\ (x_1-x_2)^2 \leq t \end{array}\right\}.$$

We plot the corresponding sets in Figure 4.1. □

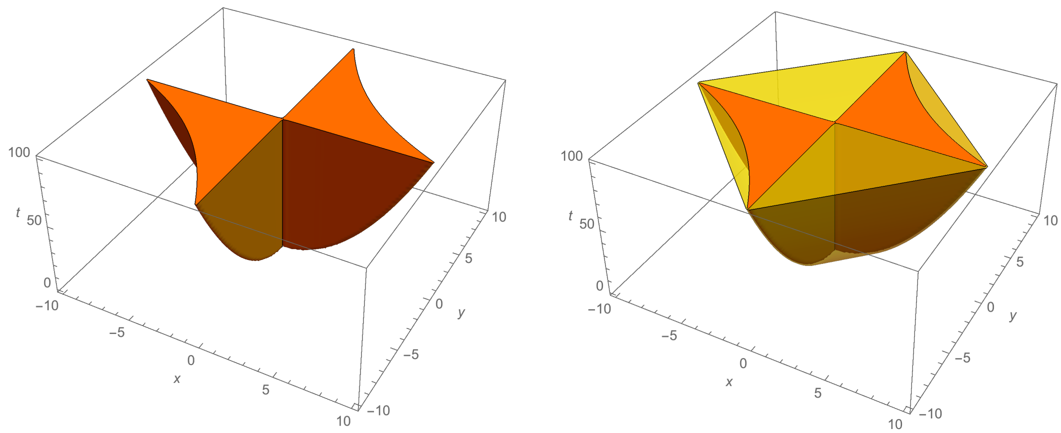

Figure 4.1: The GTRS epigraph (in orange) and its convex hull (in yellow) from Example 16

Remark 26. The assumption in Theorem 9 that there exists $\hat{\gamma} \geq 0$ such that $A[\hat{\gamma}] \succ 0$ can be relaxed to the existence of $\hat{\gamma} \geq 0$ such that $A[\hat{\gamma}] \succeq 0$ via a limiting argument. In this case, the closure of the projected SDP epigraph coincides with the *closed convex hull* of the GTRS epigraph. See [99, Theorem 2].

The assumption in Theorem 9 that A_{obj} and A_1 each have at least one negative eigenvalue can be dropped after minor modifications to some of the definitions in the current presentation. See [99, Section 2.3] for convex hull and closed convex hull descriptions of the GTRS for all possible combinations of convex/nonconvex q_{obj} and q_1. □

There have been a number of works considering interval-, equality-, or hollow-constrained variants of the TRS and GTRS [15, 59, 84, 108] (see [52, Section 3.3] and references therein for extensions of the TRS and their applications). As an immediate corollary of Theorem 9, we may also handle additional constraints which do not intersect the boundary of the feasible region. This allows us to handle multiple variants of the GTRS simultaneously.

Corollary 9. *Suppose A_{obj} and A_1 both have at least one negative eigenvalue and there exists $\gamma \geq 0$ such that $A[\gamma] \succ 0$. Suppose $\Omega \subseteq \mathbb{R}^n$ satisfies $\{x \in \mathbb{R}^n : q_1(x) = 0\} \subseteq \Omega$, then*

$$\operatorname{conv}\left(\left\{(x,t) : \begin{array}{l} q_{\text{obj}}(x) \leq t \\ q_1(x) \leq 0 \\ x \in \Omega \end{array}\right\}\right) = \left\{(x,t) : \begin{array}{l} q_-(x) \leq t \\ q_+(x) \leq t \end{array}\right\}.$$

Chapter 5

Variants of simultaneous diagonalizability of quadratic forms

This chapter is based on joint work [97] with Rujun Jiang.

A set of quadratic forms,[1] $\mathcal{A} \subseteq \mathbb{S}^n$, is simultaneously diagonalizable via congruence (SDC) if there exists a basis under which each of the quadratic forms is diagonal, i.e., if there exists an invertible $P \in \mathbb{R}^{n\times n}$ such that $P^\top AP$ is diagonal for every $A \in \mathcal{A}$. This property appears naturally when analyzing QCQPs and has important implications in this context.

The SDC property has attracted significant interest in recent years in the context of solving specific classes of QCQPs and their relaxations [56, 64, 68, 77, 101, 111, 112]. We will refer to QCQPs in which the involved quadratic forms are SDC as *diagonalizable QCQPs*. While such diagonalizable QCQPs are not easier to solve in any broad complexity-theoretic sense (indeed, binary integer programs can be cast naturally as QCQPs even only using *diagonal* quadratic constraints), they do benefit from a number of advantages over more general QCQPs: It is well known that the standard Shor semidefinite program (SDP) relaxation of a diagonalizable QCQP is equivalent to a second order cone program [101]. Consequently, the standard SDP relaxation can be computed substantially faster for diagonalizable QCQPs than for arbitrary QCQPs. This idea has been used effectively to build cheap convex relaxations within branch-and-bound frameworks [111, 112]; see also [68] for an application to portfolio optimization. Additionally, qualitative properties of the standard SDP relaxation are often easier to analyze in the context of diagonalizable QCQPs. For example, a long line of work has investigated when the SDP relaxations of certain diagonalizable QCQPs are *exact* (for various definitions of exact) and have given sufficient conditions for these properties [13, 15, 23, 52, 54–56, 66, 99]. Often, such arguments rely on conditions (such as convexity[2] or polyhedrality) of the quadratic image [83] or the set of convex Lagrange multipliers [101]. In this context, the SDC property ensures that both of these sets are polyhedral. While such conditions have been generalized beyond only diagonalizable QCQPs, these generalized sufficient conditions often become much more difficult to verify [100, 101].

This chapter presents work from [97] where we extend the reach of the SDC property by studying two new related but weaker notions of simultaneous diagonalizability. Specifically, we say that a set of quadratic forms is almost SDC (ASDC) if it is the limit of SDC sets and d-restricted SDC (d-RSDC) if it is the restriction of an SDC set in up to d-many additional dimensions. Our main contributions are a complete characterization of the ASDC pairs and the nonsingular ASDC triples. Surprisingly, we show that *every* singular pair is ASDC and that *almost every* pair is 1-RSDC.

[1] While all of our results hold with only minor modifications over both $\mathbb{R}^n$ and $\mathbb{C}^n$, we will simplify our presentation by only discussing the real setting.

[2] The convexity of the quadratic image is sometimes referred to as "hidden convexity."

5.1 Related work

Canonical forms for pairs of quadratic forms.

Weierstrass [104] and Kronecker (see [61]) proposed canonical forms for pairs of real quadratic or bilinear forms under simultaneous reductions.[3] These canonical forms were also subsequently extended to the complex case; see [62, 94–96] for historical accounts of these developments, as well as collected and simplified proofs.

While this line of work was not specifically developed to understand the SDC property, it nonetheless gives a complete characterization of the SDC property for pairs of quadratic forms. We make extensive use of these canonical forms in our work.

The SDC property for sets of three or more quadratic forms and SDC algorithms.

There has been much recent interest in understanding the SDC property for more general m-tuples of quadratic forms. In fact, the search for “sensible and ‘palpable’ conditions” for this property appeared as an open question on a short list of 14 open questions in nonlinear analysis and optimization [51].

A recent line of work has given a complete characterization of the SDC property for general m-tuples of quadratic forms: In the real symmetric setting, Jiang and Li [56] gave a complete characterization of this property under a semidefiniteness assumption. This result was then improved upon by Nguyen et al. [77] who removed the semidefiniteness assumption. Le and Nguyen [64] additionally extend these characterizations to the case of Hermitian matrices. Bustamante et al. [24] gave a complete characterization of the simultaneous diagonalizability of an m-tuple of *symmetric complex* matrices under $^\top$-congruence.[4]

We further remark that this line of work is “algorithmic” and gives numerical procedures for deciding if a given set of quadratic forms is SDC. See [64] and references therein.

The almost SDS property.

An analogous theory for the *almost* simultaneous diagonalizability of *linear operators* has been studied in the literature. In this setting, the congruence transformation is naturally replaced by a similarity transformation[5] and the SDC property is replaced by simultaneous diagonalizability *via similarity* (SDS). A widely cited theorem due to Motzkin and Taussky [74] shows that every pair of commuting linear operators, i.e., a pair of matrices in $\mathbb{C}^{n\times n}$, is almost SDS. This line of investigation was more recently picked up by O'meara and Vinsonhaler [78] who showed that triples of commuting linear operators are almost SDS under a regularity assumption on the dimensions of eigenspaces associated with the linear operators.

5.2 Preliminaries

We begin by defining formally our new variants of the SDC property.

Definition 8. A finite set $\{A_1,\dots,A_m\} \subseteq \mathbb{S}^n$ is *almost SDC* (ASDC) if for all $\epsilon > 0$, there exists $\{\tilde{A}_1,\dots,\tilde{A}_m\} \subseteq \mathbb{S}^n$ such that

- for all $i \in [m]$, the spectral norm $\|A_i - \tilde{A}_i\| \le \epsilon$, and
- $\{\tilde{A}_1,\dots,\tilde{A}_m\}$ is SDC. □

Definition 9. A finite set $\{A_1,\dots,A_m\} \subseteq \mathbb{S}^n$ is *d-restricted SDC* (d-RSDC) if there exists $\{\tilde{A}_1,\dots,\tilde{A}_m\} \subseteq \mathbb{S}^{n+d}$ such that

- for all $i \in [m]$, A_i is the top-left $n \times n$ block of $\tilde{A}_i$, and

[3] Reduction refers to one of similarity, congruence, equivalence, or strict equivalence.

[4] We emphasize that Bustamante et al. [24] consider complex symmetric matrices and adopt $^\top$-congruence as their notion of congruence.

[5] Recall that two matrices $A, B \in \mathbb{C}^{n\times n}$ are similar if there exists an invertible $P \in \mathbb{C}^{n\times n}$ such that $A = P^{-1}BP$.

- $\left\{\tilde{A}_1, \dots, \tilde{A}_m\right\}$ is SDC. □

It is not difficult to show that the ASDC and d-RSDC properties are actually properties of subspaces so that we may replace any given set $\{A_1, \dots, A_m\}$ with a different set generating the same subspace when analyzing these properties.

Definition 10. A set $\mathcal{A} \subseteq \mathbb{S}^n$ is *nonsingular* if $\operatorname{span}(\mathcal{A})$ contains a nonsingular matrix. Else, it is *singular*. □

Before moving on, we give examples how these two properties may be used to diagonalize QCQPs.

Example 17. Consider a QCQP over $\mathbb{R}^n$ of the form

$$\mathrm{Opt} \coloneqq \inf_{x\in\mathbb{R}^n} \left\{ q_{\mathrm{obj}}(x) : \begin{array}{l} q_i(x) \leq 0,\ \forall i \in [m] \\ x \in \mathcal{L} \end{array} \right\},$$

where $q_{\mathrm{obj}}, q_1, \dots, q_m : \mathbb{R}^n \to \mathbb{R}$ are quadratic functions and $\mathcal{L} \subseteq \mathbb{R}^n$ is a polytope. Define $q_{\mathrm{obj}}(x) = x^\top A_{\mathrm{obj}} x + 2b_{\mathrm{obj}}^\top x + c_{\mathrm{obj}}$ for $A_{\mathrm{obj}} \in \mathbb{S}^n$, $b_{\mathrm{obj}} \in \mathbb{R}^n$, $c_{\mathrm{obj}} \in \mathbb{R}$. Similarly define $A_i \in \mathbb{S}^n$, $b_i \in \mathbb{R}^n$, and $c_i \in \mathbb{R}$. Below, we will suppose that $\mathcal{A} \coloneqq \{A_{\mathrm{obj}}, A_1, \dots, A_m\}$ is either ASDC or d-RSDC and show how to diagonalize the above QCQP.

First, suppose $\{A_{\mathrm{obj}}, A_1, \dots, A_m\}$ is ASDC and let $\left\{\tilde{A}_{\mathrm{obj}}, \tilde{A}_1, \dots, \tilde{A}_m\right\}$ denote the set furnished by the ASDC property for a given $\epsilon > 0$. Let $\tilde{q}_{\mathrm{obj}}(x) = x^\top \tilde{A}_{\mathrm{obj}} x + 2b_{\mathrm{obj}}^\top x + c_{\mathrm{obj}}$ and similarly define $\tilde{q}_1, \dots, \tilde{q}_m$. Let R denote a bound on the radius of $\mathcal{L}$, i.e., $\mathcal{L} \subseteq B(0, R)$, and set $\delta = \epsilon R^2$. Then, define

$$\begin{aligned} \widetilde{\mathrm{Opt}} &\coloneqq \inf_{x\in\mathbb{R}^n} \left\{ \tilde{q}_{\mathrm{obj}}(x) : \begin{array}{l} \tilde{q}_i(x) \leq 0,\ \forall i \in [m] \\ x \in \mathcal{L} \end{array} \right\}, \text{ and} \\ \mathrm{Opt}_\pm &\coloneqq \inf_{x\in\mathbb{R}^n} \left\{ q_{\mathrm{obj}}(x) \pm \delta : \begin{array}{l} q_i(x) \pm \delta \leq 0,\ \forall i \in [m] \\ x \in \mathcal{L} \end{array} \right\}. \end{aligned}$$

Then, $\mathrm{Opt}_+ \geq \widetilde{\mathrm{Opt}} \geq \mathrm{Opt}_-$.

Note that $\widetilde{\mathrm{Opt}}$ is a diagonalizable QCQP. In other words, if we are willing to lose arbitrarily small additive errors in both the objective function and the constraints of a QCQP with ASDC quadratic forms, then we may approximate the QCQP by a diagonalizable QCQP.

Next, suppose $\{A_1, \dots, A_m\}$ is d-RSDC and let $\tilde{A}_1, \dots, \tilde{A}_m \in \mathbb{S}^{n+d}$ denote the furnished matrices. Define the quadratic function $\tilde{q}_{\mathrm{obj}} : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ by

$$\tilde{q}_{\mathrm{obj}}(x, y) \coloneqq \begin{pmatrix} x \\ y \end{pmatrix}^\top \tilde{A}_{\mathrm{obj}} \begin{pmatrix} x \\ y \end{pmatrix} + 2b_{\mathrm{obj}}^\top x + c_{\mathrm{obj}}$$

and similarly $\tilde{q}_1, \dots, \tilde{q}_m$. Then,

$$\mathrm{Opt} = \inf_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^d} \left\{ \tilde{q}_{\mathrm{obj}}(x, y) : \begin{array}{l} \tilde{q}_i(x, y) \leq 0,\ \forall i \in [m] \\ x \in \mathcal{L},\ y = 0 \end{array} \right\}.$$

In particular, the augmented QCQP (which by construction is diagonalizable) is an exact reformulation of the original QCQP. □

Characterization of SDC

A number of necessary and/or sufficient conditions for the SDC property have been given in the literature [24, 53, 62]. For our purposes, we will need the following two results.

Proposition 3. *Let $\mathcal{A} \subseteq \mathbb{S}^n$ and suppose $S \in \operatorname{span}(\mathcal{A})$ is nonsingular. Then, $\mathcal{A}$ is SDC if and only if $S^{-1}\mathcal{A}$ is a commuting set of diagonalizable matrices with real spectra.*

Proposition 4. *Let $\mathcal{A} \subseteq \mathbb{S}^n$ and suppose $S \in \operatorname{span}(\mathcal{A})$ is a max-rank element of $\operatorname{span}(\mathcal{A})$. Then, $\mathcal{A}$ is SDC if and only if $\operatorname{range}(A) \subseteq \operatorname{range}(S)$ for every $A \in \mathcal{A}$ and $\left\{A|_{\operatorname{range}(S)} : A \in \mathcal{A}\right\}$ is SDC.*

5.3 Variants of SDC for pairs of symmetric matrices

In this section, we will present a complete characterization of the ASDC property for pairs of symmetric matrices. As a byproduct, we will also see a sufficient condition for the 1-RSDC property for pairs of symmetric matrices.

We will switch the notation above and label our matrices $\mathcal{A} = \{A, B\}$. The analysis in [97] proceeds in two cases: when $\{A, B\}$ is nonsingular and singular respectively. In this section, we will follow this breakdown but will omit full proofs in favor of representative examples.

The examples we will present are in fact representative in a very rigorous sense. Specifically, the examples we present will treat the different types of blocks that appear in the canonical form for pairs of symmetric matrices (see [62, 95] and references therein).

The nonsingular case

The main result of this section is that if A is invertible, then $\{A, B\}$ is ASDC if and only if $A^{-1}B$ has real spectrum. We begin by examining two examples that are representative of the situation when A is invertible.

Example 18. Let

$$A = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Noting that $A^{-1}B$ is not diagonalizable, we conclude via Proposition 3 that $\{A, B\}$ is not SDC. On the other hand, let $\epsilon > 0$ and define

$$\tilde{B} = \begin{pmatrix} \epsilon & 1 \\ 1 & 1 \end{pmatrix}.$$

Now, $A^{-1}\tilde{B}$ has spectrum $1 \pm \sqrt{\epsilon}$, whence by Proposition 3 $\{A, \tilde{B}\}$ is SDC. □

Example 19. Let

$$A = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}.$$

Noting that $A^{-1}B$ has non-real eigenvalues, we conclude via Proposition 3 (and the fact that eigenvalues vary continuously) that $\{A, B\}$ is not ASDC. □

While the set of diagonalizable matrices is dense in $\mathbb{R}^{n\times n}$, it is not immediately clear that the pairs $(A, B) \in \mathbb{S}^n \times \mathbb{S}^n$ such that $A^{-1}B$ exists and is diagonalizable is dense in $\mathbb{S}^n \times \mathbb{S}^n$. The following lemma [97, Lemma 3] generalizes the above examples and shows that this is indeed the case. Specifically, it is possible to perturb B so as to "fan out" all of the nonsimple eigenvalues of $A^{-1}B$ along the real direction.

Lemma 16. *Let $\{A, B\} \subseteq \mathbb{S}^n$ and suppose A is invertible. For all $\epsilon > 0$, there exists $\tilde{B}$ such that*

- $\left\|B - \tilde{B}\right\| \le \epsilon$,
- *$A^{-1}\tilde{B}$ has a simple spectrum (whence in particular, $A^{-1}\tilde{B}$ is diagonalizable), and*
- *$A^{-1}\tilde{B}$ and $A^{-1}B$ have the same number of real eigenvalues counted with multiplicity.*

The following theorem then follows as a corollary to Lemma 16 and Proposition 3.

Theorem 10. *Let $A, B \in \mathbb{S}^n$ and suppose A is invertible. Then, $\{A, B\}$ is ASDC if and only if $A^{-1}B$ has a real spectrum.*

The singular case

The main result of this section is that *every* singular pair of symmetric matrices is ASDC. We begin with an example.

Example 20. Consider the following singular pair of symmetric matrices

$$A = \left(\begin{array}{cc|cc|c} & 1 & & & \\ 1 & & & & \\ \hline & & & 1 & \\ & & 1 & & \\ \hline & & & & 0 \end{array}\right), \qquad B = \left(\begin{array}{cc|cc|c} 1 & & & & \\ & -1 & & & \\ \hline & & 2 & & \\ & & & -2 & \\ \hline & & & & 0 \end{array}\right). \tag{5.1}$$

We will show that $\{A, B\}$ is ASDC. Towards this end, consider the following perturbation of $\{A, B\}$

$$\tilde{A} = \left(\begin{array}{cc|cc|c} & 1 & & & \\ 1 & & & & \\ \hline & & & 1 & \\ & & 1 & & \\ \hline & & & & \epsilon \end{array}\right), \qquad \tilde{B} = \left(\begin{array}{cc|cc|c} 1 & & & & \sqrt{\epsilon}u_1 \\ & -1 & & & \sqrt{\epsilon}v_1 \\ \hline & & 2 & & \sqrt{\epsilon}u_2 \\ & & & -2 & \sqrt{\epsilon}v_2 \\ \hline \sqrt{\epsilon}u_1 & \sqrt{\epsilon}v_1 & \sqrt{\epsilon}u_2 & \sqrt{\epsilon}v_2 & \epsilon z \end{array}\right) \tag{5.2}$$

where $\epsilon > 0$, $u_1, u_2, v_1, v_2, z \in \mathbb{R}$. Our goal is to set these quantities so that

$$\tilde{A}^{-1}\tilde{B} = \operatorname{Diag}(1, \ldots, 1, \sqrt{\epsilon})^{-1} \left(\begin{array}{cc|cc|c} & -1 & & & v_1 \\ 1 & & & & u_1 \\ \hline & & & -2 & v_2 \\ & & 2 & & u_2 \\ \hline u_1 & v_1 & u_2 & v_2 & z \end{array}\right) \operatorname{Diag}(1, \ldots, 1, \sqrt{\epsilon})$$

has a simple real spectrum. The above expression shows that whether $\tilde{A}^{-1}\tilde{B}$ has a simple real spectrum is invariant under changing $\epsilon > 0$ so that we may fix $\epsilon = 1$ in the following. To recap, our goal is now to set $u_1, u_2, v_1, v_2, z \in \mathbb{R}$ such that the polynomial (where terms have been collected suggestively)

$$\begin{aligned} \lambda \mapsto &- \lambda(\lambda^2+1)(\lambda^2+4) \\ &+ z\left[(\lambda^2+1)(\lambda^2+4)\right] + (v_1^2 - u_1^2)\left[\lambda^2+4\right] + (2u_1v_1)\left[\lambda(\lambda^2+4)\right] \\ &+ (2(v_2^2-u_2^2))\left[(\lambda^2+1)\right] + (2u_2v_2)\left[\lambda(\lambda^2+1)\right] \end{aligned}$$

evaluates to zero at five distinct real values. It is evident that this is possible (in fact with *any* five distinct real values) upon observing that (1) the five square-bracketed univariate polynomials in the second and third lines form a basis of the degree-4 polynomials in λ, and (2) the coefficient map $(z, u_1, u_2, v_1, v_2) \mapsto (z, v_1^2 - u_1^2, 2u_1v_1, 2(v_2^2 - u_2^2), 2u_2v_2)$ is surjective. We conclude that $\{A, B\}$ is ASDC.

For this particular example, the choice $(z, u_1, u_2, v_1, v_2) = (0, \sqrt{5/3}, -\sqrt{5/3}, 2\sqrt{5/3}, 2\sqrt{5/3})$ and $\epsilon > 0$ results in a matrix $\tilde{A}^{-1}\tilde{B}$ with spectrum $\{-2, -1, 0, 1, 2\}$. $\square$

Similar ideas to those in the above example, coupled with Lemma 16, can be used to prove the following theorems.

Theorem 11. *Let $\{A, B\} \subseteq \mathbb{S}^n$. If $\{A, B\}$ is singular, then it is ASDC.*

Theorem 12. *Let $A, B \in \mathbb{S}^n$. If A is invertible and $A^{-1}B$ has a simple spectrum, then $\{A, B\}$ is 1-RSDC. In particular, the 1-RSDC pairs are dense in $\mathbb{S}^n \times \mathbb{S}^n$.*

In fact, the proofs of both theorems are *constructive* and give an explicit procedure for constructing the corresponding SDC sets. In both cases, the SDC sets can be computed using a single eigendecomposition and linear solve.

Remark 27. In view of Example 17, Theorem 12 implies that we may diagonalize the problem of minimizing a quadratic objective function subject to a single quadratic constraint and a polytope constraint if we allow ourselves arbitrarily small errors and just one additional variable. Furthermore, if the quadratic objective function and quadratic constraint satisfy the generic property that A_{obj} is invertible and $A_{\text{obj}}^{-1}A_1$ has a simple spectrum, then we may diagonalize this QCQP with *no errors* and just one additional variable. Preliminary numerical experiments comparing the running times for solving a QCQP in this form directly versus in its lifted diagonal reformulation can be found in [97]. $\square$

5.4 Additional results

This section briefly discusses additional results in the direction of understanding the ASDC and d-RSDC properties.

The following theorem gives a characterization of the nonsingular ASDC triples.

Theorem 13. *Let $\{A, B, C\} \subseteq \mathbb{S}^n$ and suppose A is invertible. Then, $\{A, B, C\}$ is ASDC if and only if $\{A^{-1}B, A^{-1}C\}$ are a pair of commuting matrices with real spectra.*

The proof of this theorem extends an inductive argument due to Motzkin and Taussky [74] and shows that we may repeatedly perturb either $A^{-1}B$ or $A^{-1}C$ to increase the number of simple eigenvalues. In contrast to the original argument in [74], which establishes that any commuting pair $\{S, T\} \subseteq \mathbb{C}^{n\times n}$ is almost simultaneously diagonalizable *via similarity* (and thus only needs to inductively maintain commutativity of S and T), for our proof we will further need to maintain that A, B, C are symmetric matrices and that $A^{-1}B$ and $A^{-1}C$ have real spectra.

The following theorem describes an obstruction to further generalizations of Theorem 11. Specifically, in contrast to Theorem 11, which showed that singularity implies ASDC for pairs of symmetric matrices, the following theorem shows that even *triples* of symmetric matrices with "large amounts" of singularity can fail to be ASDC.

Theorem 14. *Let $\{A = I_n, B, C\} \subseteq \mathbb{S}^n$ and let $d < \operatorname{rank}([B, C])/2$. Here $[B, C] = BC - CB$ is the commutator of B and C. Then $A^{-1}B = B$ and $A^{-1}C = C$ have real spectra, but*

- *$\{A \oplus 0_d, B \oplus 0_d, C \oplus 0_d\}$ is not ASDC, and*
- *$\{A, B, C\}$ is not d-RSDC.*

Finally, one may conjecture that a nonsingular set of matrices, say $\mathcal{A}$ containing an invertible A_1, is ASDC if and only if $A_1^{-1}\mathcal{A}$ is a set of commuting matrices with real spectra. The following theorem adapts a technique introduced by O'meara and Vinsonhaler [78] for studying the almost simultaneously diagonalizable via similarity property of subsets of $\mathbb{C}^{n\times n}$ and shows that this statement is also false.

Theorem 15. *There exists a set $\mathcal{A} = \{A_1, \ldots, A_7\} \subseteq \mathbb{S}^6$ such that A_1 is invertible, $A_1^{-1}\mathcal{A}$ is a set of commuting matrices with real spectra, and $\mathcal{A}$ is not ASDC.*

Chapter 6

Future directions

This section discusses three avenues of current work and future directions related to this thesis.

The first project that we discuss below is related to Question 1 and explores applications of the framework introduced in [98, 100, 101] in the context of random QCQPs. The remaining two projects that we discuss below are related to Question 2 and are outgrowths of the completed work on the generalized trust-region subproblem [99] and sufficient conditions for exactness [101]. Specifically, these projects seek to understand how a strong form of exactness in the SDP relaxation of a QCQP may translate to an *implicit regularity* condition and improved algorithms for SDPs where this strong form of exactness holds.

6.1 Exactness in random QCQPs

This section is based on ongoing work with Fatma Kılınç-Karzan.

A number of recent exciting results on phase retrieval [25] and clustering [1, 70, 86] have shown that under various assumptions on the data (or on the parameters in a random data model), the QCQP formulation of the corresponding problem has a tight SDP relaxation. Motivated by these applications, we hope to build on work presented in Chapter 2 to analyze multiple classes random QCQPs and notions of exactness using the tools developed in [98, 100, 101]. Below, we will describe one such partial result in this direction that says that for any fixed number of constraints and $\epsilon > 0$, with probability $1 - o(1)$ as $n \to \infty$, the SDP relaxation will be "almost" exact in a rigorous sense.

Related work

Random QCQPs have served as a useful testing ground for various sufficient conditions for exactness in this area. For example, Burer and Ye [23] show that SDP exactness holds with probability $1 - o(1)$ as $n \to \infty$ for random *diagonal* QCQPs satisfying a certain geometric condition. Locatelli [67] derives a similar result for QCQPs with one quadratic constraint and one linear constraint without the explicit geometric condition present in [23]. In this direction, we hope to derive similar results for the setting where the number of constraints is arbitrary but fixed and all of the quadratic forms are defined by the (normalized) Gaussian orthogonal ensemble (see Preliminaries).

Exactness has also been observed for a number of random QCQPs arising in statistical and learning settings. For example Candès et al. [25] show that the QCQP formulation of the phase retrieval problem has an exact SDP relaxation given enough measurements. A similar story holds for various clustering problems (e.g., [1, 70, 86]) where exactness can be shown to hold with high probability as long as certain measures of *signal strength* are large enough.

Preliminaries

We briefly review and extend some of the notation and results in Chapter 2 that will be useful in this discussion. We will consider a random QCQP with random equality constraints and a unit ball constraint of the form

$$\inf_{x\in\mathbb{R}^n}\left\{q_{\mathrm{obj}}(x):\ \begin{array}{l} q_i(x)=0,\ \forall i\in[m] \\ \|x\|^2\leq 1\end{array}\right\}. \tag{6.1}$$

Here, for all $i\in[m]$, the quadratic function $q_i(x)=x^\top A_i x+2b_i^\top x+c_i$ for some $A_i\in\mathbb{S}^n$, $b_i\in\mathbb{R}^n$, and $c_i\in\mathbb{R}$. Similarly define A_{obj}, b_{obj}, c_{obj}. Let $q:\mathbb{R}^n\to\mathbb{R}^m$ such that $q(x)_i=q_i(x)$. For convenience define $q_{m+1}(x)=\|x\|^2-1$. Define $A(\eta)\coloneqq\sum_{i=1}^m\eta_i A_i$ and similarly define $b(\eta)$, $c(\eta)$.

We will define the cone of convex Lagrange multipliers

$$\Gamma\coloneqq\{(\gamma_{\mathrm{obj}},\eta,\nu)\in\mathbb{R}_+\times\mathbb{R}^m\times\mathbb{R}_+:\ \gamma_{\mathrm{obj}}A_{\mathrm{obj}}+A(\eta)+\nu I\succeq 0\}$$

and its slice at $\gamma_{\mathrm{obj}}=1$,

$$\Gamma_P\coloneqq\{(\eta,\nu)\in\mathbb{R}^m\times\mathbb{R}_+:\ (1,\eta,\nu)\in\Gamma\}.$$

One may extend the sufficient condition of Theorem 1 to the setting where Γ is arbitrary whenever the SDP relaxation is strictly feasible.

Theorem 16. *Suppose there exists $(x,X)\in\mathbb{R}^n\times\mathbb{S}^n$ such that*

$$\begin{cases}\langle A_i,X\rangle+2\langle b_i,x\rangle+c_i=0,\ \forall i\in[m],\\ \operatorname{tr}(X)<1,\ \textit{and}\\ X\succ xx^\top.\end{cases}$$

Furthermore suppose for all $(\eta,\nu)\in\Gamma_P$ and $v\in\mathbf{S}^{n-1}$, that

$$v\in\ker\left(A_{\mathrm{obj}}+A(\eta)+\nu I\right)\implies\langle b_{\mathrm{obj}}+b(\eta),v\rangle\neq 0.$$

Then, the SDP relaxation of (6.1) has a rank-one optimal solution.

Lemma 17. *Suppose the SDP relaxation of (6.1) is feasible. Then, its value is given by*

$$\min_{x\in\mathbb{R}^n}\left(q_{\mathrm{obj}}(x)+\sup_{(\eta,\nu)\in\Gamma_P}\left(\langle\eta,q(x)\rangle+\nu q_{m+1}(x)\right)\right). \tag{6.2}$$

Below, we will explore the relationship between (6.1) and (6.2) for a fixed m as n tends to infinity. We will assume $A_{\mathrm{obj}},A_1,\dots,A_m$ are i.i.d. samples from the normalized Gaussian orthogonal ensemble (see paragraph below), but will let $b_{\mathrm{obj}},b_1,\dots,b_m\in\mathbb{R}^n$ and $c_{\mathrm{obj}},c_1,\dots,c_m\in\mathbb{R}^n$ be arbitrary. We will treat $m\in\mathbb{N}$ and $\epsilon>0$ as constants and abbreviate "with probability $1-o(1)$ as $n\to\infty$" as asymptotically almost surely (*a.a.s.*).

The normalized GOE. We will assume that $A_{\mathrm{obj}},A_1,\dots,A_m$ are drawn from the normalized Gaussian Orthogonal Ensemble.

Definition 11. Let $A\in\mathbb{S}^n$ be a random matrix where: each diagonal entry $A_{i,i}$ is i.i.d. $N(0,1/2n)$; each superdiagonal entry $A_{i,j}$ is i.i.d. $N(0,1/4n)$; and each subdiagonal entry $A_{i,j}$ is defined by symmetry. We will refer to the distribution on A as the normalized Gaussian Orthogonal Ensemble (NGOE). Let

$$A\sim\mathrm{NGOE}(n)$$

denote the fact that A is drawn according to this distribution. □

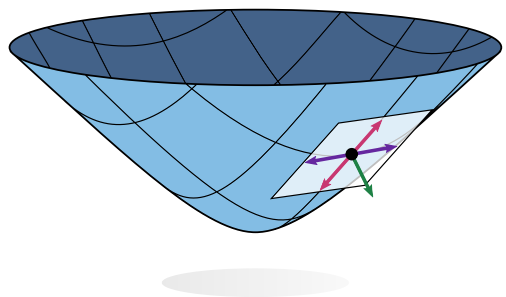

Figure 6.1: Proof sketch: we will begin by showing that in our random model, Γ will converge to the hyperboloid (the blue surface) (Lemma 18). Next, we will show that there exists a point $(\bar{\eta}, \bar{\nu})$ on the surface of the hyperboloid (the black indicated dot) almost maximizing the linear form defined by $(q(\bar{x}), q_{m+1}(\bar{x})) = (\bar{s}, \bar{t})$. In particular, the vector $(\bar{s}, \bar{t})$ will point mostly in the normal direction (indicated in green) to the hyperboloid at $(\bar{\eta}, \bar{\nu})$. We will then pick an appropriate modification $x^* = \bar{x} + \alpha v$ such that: the value of the linear map evaluated at $(\bar{\eta}, \bar{\nu})$ remains mostly unchanged, $(q(x^*), q_{m+1}(x^*))$ is still almost zero in the direction tangent to the cone (indicated in pink), the component of $(q(x^*), q_{m+1}(x^*))$ in the normal direction (indicated in green) is reduced to almost zero, and all components in the remaining directions (indicated in purple) are still almost zero. Then as the tangent, normal, and remaining directions form a basis of $\mathbb{R}^{m+1}$, we will deduce that $q_i(x^*) \approx 0$ for all $i = 1, \ldots, m+1$. We will then conclude that $q_{\text{obj}}(x^*)$ is approximately equal to its value at $(\bar{\eta}, \bar{\nu})$ (which is in turn approximately the SDP optimum).

Remark 28. A different procedure for generating the same distribution is: sample $M \in \mathbb{R}^{n \times n}$ with every entry i.i.d. $N(0, 1/2n)$, then return $A = (M + M^\top)/2$. □

Our choice of normalization here ensures that $\|A\|_2 \in [1 \pm \epsilon]$ *a.a.s.*.

The GOE is a very well-understood distribution. We will only need the following basic facts.

Fact 1. *Let* $A_1, \ldots, A_k \overset{i.i.d.}{\sim} \mathrm{NGOE}(n)$. *Let* $U \in \mathbb{R}^{k \times k}$ *be an orthogonal matrix and define*

$$\tilde{A}_i = \sum_{j=1}^{k} U_{i,j} A_j.$$

Then, $\tilde{A}_1, \ldots, \tilde{A}_k \overset{i.i.d.}{\sim} \mathrm{NGOE}(n)$.

Fact 2. *Let* $A \sim \mathrm{NGOE}(n)$. *Let* $U \in \mathbb{R}^{n \times n}$ *be an orthogonal matrix, then* $U^\top A U \sim \mathrm{NGOE}(n)$.

Fact 3. *Fix* $\epsilon > 0$. *Let* $A \sim \mathrm{NGOE}(n)$ *and let* $\lambda_1 \leq \cdots \leq \lambda_n$ *denote the spectrum of* A. *Then*

$$|\lambda_1|, |\lambda_n| \leq 1 + \epsilon \quad a.a.s.$$

and

$$\frac{|\{i \in [n] : \lambda_i \in [-1 \pm \epsilon]\}|}{n} \geq C_\epsilon \quad a.a.s.,$$

where C_ϵ *is a positive constant depending only on* ϵ.

Partial results

Below we present a partial result which states the SDP relaxation is "almost" exact for any m as $n \to \infty$. Our proof strategy will be to show that for large enough n, the set Γ_P will converge to a set with a simple analytical description. Next, using the partial Lagrangian relaxation, we will take an optimal x to the SDP relaxation and recover an almost-feasible x to the original QCQP. See Figure 6.1 for a proof sketch.

Fix $m \in \mathbb{N}$ and $\epsilon \in (0, 1/2]$. The following lemma can be thought of as saying that Γ_P is approximately the hyperboloid in $\mathbb{R}^{m+1}$.

Lemma 18. *The following holds a.a.s.:*

$$\Gamma_P \supseteq H_\epsilon \coloneqq \left\{(\eta, \nu) \in \mathbb{R}^{m+1} : \nu(1-\epsilon^2) \geq \sqrt{1 + \|\eta\|_2^2}\right\}.$$

Construct a finite collection of ordered orthonormal bases of $\mathbb{R}^{1+m}$, denoted $\mathcal{N}$, as follows:

- Let $\{(\sigma, \eta)\} \subseteq \mathbf{S}^m$ denote a finite ℓ_2-norm ϵ^2-net. Assume that $\eta \neq 0$ and $\sigma \neq 0$ for all (σ, η) in this net.
- Extend each (σ, η) in the ϵ^2-net to an orthonormal basis: Set $(\sigma^{(1)}, \eta^{(1)}) = (\sigma, \eta)$. Next, set

$$(\sigma^{(2)}, \eta^{(2)}) = \left(\left\|\eta^{(1)}\right\| \frac{\sigma^{(1)}}{\left|\sigma^{(1)}\right|}, -\left|\sigma^{(1)}\right| \frac{\eta^{(1)}}{\left\|\eta^{(1)}\right\|}\right)$$

  then extend the resulting set to an orthonormal basis $\mathcal{B}$ for $\mathbb{R}^{m+1}$.

The main property that we will need from $\mathcal{N}$ is that $\left\{(\sigma^{(1)}, \eta^{(1)}) : \mathcal{B} \in \mathcal{N}\right\}$ forms an ℓ_2-norm ϵ^2-net on $\mathbf{S}^{m-1}$. The second property—that we can parameterize span $\left\{(\sigma^{(1)}, \eta^{(1)}), (1, 0)\right\}$ using vectors from $\mathcal{B}$—is only for notational convenience to be used later. Note also that $\sigma^{(i)} = 0$ and $\left\|\eta^{(i)}\right\| = 1$ for all $i \in [3, m+1]$.

The next lemma follows from Facts 1 to 3 and states that there are large subspaces corresponding to small eigenvalues of the aggregated quadratic forms. This should be compared to Theorem 1 and Example 7 where it was assumed that there were large subspaces in the shared kernel of the aggregated quadratic forms.

Lemma 19. *The following holds a.a.s.: For every $\mathcal{B} \in \mathcal{N}$, there exists an $(m+3)$-dimensional subspace $V \subseteq \mathbb{R}^n$ such that for all unit $v \in V$,*

$$v^\top \left(\sigma^{(1)} A_{\text{obj}} + A(\eta^{(1)}) v\right) \in [-1 \pm \epsilon^2], \qquad v^\top \left(\sigma^{(i)} A_{\text{obj}} + A(\eta^{(i)})\right) v \in [\pm \epsilon^2],\ \forall i \in [2, m+1].$$

The following theorem uses Lemmas 18 and 19 to perform an inexact version of the rounding procedure proposed in Chapter 2.

Theorem 17. *Fix $\epsilon \in (0, 1/2]$. The following holds asymptotically almost surely: Let $A_{\text{obj}}, A_1, \dots, A_m \sim \mathrm{NGOE}(n)$. Then for all $b_{\text{obj}}, b_1, \dots, b_m \in \mathbb{R}^n$ and $c_{\text{obj}}, c_1, \dots, c_m \in \mathbb{R}$, we have*

$$\mathrm{Opt}_{(6.1)} \geq \mathrm{Opt}_{(6.2)} \geq \inf_{x \in \mathbb{R}^n} \left\{ q_{\text{obj}}(x) : \begin{array}{l} q_i(x) \in [\pm\epsilon],\ \forall i \in [m] \\ \|x\|^2 \leq 1 \end{array} \right\} - \epsilon.$$

In words, for any fixed $m \in \mathbb{N}$ and $\epsilon > 0$, as $n \to \infty$, the optimal value of the SDP relaxation (6.2) will be sandwiched between the QCQP optimal (6.1) and its ϵ-perturbed variant with probability $1 - o(1)$.

Open questions

In this direction, there are two specific questions we would like to investigate further. First, it seems plausible that exactness could hold in a similar random model (without adversarial b_is and c_is) *without* the need of an $\epsilon > 0$ perturbation. Exactness guarantees in this setting would extend work in [23]. Second, we would like to understand whether exactness results in statistical settings (e.g., [1, 25, 70]) may be recovered using our framework.

6.2 Implicit regularity in the GTRS

This section is based on ongoing work with Fatma Kılınç-Karzan and Yunlei Lu.

The algorithm proposed in [99] (and discussed in Chapter 4) for solving the GTRS up to an accuracy of ϵ runs in time

$$\tilde{O}\left(\frac{N}{\sqrt{\epsilon}} \log\left(\frac{n}{p}\right) \log\left(\frac{1}{\epsilon}\right)\right).$$

Recall here that N was the number of nonzero entries in A_{obj} and A_1, n was the dimension of the problem, and p was the failure probability. This dependence on ϵ may be compared to the oracle complexity lower bound for minimizing smooth convex functions using a first-order method. Alternatively, one may compare this running time with the running time of the Lanczos method for the closely related minimum eigenvalue problem,

$$O\left(\frac{N}{\sqrt{\epsilon}}\log\left(\frac{n}{p}\right)\right).$$

Surprisingly, we have reason to believe that it is possible to design algorithms for the GTRS whose dependence on ϵ scales much more efficiently: as $\approx \min\left(\frac{1}{\sqrt{\mu^*}}\log\left(\frac{1}{\epsilon}\right), \frac{1}{\sqrt{\epsilon}}\right)$. Here, μ^* is the *implicit regularity* of the given instance (see Preliminaries) and is often positive—for example, this value is positive with probability one under many of the natural random models for the GTRS. From this point of view, the minimum eigenvalue problem, for which $\mu^* = 0$, actually belongs to the *hardest* problems within the GTRS problem class.

Related work

By now, a number of interesting iterative algorithms for the GTRS have been proposed [2, 57, 58, 99]. The analyses of these algorithms fail to exploit implicit regularization and hence have guarantees of the form $\approx \frac{1}{\sqrt{\epsilon}}$. In this direction, implicit regularity has previously been observed and investigated for the trust region subproblem as well as for the cubic regularized Newton method [26]. In this setting, Carmon and Duchi [26] shows that both the TRS and cubic regularized quadratic minimization problems can be solved in time $\approx \min\left(\frac{1}{\sqrt{\mu^*}}\log\left(\frac{1}{\epsilon}\right), \frac{1}{\sqrt{\epsilon}}\right)$. In contrast to Carmon and Duchi [26] who achieve these running times based on an analysis of the Krylov subspaces, we present an explanation of this improved running time in the setting of the GTRS using "low-resolution" minimum eigenvalue computations.

Preliminaries

Recall the GTRS setup from Chapter 4: Suppose that $q_{\text{obj}}, q_1 : \mathbb{R}^n \to \mathbb{R}$ are nonconvex quadratic functions and suppose that there exists $\hat{\gamma}$ such that $A_{\text{obj}} + \hat{\gamma}A_1 \succ 0$. Here, A_{obj} and A_1 are the quadratic forms in q_{obj} and q_1 respectively. We will additionally make the regularity assumptions in Assumption 3.

Recall that the convex Lagrange multipliers $\Gamma_P \coloneqq \{\gamma \in \mathbb{R}_+ : A_{\text{obj}} + \gamma A_1 \succeq 0\} = [\gamma_-, \gamma_+]$ is a compact interval. By (4.2) and Sion's Minimax Theorem, we have that

$$\begin{aligned}\text{Opt} &= \min_{x\in\mathbb{R}^n}\max_{\gamma\in[\gamma_-,\gamma_+]}\left(q_{\text{obj}}(x) + \gamma q_1(x)\right)\\ &= \max_{\gamma\in[\gamma_-,\gamma_+]}\left[\inf_{x\in\mathbb{R}^n}\left(q_{\text{obj}}(x) + \gamma q_1(x)\right)\right].\end{aligned}$$

We will define $d : [\gamma_-, \gamma_+] \to [-\infty, \infty)$ to be the extended-real-valued function given in square brackets above. This is the objective function in the dual optimization problem.

Let γ^* maximize $\lambda_{\min}(A_{\text{obj}} + \gamma^* A_1)$ over maximizers of $d(\gamma)$ and let $\mu^* \coloneqq \lambda_{\min}(A_{\text{obj}} + \gamma^* A_1)$. We will refer to μ^* as the *implicit regularity* of the GTRS.

Note that given any interval $\gamma^* \in [\tilde{\gamma}_-, \tilde{\gamma}_+] \subseteq [\gamma_-, \gamma_+]$, it holds that

$$\min_{x\in\mathbb{R}^n}\max_{\gamma\in[\tilde{\gamma}_-,\tilde{\gamma}_+]}\left(q_{\text{obj}}(x) + \gamma q_1(x)\right)$$

is a convex minimization problem achieving the same objective value as the GTRS.

To compare with the convex reformulation of the GTRS given in (4.2), note that the quadratic functions q_- and q_+ in (4.2) do *not* admit strong convexity. In contrast, if we pick $[\tilde{\gamma}_-, \tilde{\gamma}_+]$ to be a small interval containing γ^*, then $\tilde{q}_- \coloneqq q_{\text{obj}} + \tilde{\gamma}_- q_1$ and $\tilde{q}_+ \coloneqq q_{\text{obj}} + \tilde{\gamma}_+ q_1$ both have strong convexity on the order μ^*. This strong convexity allows us to achieve first order algorithms with convergence rates akin to those for smooth strongly convex minimization.

Partial results

In ongoing work, we have designed a pair of algorithms ConstructConvex and SolveConvex exploiting the above ideas. Informally, ConstructConvex performs a binary search using *low-resolution* minimum eigenvalue calls to find $\tilde{q}_-$ and $\tilde{q}_+$ such that $\text{Opt} = \min_{x\in\mathbb{R}^n} \max(\tilde{q}_-(x), \tilde{q}_+(x))$ where $\tilde{q}_-$ and $\tilde{q}_+$ both have strong convexity on the order $\max(\mu^*, \epsilon)$. In this algorithm, eigenvalues can only be computed up to accuracy $\approx \max(\epsilon, \mu^*)$ in order to achieve the desired running time. Then, SolveConvex applies Nesterov's accelerated gradient descent scheme for minimax problems to the computed reformulation [76]. Specifically, we prove the following guarantees[1] on ConstructConvex and SolveConvex.

Theorem 18 (Informal)**.** *Suppose Assumption 3 holds. With probability at least $1-p$, ConstructConvex will either output:*

- *"Case 1," $\tilde{\gamma}_-$, $\tilde{\gamma}_+$ such that $\gamma^* \in [\tilde{\gamma}_-, \tilde{\gamma}_+]$ and $\max_{\gamma\in\{\tilde{\gamma}_-,\tilde{\gamma}_+\}} (\lambda_{\min}(A_{\text{obj}} + \gamma A_1)) = \Omega(\mu^*)$, or*
- *"Case 2," $\tilde{\gamma}$ such that $\lambda_{\min}(A_{\text{obj}} + \tilde{\gamma} A_1) = \Omega(\max(\mu^*, \epsilon))$ and $(A_{\text{obj}} + \tilde{\gamma} A_1)^{-1}(b_{\text{obj}} + \tilde{\gamma} b_1)$ is an ϵ-optimizer of the GTRS.*

in time

$$\tilde{O}\left(\frac{N\sqrt{\zeta}}{\sqrt{\max(\mu^*, \epsilon)}} \log\left(\frac{n}{p}\right) \log\left(\frac{\zeta}{\epsilon\xi}\right)^2\right).$$

Here, note that if ConstructConvex succeeds in Case 2, then we can recover an ϵ-optimizer of the GTRS via conjugate gradient descent. On the other hand, if ConstructConvex succeeds in Case 1, then we can recover an ϵ-optimizer of the GTRS via SolveConvex as follows:

Theorem 19 (Informal)**.** *Suppose Assumption 3 holds. If $\gamma^* \in [\tilde{\gamma}_-, \tilde{\gamma}_+]$ and $\max_{\gamma\in\{\tilde{\gamma}_-,\tilde{\gamma}_+\}}(\lambda_{\min}(A_{\text{obj}} + \gamma A_1)) = \Omega(\mu^*)$, then SolveConvex outputs an ϵ-optimizer of the GTRS in time*

$$O\left(\frac{N\sqrt{\zeta}}{\sqrt{\mu^*}} \log\left(\frac{\zeta}{\epsilon\xi}\right)\right).$$

We remark that while our procedure for solving the GTRS consists of two separate algorithms, if one plans to solve a single instance of the GTRS multiple times with increasing accuracy, then ConstructConvex only needs to succeed in Case 1 at most once. That is, after the first Case 1 success of ConstructConvex, we have an *exact* reformulation of the GTRS and only need to run SolveConvex to increasing accuracy.

Open questions

While many new algorithms for the GTRS have been proposed in recent years in the optimization community, including the ones mentioned in this proposal, there is a lack of numerical evidence comparing these different algorithms. In ongoing and future work, we intend to build a codebase for solving instances of the GTRS numerically using our algorithm from [99], the algorithm proposed in this section, as well as algorithms previously proposed in the literature [2, 13, 57, 58]. This is an important step to take to bring ideas from this domain closer to practically applicable.

6.3 Implicit regularity for QCQPs

This section is based on ongoing work with Fatma Kılınç-Karzan.

A number of the ideas we are currently exploring related to implicit regularity in the GTRS (Section 6.2) do not *explicitly* depend on the fact that we were working with an instance of the GTRS. In fact, some of the ideas there generalize to the setting of arbitrary QCQPs as long as we assume some sort of implicit regularity. This leads to our second proposed direction: Can we design first-order algorithms for solving the SDP relaxation of a QCQP whose running times scale in ϵ as $\approx \frac{1}{\sqrt{\mu^*}} \log\left(\frac{1}{\epsilon}\right)$ and *do not* make use of interior point methods?

[1] To simplify the presentation of these informal statements, we will assume that $\mu^* \leq \xi$.

Related work

This work hopes to reiterate a theme in a number of recent papers [2, 10, 21, 26, 37, 43, 99, 101] that *structures that imply SDP exactness often coincide or overlap with structures that may be exploited in solving SDPs.* For example, Gatermann and Parrilo [43] showed that group symmetries can be exploited to reduce the size (and the corresponding solve-times) for SDPs with known group structures. Separately, Beck [10] (along with work presented in Chapter 2) show that a particular form of group symmetry can lead to SDP exactness. Another example of this theme closely related to work in this proposal is that of the generalized trust-region subproblem, where a number of custom algorithms [2, 58, 99] have been developed. More general instantiations of this idea can be found in the Burer-Montiero method [21, 28], where a dense matrix variable is replaced by an explicit low-rank factorization, and in recent *storage-optimal* methods for solving SDPs [37].

On the technical side, the work presented in this section is related to recent work on understanding the effects of noise in accelerated gradient descent [7, 29, 88].

Preliminaries

Consider a general QCQP of the form

$$\min_{x\in\mathbb{R}^n} \left\{q_{\text{obj}}(x) :\, q_i(x) \leq 0,\, \forall i\in[m]\right\}.$$

As always, write $q_{\text{obj}}(x) = x^\top A_{\text{obj}}x + 2b_{\text{obj}}^\top x + c_{\text{obj}}$. Similarly define $A_i \in \mathbb{S}^n$, $b_i \in \mathbb{R}^n$, and $c_i \in \mathbb{R}$. We will assume that there exists $\gamma \in \mathbb{R}^m_+$ such that $A_{\text{obj}} + A(\gamma) \succ 0$. Recall here that $A(\gamma) \coloneqq \sum_{i=1}^m \gamma_i A_i$. We will further assume the SDP relaxation is strictly feasible.

These assumptions ensure strong duality in the SDP relaxation and its SDP dual and that both programs achieve their optimum values, i.e.,

$$\min_{x\in\mathbb{R}^n, X\in\mathbb{S}^n} \left\{ \left\langle \begin{pmatrix} A_{\text{obj}} & b_{\text{obj}} \\ b_{\text{obj}}^\top & c_{\text{obj}} \end{pmatrix}, \begin{pmatrix} X & x \\ x^\top & 1 \end{pmatrix} \right\rangle :\, \begin{array}{l} \left\langle \begin{pmatrix} A_i & b_i \\ b_i^\top & c_i \end{pmatrix}, \begin{pmatrix} X & x \\ x^\top & 1 \end{pmatrix} \right\rangle \leq 0,\, \forall i\in[m] \\ \begin{pmatrix} X & x \\ x^\top & 1 \end{pmatrix} \succeq 0 \end{array} \right\} \tag{6.3}$$

$$= \max_{\gamma\in\mathbb{R}^m,\, t\in\mathbb{R}} \left\{ t :\, \begin{array}{l} \begin{pmatrix} A_{\text{obj}} + A(\gamma) & b_{\text{obj}} + b(\gamma) \\ b_{\text{obj}}^\top + b(\gamma)^\top & c_{\text{obj}} + c(\gamma) - t \end{pmatrix} \succeq 0 \\ \gamma \geq 0 \end{array} \right\}. \tag{6.4}$$

Definition 12. Let γ^* denote a maximizer of $\lambda_{\min}(A_{\text{obj}} + A(\gamma))$ among maximizers of (6.3) and set $\mu^* \coloneqq \lambda_{\min}(A_{\text{obj}} + A(\gamma^*))$. □

The following lemma explains why we have at times referred to implicit regularity as a strong form of exactness.

Lemma 20. *Suppose $\mu^* > 0$ and define $x^* \coloneqq -(A_{\text{obj}} + A(\gamma^*))^{-1}(b_{\text{obj}} + b(\gamma^*))$. Then, $(x^*, x^*(x^*)^\top)$ is the unique optimum solution to the SDP relaxation. In particular, x^* is also the unique optimal solution of the QCQP and objective value exactness holds.*

Proof. Let (x, X) denote an optimal solution to the SDP and (γ^*, t^*) an optimal solution of the dual SDP for which $A_{\text{obj}} + A(\gamma^*) \succ 0$. By complementary slackness,

$$\left\langle \begin{pmatrix} A_{\text{obj}} + A(\gamma^*) & b_{\text{obj}} + b(\gamma^*) \\ (b_{\text{obj}} + b(\gamma^*))^\top & c_{\text{obj}} + c(\gamma^*) - t^* \end{pmatrix}, \begin{pmatrix} X & x \\ x^\top & 1 \end{pmatrix} \right\rangle = 0.$$

As $A_{\text{obj}} + A(\gamma^*) \succ 0$, the matrix on the left of the inner product must have rank at least n (in fact exactly n) so that the matrix on the right of the inner product has rank one. We deduce that $X = xx^\top$ and that x is the unique optimizer of $x \mapsto x^\top(A_{\text{obj}} + A(\gamma^*))x + 2(b_{\text{obj}} + b(\gamma^*))^\top x + (c_{\text{obj}} + c(\gamma^*) - t)$, i.e., $x = x^*$. ■

Partial results

In this direction, we will assume $\mu^* > 0$ and attempt to design first order algorithms for such problems with improved rates of convergence.

Again, note that we can get a convex reformulation of the QCQP by finding a set $\gamma^* \in S \subseteq \Gamma_P$.

Lemma 21. *Suppose $\mu^* > 0$ and let $\gamma^* \in S \subseteq \Gamma_P$. Then,*

$$\mathrm{Opt} = \min_{x\in\mathbb{R}^n} \sup_{\gamma\in S} \left(q_{\mathrm{obj}}(x) + \langle \gamma, q(x)\rangle\right).$$

Our approach will be based on the algorithms ConstructConvex and SolveConvex in Section 6.2. Our first algorithm will need to find a "simple" set S such that $\gamma^* \in S \subseteq \Gamma_P$ and $\min_{\gamma\in S}\left(\lambda_{\min}(A_{\mathrm{obj}} + A(\gamma))\right) = \Omega(\mu^*)$. Here, "simple" could for example mean a polytope or a ball. Our second algorithm will then need to solve the $\Omega(\mu^*)$-strongly-convex minimax problem

$$\min_{x\in\mathbb{R}^n} \sup_{\gamma\in S} \left(q_{\mathrm{obj}}(x) + \langle \gamma, q(x)\rangle\right) \tag{6.5}$$

in time $\approx \frac{1}{\sqrt{\mu^*}} \log\left(\frac{1}{\epsilon}\right)$.

We briefly discuss some partial results and ideas towards the second algorithm. A straightforward extension of Nesterov's accelerated gradient descent scheme for minimax problems shows that we can solve the convex reformulation in (6.5) in $\approx \frac{1}{\sqrt{\mu^*}} \log\left(\frac{1}{\epsilon}\right)$ iterations of a "gradient mapping" step. Each gradient mapping step requires the true optimizer of the auxiliary problem:

$$\min_{x\in\mathbb{R}^n} \max_{\gamma\in S} \left(\langle (A_{\mathrm{obj}} + A(\gamma))\bar{x} + b_{\mathrm{obj}} + b(\gamma), x - \bar{x}\rangle + q_{\mathrm{obj}}(\bar{x}) + \langle \gamma, q(\bar{x})\rangle\right) + L\left\|x - \bar{x}\right\|^2 \tag{6.6}$$

for some $L > 0$ and $\bar{x} \in \mathbb{R}^n$.

In the setting where S is a polytope, it is possible to solve this auxiliary problem exactly in time linear in n but exponential in the number of vertices in S. This route is straightforward but is unlikely to be useful when m, the number of constraints, is large. In the setting where S is a ball, it is possible to reformulate this auxiliary problem as a second-order-cone program (SOCP): Suppose $S = B(\bar{\gamma}, \delta)$. Then, (6.6) has the same optimizer as the following second-order-cone-representable program.

$$\min_{x\in\mathbb{R}^n} L\left\|x - \bar{x}\right\|^2 + 2\left\langle (A_{\mathrm{obj}} + A(\bar{\gamma}))\bar{x} + b_{\mathrm{obj}} + b(\bar{\gamma}), x - \bar{x}\right\rangle + \delta\left\|q(\bar{x}) + \left(\langle A_i\bar{x} + b_i, x - \bar{x}\rangle\right)_{i=1}^m\right\|.$$

We may then use interior point methods to solve this second-order-cone program up to an arbitrary accuracy ϵ. While this direction is more likely to scale for more constraints, it is unclear how the bounds for Nesterov's accelerated gradient descent scheme for minimax problems behave when exact optimizers are replaced with ϵ-optimizer.

Open questions

This project is still in its early stages and there are many remaining questions. First, how accurate do the gradient maps, i.e., the optimizers of (6.6), need to be for error not to accumulate in the accelerated gradient scheme? Our current approach to tackle this question is to adapt analyses of accelerated gradient descent where gradients are computed only approximately. In this direction, Cohen et al. [29] gave a bound on the expectation of the primal gap as a function of the radius of the instance and the expected noise in the standard smooth strongly convex setting. One possible direction is to extend this analysis to the setting of minimax optimization problems with noise corrupted gradients. The second major question in this project is: How can we compute the ball $S \ni \gamma^*$? In this direction, we plan to revisit first-order algorithms for general SDPs as well as possible connections with storage-optimal approaches to semidefinite programs [37].

Bibliography

[1] E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *IEEE Trans. Inform. Theory*, 62(1):471–487, 2015.

[2] S. Adachi and Y. Nakatsukasa. Eigenvalue-based algorithm and analysis for nonconvex QCQP with one constraint. *Math. Program.*, 173:79–116, 2019.

[3] J. Agler, W. Helton, S. McCullough, and L. Rodman. Positive semidefinite matrices with a given sparsity pattern. *Linear Algebra Appl.*, 107:101–149, 1988.

[4] K. M. Anstreicher and S. Burer. Computable representations for convex hulls of low-dimensional quadratic forms. *Math. Program.*, 124:33–43, 2010.

[5] C.J. Argue, F. Kılınç-Karzan, and A. L. Wang. Necessary and sufficient conditions for rank-one generated cones. *arXiv preprint*, 2007.07433, 2020.

[6] A. Atamtürk, A. Gómez, and S. Han. Sparse and smooth signal estimation: Convexification of ℓ_0 formulations. *arXiv preprint*, 1811.02655, 2018.

[7] M. Baes. Estimate sequence methods: extensions and approximations. Technical report, Institute for Operations Research, ETH, Zürich, Switzerland, 2009.

[8] X. Bao, N. V. Sahinidis, and M. Tawarmalani. Semidefinite relaxations for quadratically constrained quadratic programming: A review and comparisons. *Math. Program.*, 129:129, 2011.

[9] A. Barvinok. Feasibility testing for systems of real quadratic equations. *Discrete Comput. Geom.*, 10: 1–13, 1993.

[10] A. Beck. Quadratic matrix programming. *SIAM J. Optim.*, 17(4):1224–1238, 2007.

[11] A. Beck and M. Teboulle. A convex optimization approach for minimizing the ratio of indefinite quadratic functions over an ellipsoid. *Math. Program.*, 118:13–35, 2009.

[12] A. Beck, Y. Drori, and M. Teboulle. A new semidefinite programming relaxation scheme for a class of quadratic matrix problems. *Oper. Res. Lett.*, 40(4):298–302, 2012.

[13] A. Ben-Tal and D. den Hertog. Hidden conic quadratic representation of some nonconvex quadratic optimization problems. *Math. Program.*, 143:1–29, 2014.

[14] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization*, volume 2 of *MPS-SIAM Ser. Optim.* SIAM, 2001.

[15] A. Ben-Tal and M. Teboulle. Hidden convexity in some nonconvex quadratically constrained quadratic programming. *Math. Program.*, 72:51–63, 1996.

[16] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*, volume 28 of *Princeton Ser. Appl. Math.* Princeton University Press, 2009.

[17] D. Bienstock and A. Michalka. Polynomial solvability of variants of the Trust-Region Subproblem. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 380–390. SIAM, 2014.

[18] G. Blekherman, R. Sinn, and M. Velasco. Do sums of squares dream of free resolutions? *SIAM J. Appl. Algebra Geom.*, 1:175–199, 2017.

[19] S. Burer. A gentle, geometric introduction to copositive optimization. *Math. Program.*, 151:89–116, 2015.

[20] S. Burer and F. Kılınç-Karzan. How to convexify the intersection of a second order cone and a nonconvex quadratic. *Math. Program.*, 162:393–429, 2017.

[21] S. Burer and R. D.C. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Math. Program.*, 95:329–357, 2003.

[22] S. Burer and B. Yang. The trust region subproblem with non-intersecting linear constraints. *Math. Program.*, 149:253–264, 2014.

[23] S. Burer and Y. Ye. Exact semidefinite formulations for a class of (random and non-random) nonconvex quadratic programs. *Math. Program.*, 181:1–17, 2019.

[24] M. D. Bustamante, P. Mellon, and M. V. Velasco. Solving the problem of simultaneous diagonalization of complex symmetric matrices via congruence. *SIAM Journal on Matrix Analysis and Applications*, 41 (4):1616–1629, 2020.

[25] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *SIAM Rev.*, 57(2):225–251, 2015.

[26] Y. Carmon and J. C. Duchi. Analysis of Krylov subspace solutions of regularized nonconvex quadratic problems. pages 10728–10738, 2018.

[27] S. Ceria and J. Soares. Convex programming for disjunctive convex optimization. *Math. Program.*, 86: 595–614, 1999.

[28] D. Cifuentes and A. Moitra. Polynomial time guarantees for the Burer-Monteiro method. *arXiv preprint*, 1912.01745, 2019.

[29] M. Cohen, J. Diakonikolas, and L. Orecchia. On acceleration with noise-corrupted gradients. In *International Conference on Machine Learning*, pages 1019–1028. PMLR, 2018.

[30] A. R. Conn, N. I. Gould, and P. L. Toint. *Trust Region Methods*, volume 1 of *MPS-SIAM Ser. Optim.* SIAM, 2000.

[31] M. K. de Carli Silva and L. Tunçel. A notion of Total Dual Integrality for convex, semidefinite, and extended formulations. *SIAM J. Discrete Math.*, 34(1):470–496, 2020.

[32] E. de Klerk and R. Sotirov. Exploiting group symmetry in semidefinite programming relaxations of the quadratic assignment problem. *Math. Program.*, 122:225–246, 2010.

[33] E. de Klerk, D. V. Pasechnik, and A. Schrijver. Reduction of symmetric semidefinite programs using the regular $*$-representation. *Math. Program.*, 109:613–624, 2007.

[34] E. de Klerk, C. Dobre, and D. V. Pasechnik. Numerical block diagonalization of matrix *-algebras with application to semidefinite programming. *Math. Program.*, 129:91–111, 2011.

[35] A. P. Dempster. Covariance selection. *Biometrics*, 28:157–175, 1972.

[36] L. L. Dines. On the mapping of quadratic forms. *Bull. Amer. Math. Soc.*, 47(6):494–498, 1941.

[37] L. Ding, A. Yurtsever, V. Cevher, J. A. Tropp, and M. Udell. An optimal-storage approach to semidefinite programming using approximate complementarity. *arXiv preprint*, 1902.03373, 2019.

[38] H. Dong and J. Linderoth. On valid inequalities for quadratic programming with continuous variables and binary indicators. In *Integer Programming and Combinatorial Optimization (IPCO 2013)*, pages 169–180. Springer, 2013.

[39] J. Feng, G. Lin, R. Sheu, and Y. Xia. Duality and solutions for quadratic programming over single non-homogeneous quadratic constraint. *J. Global Optim.*, 54(2):275–293, 2012.

[40] A. L. Fradkov and V. A. Yakubovich. The S-procedure and duality relations in nonconvex problems of quadratic programming. *Vestnik Leningrad Univ. Math.*, 6:101–109, 1979.

[41] A. Frangioni and C. Gentile. Perspective cuts for a class of convex 0–1 mixed integer programs. *Math. Program.*, 106:225–236, 2006.

[42] T. Fujie and M. Kojima. Semidefinite programming relaxation for nonconvex quadratic programs. *J. Global Optim.*, 10(4):367–380, 1997.

[43] K. Gatermann and P. A. Parrilo. Symmetry groups, semidefinite programs, and sums of squares. *J. Pure Appl. Algebra*, 192(1-3):95–128, 2004.

[44] D. Gijswijt. Matrix algebras and semidefinite programming techniques for codes. *arXiv preprint*, 1007.0906, 2010.

[45] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995.

[46] G. H. Golub, P. C. Hansen, and D. P. O'Leary. Tikhonov regularization and total least squares. *SIAM J. Matrix Anal. Appl.*, 21(1):185–194, 1999.

[47] R. Grone, C. R. Johnson, E. M. Sá, and H. Wolkowicz. Positive definite completions of partial Hermitian matrices. *Linear Algebra Appl.*, 58:109–124, 1984.

[48] O. Günlük and J. Linderoth. Perspective reformulations of mixed integer nonlinear programs with indicator variables. *Math. Program.*, 124:183–205, 2010.

[49] E. Hazan and T. Koren. A linear-time algorithm for trust region problems. *Math. Program.*, 158: 363–381, 2016.

[50] R. Hildebrand. Spectrahedral cones generated by rank 1 matrices. *J. Global Optim.*, 64:349–397, 2016.

[51] J. Hiriart-Urruty. Potpourri of conjectures and open questions in nonlinear analysis and optimization. *SIAM review*, 49(2):255–273, 2007.

[52] N. Ho-Nguyen and F. Kılınç-Karzan. A second-order cone based approach for solving the Trust Region Subproblem and its variants. *SIAM J. Optim.*, 27(3):1485–1512, 2017.

[53] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, 2012.

[54] Y. Hsia and R. Sheu. Trust region subproblem with a fixed number of additional linear inequality constraints has polynomial complexity. *arXiv preprint*, (arXiv:1312.1398), 2013.

[55] V. Jeyakumar and G. Y. Li. Trust-region problems with linear inequality constraints: Exact SDP relaxation, global optimality and robust optimization. *Math. Program.*, 147:171–206, 2014.

[56] R. Jiang and D. Li. Simultaneous diagonalization of matrices and its applications in quadratically constrained quadratic programming. *SIAM J. Optim.*, 26(3):1649–1668, 2016.

[57] R. Jiang and D. Li. Novel reformulations and efficient algorithms for the Generalized Trust Region Subproblem. *SIAM J. Optim.*, 29(2):1603–1633, 2019.

[58] R. Jiang and D. Li. A linear-time algorithm for generalized trust region problems. *SIAM J. Optim.*, 30 (1):915–932, 2020.

[59] A. Joyce and B. Yang. Convex hull results on quadratic programs with non-intersecting constraints. *Optimization Online preprint*, (http://www.optimization-online.org/DB_HTML/2021/05/8421.html), July 2021.

[60] F. Kılınç-Karzan and A. L. Wang. Exactness in SDP relaxations of QCQPs: Theory and applications. Tut. in Oper. Res. INFORMS, 2021.

[61] L. Kronecker. *Collected works*. American Mathematical Society, 1968.

[62] P. Lancaster and L. Rodman. Canonical forms for Hermitian matrix pairs under strict equivalence and congruence. *SIAM Review*, 47(3):407–443, 2005.

[63] M. Laurent and S. Poljak. On a positive semidefinite relaxation of the cut polytope. *Linear Algebra Appl.*, 223-224:439–461, 1995.

[64] T. H. Le and T. N. Nguyen. Simultaneous diagonalization via congruence of hermitian matrices: some equivalent conditions and a numerical solution. *arXiv preprint*, (arXiv:2007.14034), 2020.

[65] M. Locatelli. Some results for quadratic problems with one or two quadratic constraints. *Oper. Res. Lett.*, 43(2):126–131, 2015.

[66] M. Locatelli. Exactness conditions for an SDP relaxation of the extended trust region problem. *Oper. Res. Lett.*, 10(6):1141–1151, 2016.

[67] M. Locatelli. KKT-based primal-dual exactness conditions for the Shor relaxation. *arXiv preprint*, 2011.05033, 2020.

[68] H. Luo, Y. Chen, X. Zhang, and D. Li. Effective algorithms for optimal portfolio deleveraging problem with cross impact. *arXiv preprint*, (arXiv:2012.07368), 2020.

[69] A. Megretski. Relaxations of quadratic programs in operator theory and system analysis. In *Systems, Approximation, Singular Integral Operators, and Related Topics*, pages 365–392. Birkhäuser Basel, 2001.

[70] D. G. Mixon, S. Villar, and R. Ward. Clustering subgaussian mixtures by semidefinite programming. *arXiv preprint*, 1602.06612, 2016.

[71] S. Modaresi and J. P. Vielma. Convex hull of two quadratic or a conic quadratic and a quadratic inequality. *Math. Program.*, 164:383–409, 2017.

[72] J. J. Moré. Generalizations of the Trust Region Problem. *Optimization methods and Software*, 2(3-4): 189–209, 1993.

[73] J. J. Moré and D. C. Sorensen. Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*, 4:553–572, 1983.

[74] T. S. Motzkin and O. Taussky. Pairs of matrices with property L. II. 80(2):387–401, 1955.

[75] Y. Nesterov. Quality of semidefinite relaxation for nonconvex quadratic optimization. Technical Report 1997019, Université Catholique de Louvain, Center for Operations Research and Econometrics (CORE), 1997.

[76] Y. Nesterov. *Lectures on convex optimization*, volume 137 of *Springer Optimization and Its Applications*. Springer, 2018.

[77] T. Nguyen, V. Nguyen, T. Le, and R. Sheu. On simultaneous diagonalization via congruence of real symmetric matrices. *arXiv preprint*, (arXiv:2004.06360), 2020.

[78] K. O'meara and C. Vinsonhaler. On approximately simultaneously diagonalizable matrices. *Linear Algebra Appl.*, 412:39–74, 2006.

[79] G. Pataki. On the connection of facially exposed and nice cones. *J. Math. Anal. Appl.*, 400(1):211–221, 2013.

[80] V. I. Paulsen, S. C. Power, and R. R. Smith. Schur products and matrix completions. *J. Funct. Anal.*, 85(1):151–178, 1989.

[81] E. Phan-huy-Hao. Quadratically constrained quadratic programming: Some applications and a method for solution. *Z. Oper. Res.*, 26:105–119, 1982.

[82] I. Pólik and T. Terlaky. A survey of the S-lemma. *SIAM Rev.*, 49(3):371–418, 2007.

[83] B. T. Polyak. Convexity of quadratic transformations and its use in control and optimization. 99(3): 553–583, 1998.

[84] T. K. Pong and H. Wolkowicz. The generalized trust region subproblem. *Computational Optimization and Applications*, 58(2):273–322, 2014.

[85] M. V. Ramana. Polyhedra, spectrahedra, and semidefinite programming. In *Topics in Semidefinite and Interior-Point Methods*, volume 18 of *Fields Inst. Commun.*, pages 27–38. AMS, 1997.

[86] N. Rujeerapaiboon, K. Schindler, D. Kuhn, and W. Wiesemann. Size matters: Cardinality-constrained clustering and outlier detection via conic optimization. *SIAM J. Optim.*, 29(2):1211–1239, 2019.

[87] A. Santana and S. S. Dey. The convex hull of a quadratic constraint over a polytope. *SIAM J. Optim.*, 30(4):2983–2997, 2020.

[88] M. Schmidt, N. Le Roux, and F. Bach. Convergence rates of inexact proximal-gradient methods for convex optimization. *arXiv preprint*, 1109.2415, 2011.

[89] N. Z. Shor. Dual quadratic estimates in polynomial and boolean programming. *Ann. Oper. Res.*, 25: 163–168, 1990.

[90] S. Sojoudi and J. Lavaei. Exactness of semidefinite relaxations for nonlinear optimization problems with underlying graph structure. *SIAM J. Optim.*, 24(4):1746–1778, 2014.

[91] R. J. Stern and H. Wolkowicz. Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations. *SIAM J. Optim.*, 5(2):286–313, 1995.

[92] J. F. Sturm. Error bounds for linear matrix inequalities. *SIAM J. Optim.*, 10(4):1228–1248, 2000.

[93] J. F. Sturm and S. Zhang. On cones of nonnegative quadratic functions. *Math. Oper. Res.*, 28(2): 246–267, 2003.

[94] H. W. Turnbull and A. C. Aitken. *An introduction to the theory of canonical matrices.* Dover, 1961.

[95] F. Uhlig. A canonical form for a pair of real symmetric matrices that generate a nonsingular pencil. *Linear Algebra Appl.*, 14(3):189–209, 1976.

[96] F. Uhlig. A recurring theorem about pairs of quadratic forms and extensions: A survey. *Linear Algebra Appl.*, 25:219–237, 1979.

[97] A. L. Wang and R. Jiang. New notions of simultaneous diagonalizability of quadratic forms with applications to QCQPs. *arXiv preprint*, 2101.12141, 2021.

[98] A. L. Wang and F. Kılınç-Karzan. On convex hulls of epigraphs of QCQPs. In *Integer Programming and Combinatorial Optimization (IPCO 2020)*, pages 419–432. Springer, 2020.

[99] A. L. Wang and F. Kılınç-Karzan. The generalized trust region subproblem: solution complexity and convex hull results. *Math. Program.*, 2020. doi: 10.1007/s10107-020-01560-8. Forthcoming.

[100] A. L. Wang and F. Kılınç-Karzan. A geometric view of SDP exactness in QCQPs and its applications. *arXiv preprint*, 2011.07155, 2020.

[101] A. L. Wang and F. Kılınç-Karzan. On the tightness of SDP relaxations of QCQPs. *Math. Program.*, 2021. doi: 10.1007/s10107-020-01589-9. Forthcoming.

[102] L. Wei, A. Gómez, and S. Küçükyavuz. On the convexification of constrained quadratic optimization problems with indicator variables. In *Integer Programming and Combinatorial Optimization (IPCO 2020)*, pages 433–447. Springer, 2020.

[103] L. Wei, A. Gómez, and S. Küçükyavuz. Ideal formulations for constrained convex optimization problems with indicator variables. *arXiv preprint*, 2007.00107, 2020.

[104] K. Weierstrass. Zur Theorie der quadratischen und bilinearen Formen. *Monatsber. Akad. Wiss., Berlin*, pages 310–338, 1868.

[105] H. Wolkowicz, R. Saigal, and L. Vandenberghe. *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, volume 27 of *Internat. Ser. Oper. Res. Management Sci.* Springer, 2012.

[106] Y. Xia. On minimizing the ratio of quadratic functions over an ellipsoid. *Optim.*, 64(5):1097–1106, 2015.

[107] V. A. Yakubovich. S-procedure in nonlinear control theory. *Vestnik Leningrad Univ. Math.*, pages 62–77, 1971.

[108] B. Yang, K. Anstreicher, and S. Burer. Quadratic programs with hollows. *Math. Program.*, 170:541–553, 2018.

[109] Y. Ye. Approximating quadratic programming with bound and quadratic constraints. *Math. Program.*, 84:219–226, 1999.

[110] U. Yıldıran. Convex hull of two quadratic constraints is an LMI set. *IMA J. Math. Control Inform.*, 26 (4):417–450, 2009.

[111] J. Zhou and Z. Xu. A simultaneous diagonalization based SOCP relaxation for convex quadratic programs with linear complementarity constraints. *Optim. Letters*, 13(7):1615–1630, 2019.

[112] J. Zhou, S. Chen, S. Yu, and Y. Tian. A simultaneous diagonalization-based quadratic convex reformulation for nonconvex quadratically constrained quadratic program. *Optimization*, pages 1–17, 2020.