Noname manuscript No.
(will be inserted by the editor)

Penalized Semidefinite Programming for Quadratically-Constrained Quadratic Optimization

Ramtin Madani · Mohsen Kheirandishfard · Javad Lavaei · Alper Atamtürk

Received: date / Accepted: date

Abstract In this paper, we give a new penalized semidefinite programming approach for non-convex quadratically-constrained quadratic programs (QCQPs). We incorporate penalty terms into the objective of convex relaxations in order to retrieve feasible and near-optimal solutions for non-convex QCQPs. We introduce a generalized linear independence constraint qualification (GLICQ) criterion and prove that any GLICQ regular point that is sufficiently close to the feasible set can be used to construct an appropriate penalty term and recover a feasible solution. As a consequence, we describe a heuristic sequential procedure that preserves feasibility and aims to improve the objective value at each iteration. Numerical experiments on large-scale system identification problems as well as benchmark instances from the library of quadratic programming (QPLIB) demonstrate the ability of the proposed penalized semidefinite programs in finding near-optimal solutions for non-convex QCQP.

Keywords Semidefinite programming · nonconvex optimization · nonlinear programming · convex relaxation

PACS 87.55.de

Mathematics Subject Classification (2010) 65K05 · 90-08 · 90C26 · 90C22

This work is in part supported by the NSF Award 1809454. Javad Lavaei is supported by an AFOSR YIP Award and ONR N000141712933. Alper Atamtürk is supported, in part, by grant FA9550-10-1-0168 from the Office of the Assistant Secretary of Defense for Research & Engineering and the NSF Award 1807260.

R. Madani · M. Kheirandishfard
517 Nedderman Hall, University of Texas, Arlington, TX 76019, USA
E-mail: ramtin.madani@uta.edu,mohsen.kheirandishfard@mavs.uta.edu

J. Lavaei and A. Atamtürk
4175 Etcheverry Hall, University of California, Berkeley, CA 94720, USA
Tel.: +1-510-642-4559
E-mail: lavaei@berkeley.edu,atamturk@berkeley.edu

1 Introduction

Polynomial optimization is the problem of minimizing a polynomial function within a feasible set that is characterized by polynomial functions. Physics laws and characteristics of dynamical systems are widely modeled using polynomials. As a result, polynomial optimization arises in numerous scientific and engineering applications, such as electric power systems [42, 43], imaging science [20], signal processing [40], automatic control [1, 19], quantum mechanics [11, 27], and cybersecurity [15, 16]. This paper is on a subclass of polynomial optimization, referred to as quadratically-constrained quadratic programming (QCQP), where polynomials are quadratic. The results in this paper extend to polynomial optimization by transformation to QCQP, as shown in Appendix A. The development of optimization algorithms for QCQP has been an active area of research for decades. Due to the barriers imposed by NP-hardness, the focus of some of the research efforts has shifted from designing general-purpose algorithms to specialized methods that are robust and scalable for specific application domains. Notable examples for which methods with guaranteed performance have been offered in the literature include the problems of multisensor beamforming in communication theory [23], phase retrieval in signal processing [12], and in machine learning [5, 46].

This paper advances a popular framework for the global analysis of QCQP that builds hierarchies of semidefinite programming (SDP) [50] relaxations [14, 29, 37, 45, 51, 56]. SDP has been critically important for constructing strong convex relaxations of non-convex optimization problems. In particular, forming hierarchies of SDP relaxations [14, 29, 36–38, 45, 51, 56] has been shown to yield the convex hull of non-convex problems. Geomans and Williamson [24] show that the SDP relaxation objective is within 14% of the optimal value for the MAXCUT problem on graphs with non-negative weight. SDP relaxations have played a central role in developing numerous approximation algorithms for non-convex optimization problems [25, 26, 39, 49, 62–65]. They are also used within branch-and-bound algorithms [9, 13] for non-convex optimization. One of the primary challenges for the application of SDP hierarchies beyond small-scale instances is the rapid growth of dimensionality. In response, some studies have exploited sparsity and structural patterns to boost efficiency [7, 33, 34, 47, 48]. Another direction, pursued in [1, 2, 6, 44, 53], is to use lower-complexity relaxations as alternatives to computationally demanding semidefinite programming relaxations. A relaxation is said to be *exact* if it has the same optimal objective value as the original problem. The exactness of the SDP relaxation has been verified for a variety of problems [10, 33, 35, 58].

1.1 Contributions

This paper is concerned with non-convex quadratically-constrained quadratic programs for which SDP relaxations are inexact. In order to recover feasible points for QCQP, we incorporate a linear penalty term into the objective of SDP relaxations and show that feasible and near-globally optimal points can be obtained for the original QCQP by solving the resulting penalized SDPs. The penalty term is based on an arbitrary initial point. Our first result states that if the initial point is feasible and satisfies the linear independence constraint qualification (LICQ) condition, then penalized SDP produces a unique solution that is feasible for the original QCQP and its objective value is not worse than that of the initial point. Our second result states that if the initial point is infeasible, but instead is sufficiently close to the feasible set and satisfies a generalized LICQ condition, then the unique optimal solution to penalized SDP is feasible for QCQP. Lastly, motivated by these results on constructing feasible points, we propose a heuristic sequential procedure for non-convex QCQP and demonstrate its performance on benchmark instances from the QPLIB library [21] as well as on large-scale system identification problems.

The success of sequential frameworks and penalized SDP in solving bilinear matrix inequalities (BMIs) is demonstrated in [28, 30, 32]. In [4], it is shown that penalized SDP is able to find the roots of overdetermined systems of polynomial equations. Moreover, the incorporation of penalty terms into the objective of SDP relaxations are proven to be effective for solving non-convex optimization problems in power systems [42, 43, 66, 67]. These papers show that penalizing certain physical quantities in power network optimization problems such as reactive power loss or thermal loss facilitates the recovery of feasible points from convex relaxations. In [28], a sequential framework is introduced for solving BMIs without theoretical guarantees. Papers [30, 32] investigate this approach further and offer theoretical results through the notion of generalized Mangasarian-Fromovitz regularity condition. However, these conditions are not valid in the presence of equality constraints and for general QCQPs. Motivated by the success of penalized SDP, this paper offers a theoretical framework for general QCQP and, by extension, polynomial optimization problems.

1.2 Notations

Throughout the paper, scalars, vectors, and matrices are respectively shown by italic letters, lower-case italic bold letters, and upper-case italic bold letters. The symbols $\mathbb{R}$, $\mathbb{R}^n$, and $\mathbb{R}^{n\times m}$ denote the sets of real scalars, real vectors of size n, and real matrices of size $n \times m$, respectively. The set of $n \times n$ real symmetric matrices is shown by $\mathbb{S}_n$. For a given vector $\boldsymbol{a}$ and a matrix $\boldsymbol{A}$, the symbols a_i and A_{ij} respectively indicate the i^{th} element of $\boldsymbol{a}$ and the $(i, j)^{\text{th}}$ element of $\boldsymbol{A}$. The symbols $\langle \cdot\, , \cdot \rangle$ and $\|\cdot\|_{\text{F}}$ denote the Frobenius inner product and norm of matrices, respectively. The notation $|\cdot|$ represents either the absolute value operator or cardinality of a set, depending on the context. The notation $\|\cdot\|_2$ denotes the ℓ_2 norm of vectors, matrices, and matrix pencils. The $n \times n$ identity matrix is denoted by $\boldsymbol{I}_n$. The origin of $\mathbb{R}^n$ is denoted by $\boldsymbol{0}_n$. The superscript $(\cdot)^\top$ and the symbol $\operatorname{tr}\{\cdot\}$ represent the transpose and trace operators, respectively. Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, the notation $\sigma_{\min}(\boldsymbol{A})$ represents the minimum singular value of $\boldsymbol{A}$. The notation $\boldsymbol{A} \succeq 0$ means that $\boldsymbol{A}$ is symmetric positive-semidefinite. For a pair of $n\times n$ symmetric matrices $(\boldsymbol{A}, \boldsymbol{B})$ and proper cone $\mathcal{C} \subseteq \mathbb{S}_n$, the notation $\boldsymbol{A} \succeq_{\mathcal{C}} \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B} \in \mathcal{C}$, whereas $\boldsymbol{A} \succ_{\mathcal{C}} \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ belongs to the interior of $\mathcal{C}$. Given an integer $r > 1$, define $\mathcal{C}_r$ as the cone of $n \times n$ symmetric matrices whose $r \times r$ principal submatrices are all positive semidefinite. Similarly, define $\mathcal{C}_r^*$ as the dual cone of $\mathcal{C}_r$, i.e., the cone of $n \times n$

symmetric matrices whose every $r \times r$ principal submatrix is positive semidefinite (i.e., its factor-width is bounded by r). Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and two sets of positive integers $\mathcal{S}_1$ and $\mathcal{S}_2$, define $\boldsymbol{A}\{\mathcal{S}_1, \mathcal{S}_2\}$ as the submatrix of $\boldsymbol{A}$ obtained by removing all rows of $\boldsymbol{A}$ whose indices do not belong to $\mathcal{S}_1$, and all columns of $\boldsymbol{A}$ whose indices do not belong to $\mathcal{S}_2$. Moreover, define $\boldsymbol{A}\{\mathcal{S}_1\}$ as the submatrix of $\boldsymbol{A}$ obtained by removing all rows of $\boldsymbol{A}$ that do not belong to $\mathcal{S}_1$. Given a vector $\boldsymbol{a} \in \mathbb{R}^n$ and a set $\mathcal{F} \subseteq \mathbb{R}^n$, define $d_{\mathcal{F}}(\boldsymbol{a})$ as the minimum distance between $\boldsymbol{a}$ and members of $\mathcal{F}$. Given a pair of integers (n, r), the binomial coefficient "n choose r" is denoted by C^n_r. The notations $\nabla_{\boldsymbol{x}} f(\boldsymbol{a})$ and $\nabla^2_{\boldsymbol{x}} f(\boldsymbol{a})$, respectively, represent the gradient and Hessian of the function f, with respect to the vector $\boldsymbol{x}$, at a point $\boldsymbol{a}$.

1.3 Outline

The remainder of the paper is organized as follows. In Section 2, we review the standard SDP relaxation for QCQP. Section 3 presents the main results of the paper: the penalized SDP, its theoretical analysis on producing a feasible solution along with a generalized linear independence constraint qualification, and finally the sequential penalization procedure. In Section 4 we present numerical experiments to test the effectiveness of the sequential penalization approach for non-convex QCQPs from the library of quadratic programming instances (QPLIB) as well as large-scale system identification problems. Finally, we conclude in section 5 with a few final remarks.

2 Preliminaries

In this section, we review the lifting and reformulation-linearization as well as the standard convex relaxations of QCQP that are necessary for the development of the main results on penalized SDP in Section 3. Consider a general quadratically-constrained quadratic program (QCQP):

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad q_0(\boldsymbol{x}) \tag{1a}$$

$$\text{s.t.} \quad q_k(\boldsymbol{x}) \leq 0, \quad k \in \mathcal{I} \tag{1b}$$

$$q_k(\boldsymbol{x}) = 0, \quad k \in \mathcal{E}, \tag{1c}$$

where $\mathcal{I}$ and $\mathcal{E}$ index the sets of inequality and equality constraints, respectively. For every $k \in \{0\} \cup \mathcal{I} \cup \mathcal{E}$, $q_k : \mathbb{R}^n \to \mathbb{R}$ is a quadratic function of the form $q_k(\boldsymbol{x}) \triangleq \boldsymbol{x}^\top \boldsymbol{A}_k \boldsymbol{x} + 2\boldsymbol{b}_k^\top \boldsymbol{x} + c_k$, where $\boldsymbol{A}_k \in \mathbb{S}_n$, $\boldsymbol{b}_k \in \mathbb{R}^n$, and $c_k \in \mathbb{R}$. Denote $\mathcal{F}$ as the feasible set of the QCQP (1a)–(1c). To derive the optimality conditions for a given point, it is useful to define the Jacobian matrix of the constraint functions.

Definition 1 (Jacobian Matrix) For every $\hat{\boldsymbol{x}} \in \mathbb{R}^n$, the Jacobian matrix $\mathcal{J}(\hat{\boldsymbol{x}})$ for the constraint functions $\{q_k\}_{k\in\mathcal{I}\cup\mathcal{E}}$ is

$$\mathcal{J}(\hat{\boldsymbol{x}}) \triangleq [\nabla_{\boldsymbol{x}} q_1(\hat{\boldsymbol{x}}), \ldots, \nabla_{\boldsymbol{x}} q_{|\mathcal{I}\cup\mathcal{E}|}(\hat{\boldsymbol{x}})]^\top. \tag{2a}$$

For every $\mathcal{Q} \subseteq \mathcal{I} \cup \mathcal{E}$, define $\mathcal{J}_{\mathcal{Q}}(\hat{\boldsymbol{x}})$ as the submatrix of $\mathcal{J}(\hat{\boldsymbol{x}})$ resulting from the rows that belong to $\mathcal{Q}$.

Given a feasible point for the QCQP (1a)–(1c), the well-known linear independence constraint qualification (LICQ) condition can be used as a regularity criterion.

Definition 2 (LICQ Condition) A feasible point $\hat{\boldsymbol{x}} \in \mathcal{F}$ is LICQ regular if the rows of $\mathcal{J}_{\hat{\mathcal{B}}}(\hat{\boldsymbol{x}})$ are linearly independent, where $\hat{\mathcal{B}} \triangleq \{k \in \mathcal{I} \cup \mathcal{E} \mid q_k(\hat{\boldsymbol{x}}) = 0\}$ denotes the set of binding constraints at $\hat{\boldsymbol{x}}$.

Finding a feasible point for the QCQP (1a)–(1c), however, is NP-hard as the Boolean Satisfiability Problem (SAT) is a special case. Therefore, in Section 3, we introduce the notion of generalized LICQ as a regularity condition for both feasible and infeasible points.

SDP relaxation

A common approach for tackling the non-convex QCQP (1a)–(1c) is to introduce an auxiliary variable $\boldsymbol{X} \in \mathbb{S}_n$ to represent $\boldsymbol{x}\boldsymbol{x}^\top$. Then, the objective function (1a) and constraints (1b)–(1c) can be written as linear functions of $\boldsymbol{x}$ and $\boldsymbol{X}$. For every $k \in \{0\} \cup \mathcal{I} \cup \mathcal{E}$, define $\bar{q}_k : \mathbb{R}^n \times \mathbb{S}_n \to \mathbb{R}$ as

$$\bar{q}_k(\boldsymbol{x}, \boldsymbol{X}) \triangleq \langle \boldsymbol{A}_k, \boldsymbol{X} \rangle + 2\boldsymbol{b}_k^\top \boldsymbol{x} + c_k. \tag{3}$$

Consider the following relaxation of QCQP (1a)–(1c):

$$\begin{aligned}
\underset{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{X}\in\mathbb{S}_n}{\text{minimize}} \quad & \bar{q}_0(\boldsymbol{x}, \boldsymbol{X}) && && \text{(4a)}\\
\text{s.t.} \quad & \bar{q}_k(\boldsymbol{x}, \boldsymbol{X}) \leq 0, && k \in \mathcal{I} && \text{(4b)}\\
& \bar{q}_k(\boldsymbol{x}, \boldsymbol{X}) = 0, && k \in \mathcal{E} && \text{(4c)}\\
& \boldsymbol{X} - \boldsymbol{x}\boldsymbol{x}^\top \succeq_{\mathcal{C}_r} 0 && && \text{(4d)}
\end{aligned}$$

where the additional conic constraint (4d) is a convex relaxation of the equation $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^\top$ and

$$\mathcal{C}_r \triangleq \big\{ \boldsymbol{Y} \;\big|\; \boldsymbol{Y}\{\mathcal{K}, \mathcal{K}\} \succeq 0, \;\; \forall\, \mathcal{K} \subseteq \{1, \ldots, n\} \wedge |\mathcal{K}| = r \big\}. \tag{5}$$

We refer to the convex problem (4a)–(4d) as the $r \times r$ SDP relaxation of the QCQP (1a)–(1c). The choice $r = n$ yields the well-known semidefinite programming (SDP) relaxation. Additionally, in the homogeneous case (i.e., if $\boldsymbol{b}_0 = \boldsymbol{b}_1 = \cdots = \boldsymbol{b}_{|\mathcal{I}\cup\mathcal{E}|} = \boldsymbol{0}$), the case $r = 2$ leads to the second-order conic programming (SOCP) relaxation.

In the presence of affine constraints, the reformulation-linearization technique (RLT) of Sherali and Adams [57] can be used to produce additional inequalities with respect to $\boldsymbol{x}$ and $\boldsymbol{X}$ to strengthen convex relaxations (Appendix B).

If the relaxed problem (4a)–(4d) has an optimal solution $(\overset{*}{\boldsymbol{x}}, \overset{*}{\boldsymbol{X}})$ that satisfies $\overset{*}{\boldsymbol{X}} = \overset{*}{\boldsymbol{x}}\overset{*}{\boldsymbol{x}}^\top$, then the relaxation is said to be *exact* and $\overset{*}{\boldsymbol{x}}$ is a globally optimal solution for the QCQP (1a)–(1c). The next section offers a penalization method for addressing the case where relaxations are not exact.

3 Penalized SDP

If the relaxed problem (4a)–(4d) is not exact, the resulting solution is not necessarily feasible for the original QCQP (1a)–(1c). In this case, we use an initial point $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ (either feasible or infeasible) to revise the objective function, resulting in a *penalized SDP* of the form:

$$\underset{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{X}\in\mathbb{S}_n}{\text{minimize}} \quad \bar{q}_0(\boldsymbol{x},\boldsymbol{X}) + \eta(\mathrm{tr}\{\boldsymbol{X}\} - 2\hat{\boldsymbol{x}}^\top\boldsymbol{x} + \hat{\boldsymbol{x}}^\top\hat{\boldsymbol{x}}) \tag{6a}$$

$$\text{s.t.} \quad \bar{q}_k(\boldsymbol{x},\boldsymbol{X}) \leq 0, \qquad k \in \mathcal{I} \tag{6b}$$

$$\bar{q}_k(\boldsymbol{x},\boldsymbol{X}) = 0, \qquad k \in \mathcal{E} \tag{6c}$$

$$\boldsymbol{X} - \boldsymbol{x}\boldsymbol{x}^\top \succeq_{\mathcal{C}_r} 0 \tag{6d}$$

where $\eta > 0$ is a fixed penalty parameter. Note that the penalty term $\mathrm{tr}\{\boldsymbol{X}\} - 2\hat{\boldsymbol{x}}^\top\boldsymbol{x} + \hat{\boldsymbol{x}}^\top\hat{\boldsymbol{x}}$ is nonnegative and equals zero if and only if $\boldsymbol{X} = \hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^\top$. The penalization is *tight* if problem (6a)–(6d) has a unique optimal solution $(\overset{*}{\boldsymbol{x}}, \overset{*}{\boldsymbol{X}})$ that satisfies $\overset{*}{\boldsymbol{X}} = \overset{*}{\boldsymbol{x}}\overset{*}{\boldsymbol{x}}^\top$. In the next section, we give conditions under which penalized SDP is tight.

3.1 Theoretical analysis

The following theorem guarantees that if $\hat{\boldsymbol{x}}$ is feasible and satisfies the LICQ regularity condition (in Section 2), then the solution of (6a)–(6d) is guaranteed to be feasible for the QCQP (1a)–(1c) for an appropriate choice of η.

Theorem 1 *Let $\hat{\boldsymbol{x}}$ be a feasible point for the QCQP* (1a)–(1b) *that satisfies the LICQ condition. For sufficiently large $\eta > 0$, the SDP* (6a)–(6d) *has a unique optimal solution $(\overset{*}{\boldsymbol{x}}, \overset{*}{\boldsymbol{X}})$ such that $\overset{*}{\boldsymbol{X}} = \overset{*}{\boldsymbol{x}}\overset{*}{\boldsymbol{x}}^\top$. Moreover, $\overset{*}{\boldsymbol{x}}$ is feasible for* (1a)–(1c) *and satisfies $q_0(\overset{*}{\boldsymbol{x}}) \leq q_0(\hat{\boldsymbol{x}})$.*

If $\hat{\boldsymbol{x}}$ is not feasible, but satisfies a generalized LICQ regularity condition, introduced below, and is close enough to the feasible set $\mathcal{F}$, then the penalization is still tight for large enough $\eta > 0$. This result is described formally in Theorem 2. First, we define a distance measure from an arbitrary point in $\mathbb{R}^n$ to the feasible set of the problem.

Definition 3 (Distance Function) The distance function $d_{\mathcal{F}}\colon \mathbb{R}^n \to \mathbb{R}$ is defined as

$$d_{\mathcal{F}}(\hat{\boldsymbol{x}}) \triangleq \min\{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 \,|\, \boldsymbol{x} \in \mathcal{F}\}\cdot \tag{7}$$

Definition 4 (Generalized LICQ Condition) For every $\hat{\boldsymbol{x}} \in \mathbb{R}^n$, the set of *quasi-binding* constraints is defined as

$$\hat{\mathcal{B}} \triangleq \mathcal{E} \cup \left\{ k \in \mathcal{I} \,\middle|\, q_k(\hat{\boldsymbol{x}}) + \|\nabla q_k(\hat{\boldsymbol{x}})\|_2 d_{\mathcal{F}}(\hat{\boldsymbol{x}}) + \frac{\|\nabla^2 q_k(\hat{\boldsymbol{x}})\|_2}{2}\, d_{\mathcal{F}}(\hat{\boldsymbol{x}})^2 \geq 0 \right\}\cdot \tag{8}$$

The point $\hat{\boldsymbol{x}}$ is said to satisfy the GLICQ condition if the rows of $\mathcal{J}_{\hat{\mathcal{B}}}(\hat{\boldsymbol{x}})$ are linearly independent. Moreover, the singularity function $s : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$s(\hat{\boldsymbol{x}}) \triangleq \begin{cases} \sigma_{\min}(\mathcal{J}_{\hat{\mathcal{B}}}(\hat{\boldsymbol{x}})) & \text{if } \hat{\boldsymbol{x}} \text{ satisfies GLICQ} \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where $\sigma_{\min}(\mathcal{J}_{\hat{\mathcal{B}}}(\hat{\boldsymbol{x}}))$ denotes the smallest singular value of $\mathcal{J}_{\hat{\mathcal{B}}}(\hat{\boldsymbol{x}})$.

Observe that if $\hat{\boldsymbol{x}}$ is feasible, then $d_{\mathcal{F}}(\hat{\boldsymbol{x}}) = 0$, and GLICQ condition reduces to the LICQ condition. Moreover, GLICQ is satisfied if and only if $s(\hat{\boldsymbol{x}}) > 0$.

The next definition introduces the notion of matrix pencil corresponding to the QCQP (1a)–(1c), which will be used as a sensitivity measure.

Definition 5 (Pencil Norm) For the QCQP (1a)–(1c), define the corresponding matrix pencil $\boldsymbol{P} : \mathbb{R}^{|\mathcal{I}|} \times \mathbb{R}^{|\mathcal{E}|} \to \mathbb{S}_n$ as follows:

$$\boldsymbol{P}(\boldsymbol{\gamma}, \boldsymbol{\mu}) \triangleq \sum_{k \in \mathcal{I}} \gamma_k \boldsymbol{A}_k + \sum_{k \in \mathcal{E}} \mu_k \boldsymbol{A}_k. \tag{10}$$

Moreover, define the pencil norm $\|\boldsymbol{P}\|_2$ as

$$\|\boldsymbol{P}\|_2 \triangleq \max \left\{ \|\boldsymbol{P}(\boldsymbol{\gamma}, \boldsymbol{\mu})\|_2 \,\middle|\, \|\boldsymbol{\gamma}\|_2^2 + \|\boldsymbol{\mu}\|_2^2 = 1 \right\}, \tag{11}$$

which is upperbounded by $\sqrt{\sum_{k \in \mathcal{I} \cup \mathcal{E}} \|\boldsymbol{A}_k\|_2^2}$.

Theorem 2 *Let $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ satisfy the GLICQ condition for the QCQP* (1a)–(1b)*, and assume that*

$$d_{\mathcal{F}}(\hat{\boldsymbol{x}}) < \frac{s(\hat{\boldsymbol{x}})}{2\,(1 + C_{n-1,r-1})\,\|\boldsymbol{P}\|_2} \,, \tag{12}$$

where $C_{n-1,r-1}$ is the binomial coefficient "$n-1$ choose $r-1$" and the distance function $d_{\mathcal{F}}(\cdot)$, sensitivity function $s(\cdot)$ and pencil norm $\|\boldsymbol{P}\|_2$ are given by Definitions 3, 4 and 5, respectively. If η is sufficiently large, then the convex problem (6a)–(6d) *has a unique optimal solution $(\overset{*}{\boldsymbol{x}}, \overset{*}{\boldsymbol{X}})$ such that $\overset{*}{\boldsymbol{X}} = \overset{*}{\boldsymbol{x}}\overset{*}{\boldsymbol{x}}^\top$ and $\overset{*}{\boldsymbol{x}}$ is feasible for* (1a)–(1c).

The motivation behind Theorem 2 is to show that an infeasible initial point can be used to produce feasible points. In general, it is computationally hard to calculate the exact distance to $\mathcal{F}$ and to verify GLICQ as a consequence. However, local search methods can be used in practice to find a local solution for (7), resulting in upper bounds on the distance to $\mathcal{F}$. In Section 4, we use this simple technique to verify condition (12) for benchmark cases used in our computational study.

The rest of this section is devoted to proving Theorems 1 and 2. To this end, it is convenient to consider the following optimization problem:

$$\underset{\boldsymbol{x} \in \mathbb{R}^n}{\text{minimize}} \quad q_0(\boldsymbol{x}) + \eta \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2 \tag{13a}$$

$$\text{s.t.} \quad q_k(\boldsymbol{x}) \le 0, \qquad k \in \mathcal{I} \tag{13b}$$

$$q_k(\boldsymbol{x}) = 0, \qquad k \in \mathcal{E}. \tag{13c}$$

Observe that the problem (6a) – (6d) is a convex relaxation of (13a) – (13c) and this is the motivation behind its introduction.

Consider $\alpha > 0$ for which the inequality

$$|q_0(\boldsymbol{x})| \leq \alpha \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2 + \alpha, \tag{14}$$

is satisfied for every $\boldsymbol{x} \in \mathbb{R}^n$. If $\eta > \alpha$, then the objective function (13a) is lower bounded by $-\alpha$ and its optimal value is attainable within any closed and nonempty subset of $\mathbb{R}^n$.

To prove the existence of α, assume that

$$\alpha \geq \sigma_{\max}\left(\begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{b}_0 \\ \boldsymbol{b}_0^\top & -\hat{\boldsymbol{x}}^\top \boldsymbol{A}_0 \hat{\boldsymbol{x}} - 2\boldsymbol{b}_0^\top \hat{\boldsymbol{x}} \end{bmatrix}\right) \sigma_{\min}^{-1}\left(\begin{bmatrix} \boldsymbol{I}_n & -\hat{\boldsymbol{x}} \\ -\hat{\boldsymbol{x}}^\top & \frac{1}{2} + \hat{\boldsymbol{x}}^\top \hat{\boldsymbol{x}} \end{bmatrix}\right) \tag{15a}$$

$$\alpha \geq 2|\hat{\boldsymbol{x}}^\top \boldsymbol{A}_0 \hat{\boldsymbol{x}} + 2\boldsymbol{b}_0^\top \hat{\boldsymbol{x}} + c_0|. \tag{15b}$$

Then, we have

$$|q_0(\boldsymbol{x})| = \left| \begin{bmatrix} \boldsymbol{x}^\top & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{b}_0 \\ \boldsymbol{b}_0^\top & -\hat{\boldsymbol{x}}^\top \boldsymbol{A}_0 \hat{\boldsymbol{x}} - 2\boldsymbol{b}_0^\top \hat{\boldsymbol{x}} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix} + \hat{\boldsymbol{x}}^\top \boldsymbol{A}_0 \hat{\boldsymbol{x}} + 2\boldsymbol{b}_0^\top \hat{\boldsymbol{x}} + c_0 \right| \tag{16a}$$

$$\leq \alpha \begin{bmatrix} \boldsymbol{x}^\top & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_n & -\hat{\boldsymbol{x}} \\ -\hat{\boldsymbol{x}}^\top & \frac{1}{2} + \hat{\boldsymbol{x}}^\top \hat{\boldsymbol{x}} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix} + \frac{\alpha}{2} \tag{16b}$$

$$= \alpha \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2 + \alpha \tag{16c}$$

which concludes (14).

The next lemma shows that by increasing the penalty term η, the optimal solution $\mathring{\boldsymbol{x}}$ can get as close to the initial point $\hat{\boldsymbol{x}}$ as $d_{\mathcal{F}}(\hat{\boldsymbol{x}})$. This lemma will later be used to show that $\mathring{\boldsymbol{x}}$ can inherit the LICQ property from $\hat{\boldsymbol{x}}$.

Lemma 1 *Given an arbitrary $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ and $\varepsilon > 0$, for sufficiently large $\eta > 0$, every optimal solution $\mathring{\boldsymbol{x}}$ of the problem* (13a)–(13c) *satisfies*

$$0 \leq \|\mathring{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2 - d_{\mathcal{F}}(\hat{\boldsymbol{x}}) \leq \varepsilon. \tag{17}$$

Proof Consider an optimal solution $\mathring{\boldsymbol{x}}$. Due to Definition 3, the distance between $\hat{\boldsymbol{x}}$ and every member of $\mathcal{F}$ is not less than $d_{\mathcal{F}}(\hat{\boldsymbol{x}})$, which concludes the left side of (17). Let $\boldsymbol{x}_d$ be an arbitrary member of the set $\{\boldsymbol{x} \in \mathcal{F} \mid \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 = d_{\mathcal{F}}(\hat{\boldsymbol{x}})\}$. Due to the optimality of $\mathring{\boldsymbol{x}}$, we have

$$q_0(\mathring{\boldsymbol{x}}) + \eta \|\mathring{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2^2 \leq q_0(\boldsymbol{x}_d) + \eta \|\boldsymbol{x}_d - \hat{\boldsymbol{x}}\|_2^2. \tag{18}$$

According to the inequalities (18) and (14), one can write

$$(\eta - \alpha)\|\mathring{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2^2 - \alpha \leq (\eta + \alpha)\|\boldsymbol{x}_d - \hat{\boldsymbol{x}}\|_2^2 + \alpha \tag{19a}$$

$$\Rightarrow \|\mathring{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2^2 \leq \|\boldsymbol{x}_d - \hat{\boldsymbol{x}}\|_2^2 + \frac{2\alpha}{\eta - \alpha}(1 + \|\boldsymbol{x}_d - \hat{\boldsymbol{x}}\|_2^2) \tag{19b}$$

$$\Rightarrow \|\mathring{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2^2 \leq d_{\mathcal{F}}(\hat{\boldsymbol{x}})^2 + \frac{2\alpha}{\eta - \alpha}(1 + d_{\mathcal{F}}(\hat{\boldsymbol{x}})^2), \tag{19c}$$

which concludes the right side of (17), provided that $\eta \geq \alpha + 2\alpha(1 + d_{\mathcal{F}}(\hat{\boldsymbol{x}})^2)[\varepsilon^2 + 2\varepsilon d_{\mathcal{F}}(\hat{\boldsymbol{x}})]^{-1}$.

Lemma 2 *Assume that $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ satisfies the GLICQ condition for the problem* (13a)–(13c)*. Given an arbitrary $\varepsilon > 0$, for sufficiently large $\eta > 0$, every optimal solution $\mathring{\boldsymbol{x}}$ of the problem satisfies*

$$s(\hat{\boldsymbol{x}}) - s(\mathring{\boldsymbol{x}}) \leq 2 d_{\mathcal{F}}(\hat{\boldsymbol{x}}) \|\boldsymbol{P}\|_2 + \varepsilon. \tag{20}$$

Proof Let $\hat{\mathcal{B}}$ and $\mathring{\mathcal{B}}$ denote the sets of quasi-binding constraints for $\hat{\boldsymbol{x}}$ and binding constraints for $\mathring{\boldsymbol{x}}$, respectively (based on Definition 4). Due to Lemma 1, for every $k \in \mathcal{I} \setminus \hat{\mathcal{B}}$ and every arbitrary $\varepsilon_1 > 0$, we have

$$\begin{aligned} q_k(\mathring{\boldsymbol{x}}) - q_k(\hat{\boldsymbol{x}}) &= 2(\boldsymbol{A}_k\hat{\boldsymbol{x}} + \boldsymbol{b}_k)^\top (\mathring{\boldsymbol{x}} - \hat{\boldsymbol{x}}) + (\mathring{\boldsymbol{x}} - \hat{\boldsymbol{x}})^\top \boldsymbol{A}_k (\mathring{\boldsymbol{x}} - \hat{\boldsymbol{x}}) \\ &\leq \|\nabla q_k(\hat{\boldsymbol{x}})\|_2 \|\mathring{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2 + \|\boldsymbol{A}_k\|_2 \|\mathring{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2^2 \\ &\leq \|\nabla q_k(\hat{\boldsymbol{x}})\|_2 d_{\mathcal{F}}(\hat{\boldsymbol{x}}) + \|\boldsymbol{A}_k\|_2 d_{\mathcal{F}}(\hat{\boldsymbol{x}})^2 + \varepsilon_1 < -q_k(\hat{\boldsymbol{x}}), \end{aligned} \tag{21}$$

if η is sufficiently large, which yields $\mathring{\mathcal{B}} \subseteq \hat{\mathcal{B}}$. Let $\boldsymbol{\nu} \in \mathbb{R}^{|\hat{\mathcal{B}}|}$ be the left singular vector of $\mathcal{J}_{\hat{\mathcal{B}}}(\mathring{\boldsymbol{x}})$, corresponding to the smallest singular value. Hence

$$\begin{aligned} s(\mathring{\boldsymbol{x}}) = \sigma_{\min}\{\mathcal{J}_{\mathring{\mathcal{B}}}(\mathring{\boldsymbol{x}})\} \geq \sigma_{\min}\{\mathcal{J}_{\hat{\mathcal{B}}}(\mathring{\boldsymbol{x}})\} &= \|\mathcal{J}_{\hat{\mathcal{B}}}(\mathring{\boldsymbol{x}})^\top \boldsymbol{\nu}\|_2 &\text{(22a)} \\ &\geq \|\mathcal{J}_{\hat{\mathcal{B}}}(\hat{\boldsymbol{x}})^\top \boldsymbol{\nu}\|_2 - \|[\mathcal{J}_{\hat{\mathcal{B}}}(\hat{\boldsymbol{x}}) - \mathcal{J}_{\hat{\mathcal{B}}}(\mathring{\boldsymbol{x}})]^\top \boldsymbol{\nu}\|_2 &\text{(22b)} \\ &\geq \sigma_{\min}\{\mathcal{J}_{\hat{\mathcal{B}}}(\hat{\boldsymbol{x}})\} \|\boldsymbol{\nu}\|_2 - 2\|\boldsymbol{P}\|_2 \|\hat{\boldsymbol{x}} - \mathring{\boldsymbol{x}}\|_2 \|\boldsymbol{\nu}\|_2 &\text{(22c)} \\ &\geq s(\hat{\boldsymbol{x}}) - 2\|\boldsymbol{P}\|_2 \|\hat{\boldsymbol{x}} - \mathring{\boldsymbol{x}}\|_2 &\text{(22d)} \\ &\geq s(\hat{\boldsymbol{x}}) - 2 d_{\mathcal{F}}(\hat{\boldsymbol{x}}) \|\boldsymbol{P}\|_2 - \varepsilon, &\text{(22e)} \end{aligned}$$

if η is large, which concludes the inequality (20).

In light of Lemma 2, if $\hat{\boldsymbol{x}}$ is GLICQ regular and relatively close to $\mathcal{F}$, then $\mathring{\boldsymbol{x}}$ is LICQ regular as well. This is used next to prove the existence of Lagrange multipliers.

Lemma 3 *Let $\mathring{\boldsymbol{x}}$ be an optimal solution of the problem* (13a)–(13c)*, and assume that $\mathring{\boldsymbol{x}}$ is LICQ regular. There exists a pair of dual vectors $(\mathring{\boldsymbol{\gamma}}, \mathring{\boldsymbol{\mu}}) \in \mathbb{R}_+^{|\mathcal{I}|} \times \mathbb{R}^{|\mathcal{E}|}$ that satisfies the following Karush-Kuhn-Tucker (KKT) conditions:*

$$2(\eta \boldsymbol{I} + \boldsymbol{A}_0)(\mathring{\boldsymbol{x}} - \hat{\boldsymbol{x}}) + 2(\boldsymbol{A}_0\hat{\boldsymbol{x}} + \boldsymbol{b}_0) + \mathcal{J}(\mathring{\boldsymbol{x}})^\top [\mathring{\boldsymbol{\gamma}}^\top, \mathring{\boldsymbol{\mu}}^\top]^\top = 0, \tag{23a}$$

$$\mathring{\gamma}_k q_k(\mathring{\boldsymbol{x}}) = 0, \qquad \forall k \in \mathcal{I}. \tag{23b}$$

Proof Due to the LICQ condition, there exists a pair of dual vectors $(\mathring{\boldsymbol{\gamma}}, \mathring{\boldsymbol{\mu}}) \in \mathbb{R}_+^{|\mathcal{I}|} \times \mathbb{R}^{|\mathcal{E}|}$, which satisfies the KKT stationarity and complementary slackness conditions. Due to stationarity, we have

$$\begin{aligned} 0 &= \nabla_{\boldsymbol{x}} \mathcal{L}(\mathring{\boldsymbol{x}}, \mathring{\boldsymbol{\gamma}}, \mathring{\boldsymbol{\mu}})/2 \\ &= \eta(\mathring{\boldsymbol{x}} - \hat{\boldsymbol{x}}) + (\boldsymbol{A}_0\mathring{\boldsymbol{x}} + \boldsymbol{b}_0) + \boldsymbol{P}(\mathring{\boldsymbol{\gamma}}, \mathring{\boldsymbol{\mu}})\mathring{\boldsymbol{x}} + \sum_{k \in \mathcal{I}} \mathring{\gamma}_k \boldsymbol{b}_k + \sum_{k \in \mathcal{E}} \mathring{\mu}_k \boldsymbol{b}_k \\ &= (\eta \boldsymbol{I} + \boldsymbol{A}_0)(\mathring{\boldsymbol{x}} - \hat{\boldsymbol{x}}) + (\boldsymbol{A}_0\hat{\boldsymbol{x}} + \boldsymbol{b}_0) + \mathcal{J}(\mathring{\boldsymbol{x}})^\top [\mathring{\boldsymbol{\gamma}}^\top, \mathring{\boldsymbol{\mu}}^\top]^\top / 2. \end{aligned} \tag{24}$$

Moreover, (23b) is concluded from the complementary slackness.

The next lemma bounds the Lagrange multipliers whose existence is proven previously. This bound is helpful to prove that $\overset{*}{\boldsymbol{X}} = \overset{*}{\boldsymbol{x}}\overset{*}{\boldsymbol{x}}^\top$.

Lemma 4 *Consider an arbitrary* $\varepsilon > 0$ *and suppose* $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ *satisfies the inequality*

$$s(\hat{\boldsymbol{x}}) > 2d_{\mathcal{F}}(\hat{\boldsymbol{x}})\|\boldsymbol{P}\|_2. \tag{25}$$

If η *is sufficiently large, for every optimal solution* $\overset{*}{\boldsymbol{x}}$ *of the problem* (13a)–(13c)*, there exists a pair of dual vectors* $(\overset{*}{\boldsymbol{\gamma}}, \overset{*}{\boldsymbol{\mu}}) \in \mathbb{R}_+^{|\mathcal{I}|} \times \mathbb{R}^{|\mathcal{E}|}$ *that satisfies the inequality*

$$\frac{1}{\eta}\sqrt{\|\overset{*}{\boldsymbol{\gamma}}\|_2^2 + \|\overset{*}{\boldsymbol{\mu}}\|_2^2} \leq \frac{2d_{\mathcal{F}}(\hat{\boldsymbol{x}})}{s(\hat{\boldsymbol{x}}) - 2d_{\mathcal{F}}(\hat{\boldsymbol{x}})\|\boldsymbol{P}\|_2} + \varepsilon \tag{26}$$

as well as the equations (23a) *and* (23b).

Proof Due to Lemma 3, there exists $(\overset{*}{\boldsymbol{\gamma}}, \overset{*}{\boldsymbol{\mu}}) \in \mathbb{R}_+^{|\mathcal{I}|} \times \mathbb{R}^{|\mathcal{E}|}$ that satisfies the equations (23a) and (23b). Let $\boldsymbol{\tau} \triangleq [\overset{*}{\boldsymbol{\gamma}}^\top, \overset{*}{\boldsymbol{\mu}}^\top]^\top$ and let $\overset{*}{\mathcal{B}}$ be the set of binding constraints for $\overset{*}{\boldsymbol{x}}$. Due to equations (23a) and (23b), one can write

$$2(\eta\boldsymbol{I} + \boldsymbol{A}_0)(\overset{*}{\boldsymbol{x}} - \hat{\boldsymbol{x}}) + 2(\boldsymbol{A}_0\hat{\boldsymbol{x}} + \boldsymbol{b}_0) + \mathcal{J}_{\overset{*}{\mathcal{B}}}(\overset{*}{\boldsymbol{x}})^\top\boldsymbol{\tau}\{\overset{*}{\mathcal{B}}\} = 0. \tag{27}$$

Let $\phi \triangleq s(\hat{\boldsymbol{x}}) - 2d_{\mathcal{F}}(\hat{\boldsymbol{x}})\|\boldsymbol{P}\|_2$ and define

$$\varepsilon_1 \triangleq \phi \times \frac{\varepsilon - 2\eta^{-1}\phi^{-1}(\|\boldsymbol{A}_0\hat{\boldsymbol{x}}+\boldsymbol{b}_0\|_2 + d_{\mathcal{F}}(\hat{\boldsymbol{x}})\|\boldsymbol{A}_0\|_2)}{\varepsilon + 2 + 2\eta^{-1}\|\boldsymbol{A}_0\|_2 + 2\phi^{-1}d_{\mathcal{F}}(\hat{\boldsymbol{x}})}. \tag{28}$$

If η is sufficiently large, ε_1 is positive and based on Lemmas 1 and 2, we have

$$\begin{aligned}
\frac{\|\boldsymbol{\tau}\|_2}{\eta} = \frac{\|\boldsymbol{\tau}\{\overset{*}{\mathcal{B}}\}\|_2}{\eta} &\leq \frac{2\|(\eta\boldsymbol{I} + \boldsymbol{A}_0)(\overset{*}{\boldsymbol{x}} - \hat{\boldsymbol{x}}) + (\boldsymbol{A}_0\hat{\boldsymbol{x}} + \boldsymbol{b}_0)\|_2}{\eta\sigma_{\min}\{\mathcal{J}_{\overset{*}{\mathcal{B}}}(\overset{*}{\boldsymbol{x}})\}} \\
&\leq \frac{2\eta\|\overset{*}{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2 + 2\|\boldsymbol{A}_0\|_2\|\overset{*}{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2 + 2\|\boldsymbol{A}_0\hat{\boldsymbol{x}} + \boldsymbol{b}_0\|_2}{\eta s(\overset{*}{\boldsymbol{x}})} \\
&\leq \frac{2(d_{\mathcal{F}}(\hat{\boldsymbol{x}})+\varepsilon_1)+2\eta^{-1}[\|\boldsymbol{A}_0\|_2(d_{\mathcal{F}}(\hat{\boldsymbol{x}})+\varepsilon_1)+\|\boldsymbol{A}_0\hat{\boldsymbol{x}}+\boldsymbol{b}_0\|_2]}{s(\hat{\boldsymbol{x}}) - 2d_{\mathcal{F}}(\hat{\boldsymbol{x}})\|\boldsymbol{P}\|_2 - \varepsilon_1} \\
&= \frac{2d_{\mathcal{F}}(\hat{\boldsymbol{x}})}{s(\hat{\boldsymbol{x}}) - 2d_{\mathcal{F}}(\hat{\boldsymbol{x}})\|\boldsymbol{P}\|_2} + \varepsilon,
\end{aligned} \tag{29}$$

where the last equality is a result of the equation (28).

The next two lemmas provide sufficient conditions for $\overset{*}{\boldsymbol{X}} = \overset{*}{\boldsymbol{x}}\overset{*}{\boldsymbol{x}}^\top$ with respect to the Lagrange multipliers that will be used later to prove Theorems 1 and 2.

Lemma 5 *Consider an optimal solution* $\overset{*}{\boldsymbol{x}}$ *of the problem* (13a)–(13c)*, and a pair of dual vectors* $(\overset{*}{\boldsymbol{\gamma}}, \overset{*}{\boldsymbol{\mu}}) \in \mathbb{R}_+^{|\mathcal{I}|} \times \mathbb{R}^{|\mathcal{E}|}$ *that satisfies the conditions* (23a) *and* (23b)*. If the matrix inequality*

$$\eta\boldsymbol{I} + \boldsymbol{A}_0 + \boldsymbol{P}(\overset{*}{\boldsymbol{\gamma}}, \overset{*}{\boldsymbol{\mu}}) \succ_{\mathcal{D}_r} 0, \tag{30}$$

holds true, then the pair $(\overset{*}{\boldsymbol{x}}, \overset{*}{\boldsymbol{x}}\overset{*}{\boldsymbol{x}}^\top)$ *is the unique primal solution to the penalized SDP* (6a)–(6d).

Proof Let $\boldsymbol{\Lambda} \in \mathbb{S}_n^+$ denotes the dual variable associated with the conic constraint (6d). Then, the KKT conditions for the problem (6a)-(6d) can be written as follows:

$$\nabla_{\boldsymbol{x}}\ \bar{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{X}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = 2\left(\boldsymbol{\Lambda x} - \eta\hat{\boldsymbol{x}} + \boldsymbol{b}_0 + \sum_{k\in\mathcal{I}} \mathring{\gamma}_k \boldsymbol{b}_k + \sum_{k\in\mathcal{E}} \mathring{\mu}_k \boldsymbol{b}_k\right) = 0, \tag{31a}$$

$$\nabla_{\boldsymbol{X}}\ \bar{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{X}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \eta\boldsymbol{I} + \boldsymbol{A}_0 + \boldsymbol{P}(\boldsymbol{\gamma}, \boldsymbol{\mu}) - \boldsymbol{\Lambda} = 0, \tag{31b}$$

$$\gamma_k q_k(\boldsymbol{x}) = 0, \qquad \forall k \in \mathcal{I} \tag{31c}$$

$$\langle \boldsymbol{\Lambda},\, \boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{X}\rangle = 0, \tag{31d}$$

where $\bar{\mathcal{L}} : \mathbb{R}^n \times \mathbb{S}_n \times \mathbb{R}^{|\mathcal{I}|} \times \mathbb{R}^{|\mathcal{E}|} \times \mathbb{S}_n \to \mathbb{R}$ is the Lagrangian function, equations (31a) and (31b) account for stationarity with respect to $\boldsymbol{x}$ and $\boldsymbol{X}$, respectively, and equations (31c) and (31d) are the complementary slackness conditions for the constraints (6b) and (6d), respectively. Define

$$\mathring{\boldsymbol{\Lambda}} \triangleq \eta\boldsymbol{I} + \boldsymbol{A}_0 + \boldsymbol{P}(\mathring{\boldsymbol{\gamma}}, \mathring{\boldsymbol{\mu}}). \tag{32}$$

Due to Lemma (3), if η is sufficiently large, $\mathring{\boldsymbol{x}}$ and $(\mathring{\boldsymbol{\gamma}}, \mathring{\boldsymbol{\mu}})$ satisfy the equations (23a) and (23b), which yield the optimality conditions (31a)-(31d), if $\boldsymbol{x} = \mathring{\boldsymbol{x}}$, $\boldsymbol{X} = \mathring{\boldsymbol{x}}\mathring{\boldsymbol{x}}^\top$, $\boldsymbol{\gamma} = \mathring{\boldsymbol{\gamma}}$, $\boldsymbol{\mu} = \mathring{\boldsymbol{\mu}}$, and $\boldsymbol{\Lambda} = \mathring{\boldsymbol{\Lambda}}$. Therefore, the pair $(\mathring{\boldsymbol{x}}, \mathring{\boldsymbol{x}}\mathring{\boldsymbol{x}}^\top)$ is a primal optimal points for the penalized SDP (6a)-(6d).

Since the KKT conditions hold for every pair of primal and dual solutions, we have

$$\mathring{\boldsymbol{x}} = \mathring{\boldsymbol{\Lambda}}^{-1}\left(\eta\hat{\boldsymbol{x}} - \boldsymbol{b}_0 - \sum_{k\in\mathcal{I}} \mathring{\gamma}_k \boldsymbol{b}_k - \sum_{k\in\mathcal{E}} \mathring{\mu}_k \boldsymbol{b}_k\right) \tag{33}$$

and $\mathring{\boldsymbol{X}} = \mathring{\boldsymbol{x}}\mathring{\boldsymbol{x}}^\top$, according to the equations (31a) and (31d), respectively, which implies the uniqueness of the solution.

Lemma 6 *Consider an optimal solution $\mathring{\boldsymbol{x}}$ of the problem* (13a)-(13c)*, and a pair of dual vectors $(\mathring{\boldsymbol{\gamma}}, \mathring{\boldsymbol{\mu}}) \in \mathbb{R}_+^{|\mathcal{I}|} \times \mathbb{R}^{|\mathcal{E}|}$ that satisfies the conditions* (23a) *and* (23b)*. If the inequality,*

$$\frac{1}{\eta}\sqrt{\|\mathring{\boldsymbol{\gamma}}\|_2^2 + \|\mathring{\boldsymbol{\mu}}\|_2^2} < \frac{1}{C_{n-1,r-1}\|\boldsymbol{P}\|_2} - \frac{\|\boldsymbol{A}_0\|_2}{\eta\|\boldsymbol{P}\|_2} \tag{34}$$

holds true, then the pair $(\mathring{\boldsymbol{x}}, \mathring{\boldsymbol{x}}\mathring{\boldsymbol{x}}^\top)$ is the unique primal solution to the penalized SDP (6a)–(6d).

Proof Based on Lemma 5, it suffices to prove the conic inequality (30). Define

$$\boldsymbol{K} \triangleq \boldsymbol{A}_0 + \boldsymbol{P}(\mathring{\boldsymbol{\gamma}}, \mathring{\boldsymbol{\mu}}). \tag{35}$$

It follows that

$$\|\boldsymbol{K}\|_2 \leq \|\boldsymbol{A}_0\|_2 + \sum_{k\in\mathcal{I}} \mathring{\gamma}_k \|\boldsymbol{A}_k\|_2 + \sum_{k\in\mathcal{E}} \mathring{\mu}_k \|\boldsymbol{A}_k\|_2, \tag{36a}$$

$$\leq \|\boldsymbol{A}_0\|_2 + \|\boldsymbol{P}\|_2\sqrt{\|\mathring{\boldsymbol{\gamma}}\|_2^2 + \|\mathring{\boldsymbol{\mu}}\|_2^2}\,. \tag{36b}$$

Let $\mathcal{R}$ be the set of all r-member subsets of $\{1, 2, \ldots, n\}$. Hence,

$$\eta \boldsymbol{I} + \boldsymbol{K} = \sum_{\mathcal{K} \in \mathcal{R}} \boldsymbol{I}\{\mathcal{K}\}^\top \boldsymbol{R}_\mathcal{K} \, \boldsymbol{I}\{\mathcal{K}\}, \tag{37}$$

where

$$\boldsymbol{R}_\mathcal{K} = \binom{n-1}{r-1}^{-1} [\eta \boldsymbol{I}\{\mathcal{K}, \mathcal{K}\} + \boldsymbol{K}\{\mathcal{K}, \mathcal{K}\}]. \tag{38}$$

Due to the inequalities (34) and (36), we have $\boldsymbol{R}_\mathcal{K} \succ 0$ for every $\mathcal{K} \in \mathcal{R}$, which proves that $\eta \boldsymbol{I} + \boldsymbol{K} \succ_{\mathcal{D}_r} 0$.

Proof (Theorem 2) Let $\overset{*}{\boldsymbol{x}}$ be an optimal solution of the problem (13a)–(13c). According to the assumption (12), the inequality (25) holds true, and due to Lemma 4, if η is sufficiently large, there exists a corresponding pair of dual vectors $(\overset{*}{\boldsymbol{\gamma}}, \overset{*}{\boldsymbol{\mu}})$ that satisfies the inequality (26). Now, according to the inequality (12), we have

$$\frac{2 d_\mathcal{F}(\hat{\boldsymbol{x}})}{s(\hat{\boldsymbol{x}}) - 2 d_\mathcal{F}(\hat{\boldsymbol{x}}) \|\boldsymbol{P}\|_2} \leq \frac{1}{C_{n-1,r-1} \|\boldsymbol{P}\|_2} \tag{39}$$

and therefore (26) concludes (34). Hence, according to Lemma 6, the pair $(\overset{*}{\boldsymbol{x}}, \overset{*}{\boldsymbol{x}}\overset{*}{\boldsymbol{x}}^\top)$ is the unique primal solution to the penalized SDP (6a)–(6d).

Proof (Theorem 1) If $\hat{\boldsymbol{x}}$ is feasible, then $d_\mathcal{F}(\hat{\boldsymbol{x}}) = 0$. Therefore, the tightness of the penalization for Theorem 1 is a direct consequence of Theorem 2. Denote the unique optimal solution of the penalized SDP as $(\overset{*}{\boldsymbol{x}}, \overset{*}{\boldsymbol{x}}\overset{*}{\boldsymbol{x}}^\top)$. Then it is straightforward to verify the inequality $q_0(\overset{*}{\boldsymbol{x}}) \leq q_0(\hat{\boldsymbol{x}})$ by evaluating the objective function (6a) at the point $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^\top)$.

3.2 Sequential penalization procedure

In practice, the penalized SDP (6a)–(6d) can be initialized by a point that may not satisfy the conditions of Theorem 1 or Theorem 2 as these conditions are only sufficient, but not necessary. If the chosen initial point $\hat{\boldsymbol{x}}$ does not result in a tight penalization, the penalized SDP(6a)–(6d) can be solved sequentially by updating the initial point until a feasible and near-optimal point is obtained. This heuristic procedure is described in Algorithm 1.

Algorithm 1 Sequential Penalized SDP.

initiate $\{q_k\}_{k \in \{0\} \cup \mathcal{I} \cup \mathcal{E}}$, $r \geq 2$, $\hat{\boldsymbol{x}} \in \mathbb{R}^n$, and the fixed parameter $\eta > 0$
while stopping criterion is not met **do**
  solve the problem (6a)–(6d) with the initial point $\hat{\boldsymbol{x}}$ to obtain $(\overset{*}{\boldsymbol{x}}, \overset{*}{\boldsymbol{X}})$
  set $\hat{\boldsymbol{x}} \leftarrow \overset{*}{\boldsymbol{x}}$
end while
return $\overset{*}{\boldsymbol{x}}$

According to Theorem (2), once $\hat{\boldsymbol{x}}$ is close enough to the feasible set $\mathcal{F}$, the penalization becomes tight, i.e., a feasible solution $\mathring{\boldsymbol{x}}$ is recovered as the unique optima solution to (6a)–(6d). Afterwards, in the subsequent iterations, according to Theorem (1), feasibility is preserved and the objective value does not increase.

The following example illustrates Algorithm 1 for polynomial optimization.

Example 1 Consider the following three-dimensional polynomial optimization:

$$\underset{a,b,c\in\mathbb{R}}{\text{minimize}} \quad a \tag{40a}$$

$$\text{s.t.} \quad a^5 - b^4 - c^4 + 2a^3 + 2a^2b - 2ab^2 + 6abc - 2 = 0 \tag{40b}$$

To derive a QCQP reformulation of the problem (40a)–(40b), we consider a variable $x \in \mathbb{R}^8$, whose elements account for the monomials a, b, c, a^2, b^2, c^2, ab, and a^3, respectively. This leads to the following QCQP:

$$\underset{\boldsymbol{x}\in\mathbb{R}^8}{\text{minimize}} \quad x_1 \tag{41a}$$

$$\text{s.t.} \quad x_4x_8 - x_5^2 - x_6^2 + 2x_1x_4 + 2x_2x_4 - 2x_1x_5 + 6x_3x_7 - 2 = 0, \tag{41b}$$

$$x_4 - x_1^2 = 0, \;\; x_5 - x_2^2 = 0, \;\; x_6 - x_3^3 = 0, \tag{41c}$$

$$x_7 - x_1x_2 = 0, \;\; x_8 - x_1x_4 = 0. \tag{41d}$$

The transformation of the polynomial optimization to QCQP is standard and it is described in Appendix A for completeness. The global optimal objective value of the above QCQP equals -2.0198 and the lower-bound, offered by the standard SDP relaxation equals -89.8901. In order to solve the above QCQP, we run Algorithm 1, equipped with the SDP relaxation (no additional valid inequalities) and penalty term $\eta = 0.025$. The trajectory with three different initializations $\hat{\boldsymbol{x}}^1 = [0,0,0,0,0,0,0]^\top$, $\hat{\boldsymbol{x}}^2 = [-3,0,2,9,0,4,0,27]^\top$, and $\hat{\boldsymbol{x}}^3 = [0,4,0,0,16,0,0,0]^\top$ are given in Table 1 and shown in Fig. 1. In all three cases, the algorithm achieves feasibility in 1–8 iterations. Moreover, a feasible solution with less than 0.2% gap from global optimality is attained within 10 iterations in all three cases. The example illustrates that Algorithm 1 is not sensitive to the choice of initial point.

Table 1: Trajectory of Algorithm 1 for three different initializations.

Iteration	$\mathring{\boldsymbol{x}}^1$				$\mathring{\boldsymbol{x}}^2$				$\mathring{\boldsymbol{x}}^3$			
	a (obj.)	b	c	$\mathrm{tr}\{\mathring{\boldsymbol{X}} - \mathring{\boldsymbol{x}}\mathring{\boldsymbol{x}}^\top\}$	a (obj.)	b	c	$\mathrm{tr}\{\mathring{\boldsymbol{X}} - \mathring{\boldsymbol{x}}\mathring{\boldsymbol{x}}^\top\}$	a (obj.)	b	c	$\mathrm{tr}\{\mathring{\boldsymbol{X}} - \mathring{\boldsymbol{x}}\mathring{\boldsymbol{x}}^\top\}$
0	0.0000	0.0000	0.0000	-	-3.0000	0.0000	2.0000	-	0.0000	4.0000	0.0000	-
1	-1.2739	0.6601	-0.4697	2.1884	-2.5377	1.2831	-0.7380	138.9796	-1.5721	2.6848	-0.9492	39.2455
2	-1.5173	1.1445	-1.0128	$< 10^{-11}$	-2.4389	2.0715	-1.3946	51.1170	-1.5749	2.7588	-1.3854	13.5140
3	-1.6882	1.3773	-1.2015	$< 10^{-11}$	-2.2889	2.2685	-1.7098	23.0050	-1.6678	2.6583	-1.5228	0.9995
4	-1.8021	1.5739	-1.3561	$< 10^{-11}$	-2.1878	2.3416	-1.8442	11.4963	-1.8322	2.6083	-1.5587	$< 10^{-11}$
5	-1.8824	1.7447	-1.4873	$< 10^{-11}$	-2.1194	2.3621	-1.9007	5.9206	-1.9460	2.5261	-1.6624	$< 10^{-11}$
6	-1.9386	1.8930	-1.5992	$< 10^{-11}$	-2.0733	2.3611	-1.9250	2.9082	-2.0002	2.4391	-1.7847	$< 10^{-11}$
7	-1.9760	2.0180	-1.6923	$< 10^{-11}$	-2.0423	2.3526	-1.9352	1.1594	-2.0156	2.3824	-1.8598	$< 10^{-11}$
8	-1.9985	2.1175	-1.7656	$< 10^{-11}$	-2.0214	2.3426	-1.9393	0.0938	-2.0189	2.3532	-1.8938	$< 10^{-11}$
9	-2.0104	2.1907	-1.8193	$< 10^{-11}$	-2.0197	2.3352	-1.9302	$< 10^{-11}$	-2.0196	2.3387	-1.9079	$< 10^{-11}$
10	-2.0160	2.2408	-1.8559	$< 10^{-11}$	-2.0198	2.3304	-1.9240	$< 10^{-11}$	-2.0197	2.3313	-1.9135	$< 10^{-11}$

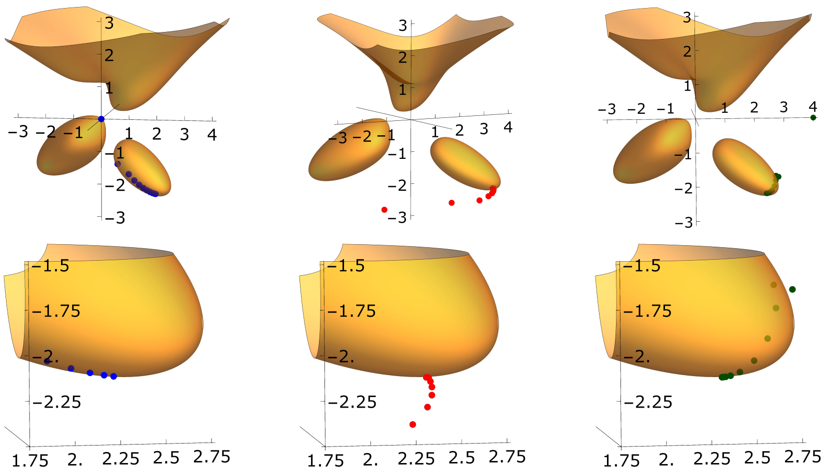

Fig. 1: Trajectory of Algorithm 1 for three different initializations. The yellow surface represents the feasible set and the blue, red and green points correspond to $\hat{x}^1$, $\hat{x}^2$ and $\hat{x}^3$, respectively.

4 Numerical experiments

In this section we describe numerical experiments to test the effectiveness of the sequential penalization method for non-convex QCQPs from the library of quadratic programming instances (QPLIB) [21] as well as large-scale system identification problems [17].

4.1 QPLIB problems

The experiments are performed on a desktop computer with a 12-core 3.0GHz CPU and 256GB RAM. MOSEK v8.1 [3] is used through MATLAB 2017a to solve the resulting SDPs. The size and number of constraints for each QPLIB instance are reported in Table 2.

Although computing the exact distance to the feasible set is difficult, one can find an upper bound for $d_{\mathcal{F}}$ via simple local search to verify the sufficient condition of Theorem 2. Using MATLAB's fmincon solver, we verified that the (infeasible) SDP relaxation solution for 10 of the 24 instances used in this study satisfied the sufficient condition of Theorem 2. Hence, for these 10 instances (highligted with stars in Table 5), Theorem 2 guarantees recovering a feasible solution via penalized SDP in a single iteration.

For example, consider QPLIB instance 1773 and its SDP relaxation solution $\boldsymbol{x}^{\mathrm{SDP}}$. Using MATLAB's fmincon solver, we verify that

$$d_{\mathcal{F}}(\boldsymbol{x}^{\mathrm{SDP}}) < 0.0327$$

which leads to 7 quasi-binding constraints (including lower and upper-bounds) and

$$s(\boldsymbol{x}^{\mathrm{SDP}}) > 3.3712.$$

Moreover, we have $\|P\| = 1.7626$, which implies that (12) is satisfied by $\boldsymbol{x}^{\mathrm{SDP}}$.

4.1.1 Sequential penalization

Tables 3–6 report results with Algorithm 1 for 2×2 SDP, 2×2 SDP+RLT, SDP, and SDP+RLT relaxations, respectively. The following valid inequalities are imposed on all of the convex relaxations:

$$X_{kk} - (x_k^{\mathrm{lb}} + x_k^{\mathrm{ub}})x_k + x_k^{\mathrm{lb}} x_k^{\mathrm{ub}} \leq 0, \qquad \forall k \in \{1,\dots,n\} \tag{42a}$$

$$X_{kk} - (x_k^{\mathrm{ub}} + x_k^{\mathrm{ub}})x_k + x_k^{\mathrm{ub}} x_k^{\mathrm{ub}} \geq 0, \qquad \forall k \in \{1,\dots,n\} \tag{42b}$$

$$X_{kk} - (x_k^{\mathrm{lb}} + x_k^{\mathrm{lb}})x_k + x_k^{\mathrm{lb}} x_k^{\mathrm{lb}} \geq 0, \qquad \forall k \in \{1,\dots,n\} \tag{42c}$$

where $\boldsymbol{l}, \boldsymbol{u} \in \mathbb{R}^n$ are given lower and upper bounds on $\boldsymbol{x}$. Problem (4a)–(4d) is solved with the following four settings:

- 2×2 *SDP relaxation:* $r = 2$ and valid inequalities (42a) – (42c).
- 2×2 *SDP+RLT relaxation:* $\mathcal{V} = \mathcal{H} \times \mathcal{H}$ and $r = 2$.
- *SDP relaxation:* $r = n$ and valid inequalities (42a) – (42c).
- *SDP+RLT relaxation:* $\mathcal{V} = \mathcal{H} \times \mathcal{H}$ and $r = n$,

where $\mathcal{V}$ is defined in Appendix B. Let $(\overset{*}{\boldsymbol{x}}, \overset{*}{\boldsymbol{X}})$ denote the optimal solution of the convex relaxation (4a)–(4d). We use the point $\hat{\boldsymbol{x}} = \overset{*}{\boldsymbol{x}}$ as the initial point of the algorithm.

The penalty parameter η is chosen via bisection as the smallest number of the form $\alpha \times 10^{\beta}$, which results in a tight penalization during the first six iterations, where $\alpha \in \{1, 2, 5\}$ and β is an integer. In all of the experiments, the value of η has remained static throughout Algorithm 1. Denote the sequence of penalized SDP solutions obtained by Algorithm 1 as

$$(\boldsymbol{x}^{(1)}, \boldsymbol{X}^{(1)}), \;\; (\boldsymbol{x}^{(2)}, \boldsymbol{X}^{(2)}), \;\; (\boldsymbol{x}^{(3)}, \boldsymbol{X}^{(3)}), \;\; \dots$$

The smallest i such that

$$\mathrm{tr}\{\boldsymbol{X}^{(i)} - \boldsymbol{x}^{(i)}(\boldsymbol{x}^{(i)})^{\top}\} < 10^{-7} \tag{43}$$

is denoted by i^{feas}, i.e., it is the number of iterations that Algorithm 1 needs to attain a tight penalization. Moreover, the smallest i such that

$$\frac{q_0(\boldsymbol{x}^{(i-1)}) - q_0(\boldsymbol{x}^{(i)})}{|q_0(\boldsymbol{x}^{(i)})|} \leq 5 \times 10^{-4} \tag{44}$$

is denoted by i^{stop}, and $\mathrm{UB} \triangleq q_0(\boldsymbol{x}^{(i^{\mathrm{stop}})})$. The following formula is used to calculate the percentage gaps from the optimal costs reported by the QPLIB library:

$$\mathrm{GAP}(\%) = 100 \times \frac{q_0^{\mathrm{stop}} - q_0^{\mathrm{opt}}}{|q_0^{\mathrm{opt}}|}. \tag{45}$$

Moreover, t(s) denotes the cumulative solver time in seconds for i^{stop} iterations. Our results are compared with the global solvers Baron [59] and Couenne [8] by setting the maximum solver time equal to the time spent by Algorithm 1. We ran Baron and

Table 2: QPLIB benchmark problems.

Inst	Total Var	Quad Cons	Total Cons	Inst	Total Var	Quad Cons	Total Cons	Inst	Total Var	Quad Cons	Total Cons	Inst	Total Var	Quad Cons	Total Cons
0343	50	0	1	1353	50	1	6	1535	60	60	66	1773	60	1	7
0911	50	50	50	1423	40	20	24	1619	50	25	30	1886	50	50	50
0975	50	10	10	1437	50	1	11	1661	60	1	13	1913	48	48	48
1055	40	20	20	1451	60	60	66	1675	60	1	13	1922	30	60	60
1143	40	20	24	1493	40	1	5	1703	60	30	36	1931	40	40	40
1157	40	1	9	1507	30	30	33	1745	50	50	55	1967	50	75	75

Table 3: Sequential penalized 2×2 SDP.

Inst	**Sequential penalized 2×2 SDP**						**BARON**			**COUENNE**		
	η	i^{feas}	i^{stop}	t(s)	UB	GAP(%)	LB	UB	GAP(%)	LB	UB	GAP(%)
0343†	5e+2	1	100	75.27	-5.882	7.89	-95.372	-6.386	0.00	-7668.005	-6.386	0.00
0911	1e+1	1	29	22.91	-30.675	4.58	-172.777	0.000	100	-172.777	-31.026	3.49
0975	5e+0	6	18	46.36	-36.434	3.75	-47.428	-37.801	0.14	-171.113	-37.213	1.69
1055	1e+1	1	22	14.39	-32.620	1.26	-37.841	-33.037	0.00	-199.457	-33.037	0.00
1143	2e+1	1	44	25.68	-55.417	3.20	-69.522	-57.247	0.00	-384.45	-56.237	1.76
1157	2e+0	2	9	9.01	-10.938	0.10	-11.414	-10.948	0.00	-80.51	-10.948	0.00
1353	5e+0	1	48	84.90	-7.700	0.19	-7.925	-7.714	0.00	-73.28	-7.714	0.00
1423	5e+0	1	29	17.44	-14.684	1.90	-16.313	-14.968	0.00	-76.13	-14.871	0.65
1437	5e+0	1	36	54.57	-7.785	0.06	-9.601	-7.789	0.00	-87.58	-7.789	0.00
1451	2e+1	4	21	20.86	-85.598	2.26	-135.140	-87.577	0.00	-468.04	-86.860	0.82
1493	2e+1	1	18	14.49	-41.910	2.90	-47.239	-43.160	0.00	-395.69	-43.160	0.00
1507	2e+0	1	15	8.98	-8.289	0.15	-49.709	-8.301	0.00	-44.37	-8.301	0.00
1535	5e+0	1	26	28.16	-10.948	5.51	-13.407	-11.397	1.63	-107.86	-11.398	1.63
1619	5e+0	1	39	32.34	-9.210	0.08	-10.302	-9.217	0.00	-74.55	-9.217	0.00
1661	5e+0	1	32	87.50	-15.666	1.81	-19.667	-15.955	0.00	-139.25	-15.955	0.00
1675	2e+1	1	21	36.38	-75.485	0.24	-96.864	-75.669	0.00	-435.48	-75.669	0.00
1703	5e+1	2	30	31.82	-130.902	1.43	-180.935	-132.802	0.00	-929.92	-132.802	0.00
1745	2e+1	1	26	22.15	-71.704	0.93	-77.465	-72.377	0.00	-317.99	-72.377	0.00
1773	5e+0	1	56	148.79	-14.154	3.34	-21.581	-14.642	0.00	-118.65	-14.642	0.00
1886	2e+1	1	34	26.82	-78.604	0.09	-135.615	-78.672	0.00	-324.87	-78.672	0.00
1913	1e+1	1	28	21.91	-51.889	0.42	-68.555	-52.109	0.00	-164.26	-51.478	1.21
1922	1e+1	1	23	11.16	-35.437	1.43	-121.872	-35.951	0.00	-123.2	-35.951	0.00
1931	1e+1	1	13	8.78	-53.684	3.64	-85.196	-55.709	0.00	-204.08	-54.290	2.55
1967	5e+1	1	32	27.23	-105.570	1.87	-136.098	0.000	100	-622.57	-107.581	0.00
Avg	33.9	1.4	31.2	36.68		2.04			8.41			0.58
Max	500	6	100	148.79		7.89			100			3.34

Couenne through GAMS v25.1.2 [22]. The resulting lower bounds, upper bounds and percentage GAPs are reported in Tables 3–6.

We observe in Tables 3–6 that penalized 2×2 SDP, 2×2 SDP+RLT, SDP, and SDP+RLT have all successfully recover feasible points with an average of at most 2.07% optimality gap at termination. Penalized SDP, and SDP+RLT consistently achieve less than 3% optimality gap for all cases. Sequential SDP requires fewer of iterations compared to sequential 2×2 SDP to meet the stopping criterion (44) and results in overall faster solution times even though each iteration takes longer. For all formulations, a feasible point is obtained on average in about two iterations of Algorithm 1. With penalized 2×2 SDP (+RLT), at termination, the average optimality gap is lower than with Baron, but higher than with Couenne. On the other hand, with the stronger penalized SDP (+RLT) approach the average as well as the maximum optimality gap is somewhat lower than with both Baron and Couenne at

Table 4: Sequential penalized 2×2 SDP+RLT.

Inst	Sequential 2×2 SDP+RLT						BARON			COUENNE		
	η	i^{feas}	i^{stop}	t(s)	UB	GAP(%)	LB	UB	GAP(%)	LB	UB	GAP(%)
0343	1e+2	4	24	25.23	-5.945	6.91	-95.372	-6.386	0.00	-7668.005	-6.386	0.00
0911	1e+1	1	33	27.69	-30.923	3.81	-172.777	-32.148	0.00	-172.777	-31.026	3.49
0975	5e+0	6	15	4.10	-36.300	13.17	-47.428	-37.794	0.16	-171.113	-36.812	2.75
1055	1e+1	1	24	16.78	-32.666	1.12	-37.841	-33.037	0.00	-199.457	-33.037	0.00
1143	2e+1	1	30	32.66	-55.507	3.04	-69.522	-57.247	0.00	-384.45	-56.237	1.76
1157	2e+0	1	0	1.14	-10.948	0.00	-11.414	-10.948	0.00	-80.51	-10.948	0.00
1353	1e+0	3	11	19.41	-7.711	0.05	-7.925	-7.714	0.00	-73.28	-7.714	0.00
1423	2e+0	3	14	16.41	-14.730	1.59	-16.313	-14.968	0.00	-76.13	-14.871	0.65
1437	5e-1	4	8	21.62	-7.788	0.02	-9.601	-7.789	0.00	-87.58	-7.789	0.00
1451	2e+1	2	36	100.50	-87.502	0.09	-135.140	-87.577	0.00	-468.04	-87.283	0.34
1493	1e+1	3	13	13.69	-41.804	3.14	-47.239	-43.160	0.00	-395.69	-43.160	0.00
1507	1e+0	6	13	10.31	-8.295	0.08	-49.709	-8.301	0.00	-44.37	-8.301	0.00
1535	2e+0	3	23	83.47	-11.241	2.98	-13.407	-11.586	0.00	-107.86	-11.398	1.62
1619	2e+0	3	20	35.62	-9.213	0.05	-10.302	-9.217	0.00	-74.55	-9.217	0.00
1661	1e+0	3	8	35.85	-15.666	1.81	-19.667	-15.955	0.00	-139.25	-15.955	0.00
1675	1e+1	3	11	41.30	-75.537	0.17	-96.864	-75.669	0.00	-435.48	-75.669	0.00
1703	2e+1	5	22	62.63	-131.330	1.11	-180.935	-132.802	0.00	-929.92	-132.802	0.00
1745	5e+0	4	19	40.44	-72.351	0.04	-77.465	-72.377	0.00	-317.99	-72.377	0.00
1773	5e+0	1	56	120.65	-14.176	3.19	-21.581	-14.642	0.00	-118.65	-14.642	0.00
1886	2e+1	1	35	28.19	-78.620	0.07	-135.615	-78.672	0.00	-324.87	-78.672	0.00
1913	5e+0	4	18	15.10	-51.879	0.44	-68.555	-52.109	0.00	-164.26	-51.348	1.46
1922	1e+1	1	26	13.22	-35.451	1.39	-121.872	-35.951	0.00	-123.2	-35.951	0.00
1931	1e+1	1	13	8.59	-53.709	3.59	-85.196	-55.709	0.00	-204.08	-54.290	2.55
1967	5e+1	1	38	33.01	-105.616	1.83	-136.098	0.000	100	-622.57	-107.581	0.00
Avg	13.4	2.7	21.3	33.65		2.07			4.17			0.61
Max	100	6	56	120.65		13.17			100			3.49

termination. Overall, the proposed sequential penalization approach is successful in recovering high quality feasible solutions fast and its performance is comparable with the nonconvex optimizers Baron and Couenne for this data set.

Figures 2, shows the convergence of Algorithm 1 for instance 1507. The choice of η for all curves are taken from the corresponding rows of the Tables 3, 4, 5, and 6.

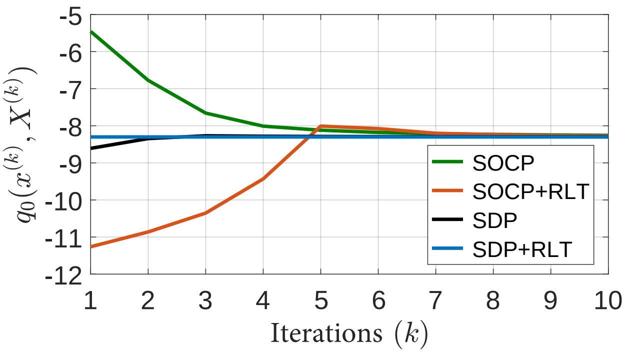

Fig. 2: Convergence of sequential 2×2 SDP, 2×2 SDP+RLT, SDP, and SDP+RLT for inst. 1507.

Table 5: Sequential penalized SDP.

Inst	Sequential SDP						BARON			COUENNE		
	η	i^{feas}	i^{stop}	t(s)	UB	GAP(%)	LB	UB	GAP(%)	LB	UB	GAP(%)
0343*	1e+2	1	53	29.24	-6.379	0.12	-95.372	-6.386	0.00	-7668.005	-6.386	0.00
0911	2e+0	1	9	5.19	-31.811	1.05	-172.777	0.000	100	-172.777	-31.026	3.49
0975	2e+0	2	13	8.18	-37.845	0.02	-47.428	-37.794	0.16	-171.113	-36.812	2.75
1055	5e+0	1	8	4.36	-32.528	1.54	-37.841	-33.037	0.00	-199.457	-33.037	0.00
1143	5e+0	4	15	7.89	-55.606	2.87	-69.522	-57.247	0.00	-384.45	-53.367	6.78
1157	1e+0	1	5	3.15	-10.945	0.03	-11.414	-10.948	0.00	-80.51	-10.948	0.00
1353*	1e+0	1	10	6.12	-7.712	0.03	-7.925	-7.714	0.00	-73.28	-7.714	0.00
1423*	1e+0	1	5	3.28	-14.676	1.95	-16.313	-14.968	0.00	-76.13	-14.078	5.94
1437*	1e+0	1	7	4.30	-7.787	0.03	-9.601	-7.789	0.00	-87.58	-7.789	0.00
1451†	5e+0	2	6	5.09	-85.972	1.83	-135.140	-	-	-468.04	-	-
1493*	5e+0	1	6	4.10	-43.160	0.00	-47.239	-43.160	0.00	-395.69	-43.160	0.00
1507	5e-1	3	6	3.28	-8.291	0.12	-49.709	-8.301	0.00	-44.37	-8.301	0.00
1535	1e+0	1	16	13.05	-11.363	1.93	-13.407	-11.397	1.63	-107.86	-11.398	1.63
1619*	1e+0	1	7	4.64	-9.213	0.05	-10.302	-9.217	0.00	-74.55	-9.217	0.00
1661*	1e+0	1	12	7.57	-15.955	0.00	-19.667	-15.955	0.00	-139.25	-15.955	0.00
1675*	5e+0	1	5	3.75	-75.550	0.16	-96.864	-75.669	0.00	-435.48	-75.669	0.00
1703	1e+1	1	10	6.96	-132.539	0.20	-180.935	-131.466	1.01	-929.92	-	-
1745†	5e+0	1	8	4.75	-71.828	0.76	-77.465	-72.377	0.00	-317.99	-72.377	0.00
1773*	1e+0	1	8	5.44	-14.633	0.06	-21.581	-14.642	0.00	-118.65	-14.636	0.04
1886	5e+0	2	9	5.84	-78.659	0.02	-135.615	-49.684	36.84	-324.87	-78.672	0.00
1913	5e+0	1	20	12.48	-51.866	0.47	-68.555	-52.109	0.00	-164.26	-51.348	1.46
1922*	5e+0	1	7	4.34	-35.452	1.39	-121.872	-35.916	0.10	-123.2	-35.951	0.00
1931	5e+0	1	10	5.87	-54.894	1.46	-85.196	-55.709	0.00	-204.08	-54.290	2.55
1967	1e+1	1	6	5.49	-104.752	2.63	-136.098	0.000	100	-622.57	-107.581	0.00
Avg	7.6	1.3	11.1	6.92		0.76			10.85			1.12
Max	100	4	53	29.24		2.87			100			6.78

† Rows 1751 and 1745 are excluded from average and maximum computations due to missing entries.

* $i^{\text{feas}} = 1$ is guaranteed by Theorem 2.

4.1.2 Choice of the penalty parameter η

In this experiment the sensitivity of different penalization methods to the choice of the penalty parameter η is tested. To this end, one iteration of the penalized SDP (6a)-(6d) is solved for a wide range of η values. The benchmark case 1143 is used for this experiment. If η is small, none of the proposed penalized SDPs are tight for the case 1143. As the value of η increases, the feasibility violation $\operatorname{tr}\{\overset{*}{\boldsymbol{X}} - \overset{*}{\boldsymbol{x}}\overset{*}{\boldsymbol{x}}^{\top}\}$ abruptly vanishes once crossing $\eta = 1.9$, $\eta = 7.7$, and $\eta = 19.6$, for the penalized 2×2 SDP, SDP and SDP+RLT, respectively. Remarkably, if $\overset{*}{\boldsymbol{x}}^{\mathrm{SDP+RLT}}$ is used as the initial point and $\eta \simeq 2$, then the penalized SDP+RLT (6a)–(6d) produces a feasible point for the benchmark case 1143 whose objective value is within 0.2% of the reported optimal cost $q_0(\boldsymbol{x}^{\mathrm{QPLIB}})$.

Additionally, Figure 3 shows the result of one iteration of penalized SDP for a wide range of η values, on QPLIB instances 1423, 1675, and 1967. As demonstrated by the figures, the resulting objective values of penalized SDP grow slowly beyond certain η. This indicates that the proposed approach is not very sensitive to the choice of η and a wide range of η values can be used for penalization.

Table 6: Sequential penalized SDP+RLT.

Inst	**Sequential SDP+RLT**						**BARON**			**COUENNE**		
	η	i^{feas}	i^{stop}	t(s)	UB	GAP(%)	LB	UB	GAP(%)	LB	UB	GAP(%)
0343	0e+0	0	0	1.42	-6.386	0.00	-95.372	-6.386	0.00	-7668.005	-6.386	0.00
0911	2e-1	4	5	13.08	-32.147	0.00	-172.777	0.000	100	-172.777	-31.026	3.49
0975	2e-1	3	5	12.75	-37.852	0.00	-47.428	-37.794	0.16	-171.113	-36.812	2.75
1055	1e+0	5	8	9.56	-32.874	0.49	-37.841	-33.037	0.00	-199.457	-33.037	0.00
1143	5e-1	4	5	7.27	-57.241	0.01	-69.522	-57.247	0.00	-384.45	-53.367	6.78
1157	0e+0	0	0	0.88	-10.948	0.00	-11.414	-10.948	0.00	-80.51	-10.948	0.00
1353	0e+0	0	0	0.45	-7.714	0.00	-7.925	-7.714	0.00	-73.28	-7.714	0.00
1423	2e-1	1	2	2.82	-14.929	0.25	-16.313	-14.968	0.00	-76.13	-14.078	5.94
1437	1e-2	1	2	7.02	-7.789	0.00	-9.601	-7.789	0.00	-87.58	-7.789	0.00
1451	2e+0	2	5	24.45	-87.573	0.01	-135.140	-87.577	0.00	-468.04	-86.860	0.82
1493	5e-1	1	2	2.76	-43.160	0.00	-47.239	-43.160	0.00	-395.69	-43.160	0.00
1507	0e+0	0	0	0.61	-8.301	0.00	-49.709	-8.301	0.00	-44.37	-8.301	0.00
1535	5e-1	1	10	38.01	-11.536	0.43	-13.407	-11.397	1.63	-107.86	-11.398	1.62
1619	0e+0	0	0	2.38	-9.217	0.00	-10.302	-9.217	0.00	-74.55	-9.217	0.00
1661	1e-1	1	2	12.88	-15.955	0.00	-19.667	-15.955	0.00	-139.25	-15.955	0.00
1675	5e-1	4	0	4.22	-75.669	0.00	-96.864	-75.669	0.00	-435.48	-75.669	0.00
1703	2e+0	1	3	13.50	-72.376	0.00	-77.465	-	-	-317.99	-72.377	0.00
1773	2e-1	3	4	18.01	-14.626	0.11	-21.581	-14.642	0.00	-118.65	-14.636	0.04
1886	2e+0	2	4	9.05	-78.643	0.04	-135.615	-78.672	0.00	-324.87	-78.672	0.00
1913	1e+0	2	6	11.49	-52.108	0.00	-68.555	-52.109	0.00	-164.26	-51.348	1.46
1922	2e+0	1	5	3.35	-35.556	1.10	-121.872	-35.741	0.58	-123.2	-35.951	0.00
1931	1e+0	1	2	2.99	-55.674	0.06	-85.196	-53.760	3.50	-204.08	-54.290	2.55
1967	5e+0	1	8	16.11	-107.052	0.49	-136.098	0.000	100	-622.57	-107.581	0.00
Avg	0.8	1.7	3.39	9.35		0.14			8.96			1.11
Max	5	5	10	38		1.1			100			6.78

† Row 1745 is excluded from average and maximum computations due to missing entries.

4.2 Large-scale system identification problems

The advantage of the panelized SDP approach becomes clearer for large scale sparse QCQPs that are beyond the capabilities of the state-of-the-art non-convex solvers.

Following [17], this case study is concerned with the problem of identifying the parameters of linear dynamical systems given limited observations and non-uniform snapshots of state vectors. Optimization is an important tool for problems involving dynamical systems such as the identification of transfer functions and control synthesis [18, 31, 54, 60, 61]. One of these computationally-hard problems is system identification based solely on data (without intrusive means) which has been widely studied in the literature of control [52, 55]. In this case study, system identification is cast as a non-convex QCQP and evaluate the ability of the proposed penalized SDP in solving very large scale instances of this problem.

Consider a discrete-time linear system described by the system of equations:

$$\boldsymbol{z}[\tau+1] = \boldsymbol{A}\boldsymbol{z}[\tau] + \boldsymbol{B}\boldsymbol{u}[\tau] + \boldsymbol{w}[\tau] \qquad \tau = 1, 2, \ldots, T-1 \tag{46a}$$

where

- $\{\boldsymbol{z}[\tau] \in \mathbb{R}^n\}_{\tau=1}^{T}$ are the state vectors that are known at times $\tau \in \{\tau_1, \ldots, \tau_o\}$,
- $\{\boldsymbol{u}[\tau] \in \mathbb{R}^m\}_{\tau=1}^{T}$ are the known control command vectors.

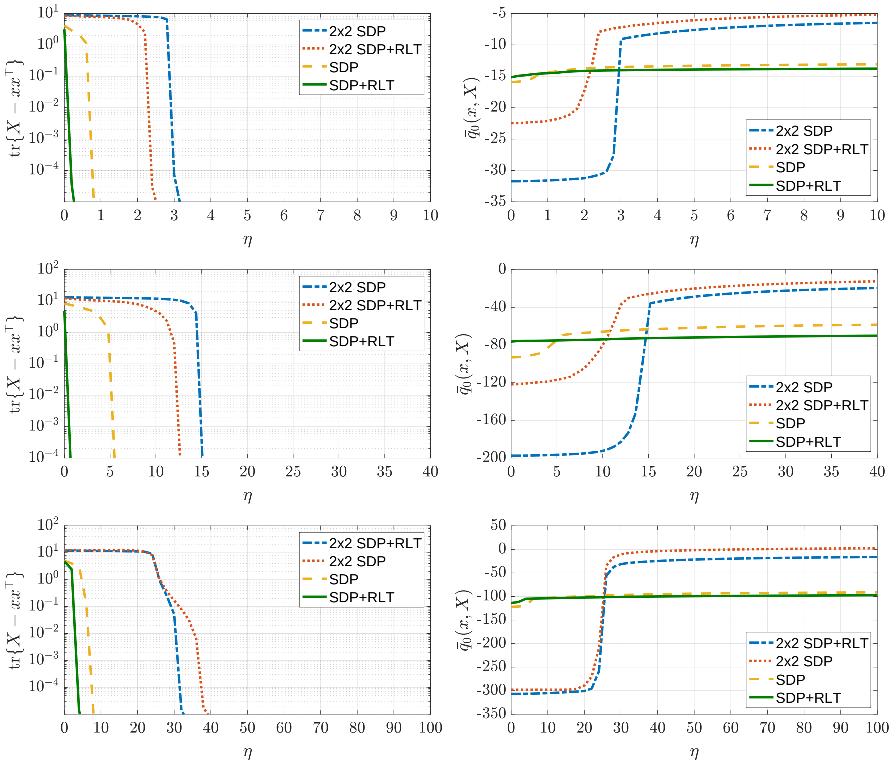

Fig. 3: The effect of η on the performance of penalized 2×2 SDP, 2×2 SDP+RLT, SDP, and SDP+RLT for cases QPLIB 1423, 1675, and 1967.

– $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{B} \in \mathbb{R}^{n \times m}$ are fixed unknown matrices, and
– $\{\boldsymbol{w}[\tau] \in \mathbb{R}^n\}_{\tau=1}^{T}$ account for the unknown disturbance vectors.

Our goal is to estimate the pair of ground truth matrices $(\bar{\boldsymbol{A}}, \bar{\boldsymbol{B}})$, given a sample trajectory of the control commands $\{\bar{\boldsymbol{u}}[\tau] \in \mathbb{R}^n\}_{\tau=1}^{T}$ and the incomplete state vectors $\{\bar{\boldsymbol{z}}[\tau] \in \mathbb{R}^n\}_{\tau \in \{\tau_1, \ldots, \tau_o\}}$. To this end, we employ the minimum least absolute value estimator which amounts to the following QCQP:

$$\underset{\substack{\{\boldsymbol{y}[\tau] \in \mathbb{R}^n\}_{\tau=1}^{T-1} \\ \{\boldsymbol{z}[\tau] \in \mathbb{R}^n\}_{\tau=1}^{T} \\ \boldsymbol{A} \in \mathbb{R}^{n \times n} \\ \boldsymbol{B} \in \mathbb{R}^{n \times m}}}{\text{minimize}} \quad \sum_{\tau=1}^{T-1} \boldsymbol{1}_n^\top \boldsymbol{y}[\tau] \tag{47a}$$

$$\text{subject to} \quad \boldsymbol{y}[\tau] \geq +\boldsymbol{z}[\tau+1] - \boldsymbol{A}\boldsymbol{z}[\tau] - \boldsymbol{B}\bar{\boldsymbol{u}}[\tau] \qquad \tau \in \{1, 2, \ldots, T-1\} \tag{47b}$$

$$\boldsymbol{y}[\tau] \geq -\boldsymbol{z}[\tau+1] + \boldsymbol{A}\boldsymbol{z}[\tau] + \boldsymbol{B}\bar{\boldsymbol{u}}[\tau] \qquad \tau \in \{1, 2, \ldots, T-1\} \tag{47c}$$

$$\boldsymbol{z}[\tau] = \bar{\boldsymbol{z}}[\tau] \qquad \tau \in \{\tau_1, \ldots, \tau_o\}. \tag{47d}$$

For every $\tau \in \{1, 2, \ldots, T-1\}$, the auxiliary variable $\boldsymbol{y}[\tau] \in \mathbb{R}^n$ accounts for $|\boldsymbol{z}[\tau+1] - \boldsymbol{A}\boldsymbol{z}[\tau] - \boldsymbol{B}\bar{\boldsymbol{u}}[\tau]|$. This relation is imposed through the pair of constraints (47b) and (47c).

The problem (47a)–(47d), can be cast in the form of (1a)–(1c), with respect to the vector

$$\boldsymbol{x} \triangleq [\boldsymbol{z}[1]^\top, \ldots, \boldsymbol{z}[T]^\top, \mathrm{vec}\{\boldsymbol{A}\}^\top, \alpha\, \boldsymbol{y}[1]^\top, \ldots, \alpha\, \boldsymbol{y}[T-1]^\top, \alpha\, \mathrm{vec}\{\boldsymbol{B}\}^\top], \quad (48)$$

where α is a preconditioning constant. To solve the resulting problem, we use the sequential Algorithm 1 equipped with the 2×2 SDP relaxation and the initial point $\hat{\boldsymbol{x}} = \boldsymbol{0}$.

We consider system identification problems with $n = 25$, $m = 20$, $T = 500$ and $o = 400$. In every experiment, $\{\tau_1, \ldots, \tau_o\}$ is a uniformly selected subset of $\{1, 2, \ldots, T\}$. The resulting QCQP variable $\boldsymbol{x}$ is 23605-dimensional and the problem is 16100-dimensional if we exclude the known state vectors $\{\bar{\boldsymbol{z}}[\tau] \in \mathbb{R}^n\}_{\tau \in \{\tau_1, \ldots, \tau_o\}}$. Due to sparsity of the QCQP (47a)–(47d) each iteration of the penalized 2×2 SDP is solved within 30 minutes, by omitting the elements of the lifted variable $\boldsymbol{X}$ that do not appear in the objective and constraints. All of the convex programs are solved using MOSEK v8.1 [3] through MATLAB 2017a and on a desktop computer with a 12-core 3.0GHz CPU and 256GB RAM. Due to the sheer size of the problem, we were only able to solve instances with $T \leq 70$ using Baron and Couenne, neither of which resulted in successful recovery of the unknown matrices due to limited data points.

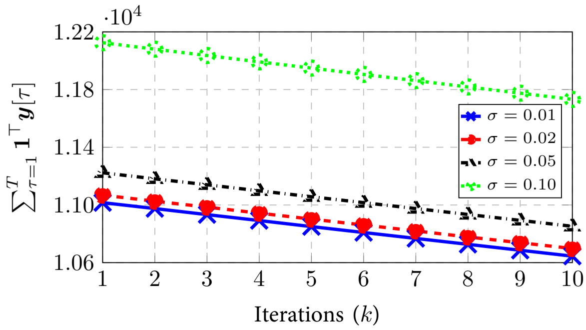

Fig. 4: Convergence of the sequential penalized 2×2 SDP for large-scale system identification with different disturbance levels.

The ground truth values are chosen as follows:

- The elements of $\bar{\boldsymbol{A}} \in \mathbb{R}^{25 \times 25}$ have zero-mean Gaussian distribution and the matrix is scaled in such a way that the largest singular value is equal to 0.5.
- Every element of $\bar{\boldsymbol{B}} \in \mathbb{R}^{25 \times 20}$, $\{\bar{\boldsymbol{u}}[\tau] \in \mathbb{R}^{20}\}_{\tau=1}^{T}$ and $\bar{\boldsymbol{z}}[1] \in \mathbb{R}^{25}$ have standard normal distribution.
- The elements of $\{\bar{\boldsymbol{w}}[\tau] \in \mathbb{R}^{25}\}_{\tau=1}^{T-1}$ have independent zero-mean Gaussian distribution with the standard deviation $\sigma \in \{0.01, 0.02, 0.05, 0.10\}$.

For each experiment, we ran Algorithm 1 for 10 iterations. The preconditioning and penalty terms are set to $\alpha = 10^{-3}$ and $\eta = 40$, respectively. For each $\sigma \in \{0.01, 0.02, 0.05, 0.10\}$, we have run 10 random experiments resulting in the average recovery errors 0.0005, 0.0010, 0.0026, and 0.0062, respectively, for $\frac{1}{n}\|\bar{\boldsymbol{A}} - \boldsymbol{A}^{(10)}\|_F$, and the average errors 0.0014, 0.0028, 0.0070, and 0.0141, respectively, for $(mn)^{-\frac{1}{2}}\|\bar{\boldsymbol{B}} - \boldsymbol{B}^{(10)}\|_F$. In all of the trials, a feasible point is obtained in the first iteration of Algorithm 1. Figure 4 illustrates the convergence behavior of the objective functions for one of the trials for each disturbance level.

5 Conclusions

This paper introduces a penalized SDP approach for recovering feasible and near-optimal solutions to nonconvex quadratically-constrained quadratic programming (QCQP) problems. Given an arbitrary initial point (feasible or infeasible) for the original QCQP, penalized semidefinite programs are formulated by adding a linear term to the objective. A generalized linear independence constraint qualification (LICQ) condition is introduced as a regularity criterion for initial points, and it is shown that the solution of penalized SDP is feasible for QCQP if the initial point is regular and close to the feasible set. We show that the proposed penalized SDPs can be solved sequentially in order to improve the objective of the feasible solution. Numerical experiments on QPLIB benchmark cases demonstrate that the proposed sequential approach performs comparable to nonconvex optimizers Baron and Couenne. More importantly, it allows one to solve large scale sparse QCQPs, which are beyond the capabilities of the state-of-the are solvers, as demonstrated on large-scale system identification problems.

Acknowledgements The authors are grateful to GAMS Development Corporation for providing them with unrestricted access to a full set of solvers throughout the project.

References

1. Ahmadi AA, Majumdar A (2018) DSOS and SDSOS optimization: more tractable alternatives to sum of squares and semidefinite optimization. SIAM Journal on Applied Algebraic Geometry,
2. Alizadeh F, Goldfarb D (2003) Second-order cone programming. Mathematical Programming 95(1):3–51
3. ApS M (2017) The MOSEK optimization toolbox for MATLAB manual. Version 8.1. URL `http://docs.mosek.com/8.1/toolbox/index.html`
4. Ashraphijuo M, Madani R, Lavaei J (2016) Characterization of rank-constrained feasibility problems via a finite number of convex programs. In: 2016 IEEE 55th Conference on Decision and Control (CDC), IEEE, pp 6544–6550
5. Atamtürk A, Gómez A (2019) Rank-one convexification for sparse regression. arXiv preprint arXiv:190110334 BCOL Research Report 19.01, IEOR, UC Berkeley

6. Atamtürk A, Narayanan V (2007) Cuts for conic mixed-integer programming. In: Fischetti M, Williamson DP (eds) Integer Programming and Combinatorial Optimization, Springer, Berlin, Heidelberg, pp 16–29
7. Bao X, Sahinidis NV, Tawarmalani M (2011) Semidefinite relaxations for quadratically constrained quadratic programming: A review and comparisons. Mathematical Programming 129:129–157
8. Belotti P (2013) COUENNE: A user's manual. Tech. rep., Technical report, Lehigh University
9. Burer S, Vandenbussche D (2008) A finite branch-and-bound algorithm for nonconvex quadratic programming via semidefinite relaxations. Mathematical Programming 113(2):259–282
10. Burer S, Ye Y (2018) Exact semidefinite formulations for a class of (random and non-random) nonconvex quadratic programs. arXiv preprint arXiv:180202688
11. Burgdorf S, Laurent M, Piovesan T (2015) On the closure of the completely positive semidefinite cone and linear approximations to quantum colorings. arXiv preprint arXiv:150202842
12. Candes EJ, Strohmer T, Voroninski V (2013) Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. Communications on Pure and Applied Mathematics 66(8):1241–1274
13. Chen C, Atamtürk A, Oren SS (2017) A spatial branch-and-cut method for nonconvex QCQP with bounded complex variables. Mathematical Programming 165(2):549–577
14. Chen J, Burer S (2012) Globally solving nonconvex quadratic programming problems via completely positive programming. Mathematical Programming Computation 4(1):33–52
15. Cid C, Murphy S, Robshaw MJ (2005) Small scale variants of the aes. In: International Workshop on Fast Software Encryption, Springer, pp 145–162
16. Courtois NT, Pieprzyk J (2002) Cryptanalysis of block ciphers with overdefined systems of equations. In: International Conference on the Theory and Application of Cryptology and Information Security, Springer, pp 267–287
17. Fattahi S, Sojoudi S (2018) Data-driven sparse system identification. In: 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton), IEEE
18. Fattahi S, Fazelnia G, Lavaei J, Arcak M (2018) Transformation of optimal centralized controllers into near-globally optimal static distributed controllers. IEEE Transactions on Automatic Control 64(1):66–80
19. Fazelnia G, Madani R, Kalbat A, Lavaei J (2017) Convex relaxation for optimal distributed control problems. IEEE Transactions on Automatic Control 62(1):206–221
20. Fogel F, Waldspurger I, dAspremont A (2013) Phase retrieval for imaging problems. Mathematical Programming Computation pp 1–25
21. Furini F, Traversi E, Belotti P, Frangioni A, Gleixner A, Gould N, Liberti L, Lodi A, Misener R, Mittelmann H, Sahinidis N, Vigerske S, Wiegele A (2019) QPLIB: A library of quadratic programming instances. Mathematical Programming Computations 11:237–310, URL http://qplib.zib.de/

22. GAMS Development Corporation (2013) General Algebraic Modeling System (GAMS) Release 24.2.1. Washington, DC, USA, URL http://www.gams.com/
23. Gershman AB, Sidiropoulos ND, Shahbazpanahi S, Bengtsson M, Ottersten B (2010) Convex optimization-based beamforming: From receive to transmit and network designs. IEEE Signal Process Mag 27(3):62–75
24. Goemans MX, Williamson DP (1995) Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. Journal of the ACM (JACM) 42(6):1115–1145
25. He S, Luo Z, Nie J, Zhang S (2008) Semidefinite relaxation bounds for indefinite homogeneous quadratic optimization. SIAM Journal on Optimization 19:503–523
26. He S, Li Z, Zhang S (2010) Approximation algorithms for homogeneous polynomial optimization with quadratic constraints. Mathematical Programming 125:353–383
27. Hilling JJ, Sudbery A (2010) The geometric measure of multipartite entanglement and the singular values of a hypermatrix. Journal of Mathematical Physics 51(7):072102
28. Ibaraki S, Tomizuka M (2001) Rank minimization approach for solving BMI problems with random search. In: Proceedings of the 2001 American Control Conference.(Cat. No. 01CH37148), IEEE, vol 3, pp 1870–1875
29. Josz C, Molzahn DK (2018) Lasserre hierarchy for large scale polynomial optimization in real and complex variables. SIAM Journal on Optimization 28(2):1017–1048
30. Kheirandishfard M, Zohrizadeh F, Adil M, Madani R (2018) Convex relaxation of bilinear matrix inequalities part II: Applications to optimal control synthesis. In: IEEE 57th Annual Conference on Decision and Control (CDC)
31. Kheirandishfard M, Zohrizadeh F, Adil M, Madani R (2018) Convex relaxation of bilinear matrix inequalities part ii: Applications to optimal control synthesis. In: 2018 IEEE Conference on Decision and Control (CDC), IEEE, pp 75–82
32. Kheirandishfard M, Zohrizadeh F, Madani R (2018) Convex relaxation of bilinear matrix inequalities part I: Theoretical results. In: IEEE 57th Annual Conference on Decision and Control (CDC),
33. Kim S, Kojima M (2003) Exact solutions of some nonconvex quadratic optimization problems via SDP and SOCP relaxations. Computational Optimization and Applications 26(2):143–154
34. Kim S, Kojima M, Yamashita M (2003) Second order cone programming relaxation of a positive semidefinite constraint. Optimization Methods and Software 18:535–541
35. Lasserre JB (2001) An explicit exact SDP relaxation for nonlinear 0-1 programs. In: Integer Programming and Combinatorial Optimization, Springer, pp 293–303
36. Lasserre JB (2001) Global optimization with polynomials and the problem of moments. SIAM Journal on Optimization 11:796–817
37. Lasserre JB (2006) Convergent SDP-relaxations in polynomial optimization with sparsity. SIAM Journal on Optimization 17:822–843

38. Lovász L, Schrijver A (1991) Cones of matrices and set-functions and 0–1 optimization. SIAM Journal on Optimization 1(2):166–190
39. Luo Z, Sidiropoulos N, Tseng P, Zhang S (2007) Approximation bounds for quadratic optimization with homogeneous quadratic constraints. SIAM Journal on Optimization 18:1–28
40. Luo ZQ, Ma Wk, So AMC, Ye Y, Zhang S (2010) Semidefinite relaxation of quadratic optimization problems. IEEE Signal Processing Magazine 27(3):20
41. Madani R, Fazelnia G, Lavaei J (2014) Rank-2 matrix solution for semidefinite relaxations of arbitrary polynomial optimization problems. Preprint
42. Madani R, Sojoudi S, Lavaei J (2015) Convex relaxation for optimal power flow problem: Mesh networks. IEEE Transactions on Power Systems 30(1):199–211
43. Madani R, Ashraphijuo M, Lavaei J (2016) Promises of conic relaxation for contingency-constrained optimal power flow problem. IEEE Transactions on Power Systems 31(2):1297–1307
44. Majumdar A, Ahmadi AA, Tedrake R (2014) Control and verification of high-dimensional systems with DSOS and SDSOS programming. In: Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on, IEEE, pp 394–401
45. Mohammad-Nezhad A, Terlaky T (2017) A rounding procedure for semidefinite optimization. Submitted to Operations Research Letters
46. Mu C, Zhang Y, Wright J, Goldfarb D (2016) Scalable robust matrix recovery: Frank–Wolfe meets proximal methods. SIAM Journal on Scientific Computing 38(5):A3291–A3317
47. Muramatsu M, Suzuki T (2003) A new second-order cone programming relaxation for max-cut problems. Journal of the Operations Research Society of Japan 46:164–177
48. Natarajan K, Shi D, Toh KC (2013) A penalized quadratic convex reformulation method for random quadratic unconstrained binary optimization. Optimization Online
49. Nesterov Y (1998) Semidefinite relaxation and nonconvex quadratic optimization. Optimization Methods and Software 9:141–160
50. Nesterov Y, Nemirovskii AS, Ye Y (1994) Interior-point polynomial algorithms in convex programming. SIAM
51. Papp D, Alizadeh F (2013) Semidefinite characterization of sum-of-squares cones in algebras. SIAM Journal on Optimization 23(3):1398–1423
52. Pereira J, Ibrahimi M, Montanari A (2010) Learning networks of stochastic differential equations. In: Advances in Neural Information Processing Systems, pp 172–180
53. Permenter F, Parrilo P (2014) Partial facial reduction: simplified, equivalent sdps via approximations of the psd cone. Mathematical Programming pp 1–54
54. Rotkowitz M, Lall S (2005) A characterization of convex problems in decentralized control. IEEE transactions on Automatic Control 50(12):1984–1996
55. Sarkar T, Rakhlin A (2018) How fast can linear dynamical systems be learned? arXiv preprint arXiv:181201251
56. Sherali HD, Adams WP (1990) A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. SIAM Journal on Discrete Mathematics 3(3):411–430

57. Sherali HD, Adams WP (2013) A reformulation-linearization technique for solving discrete and continuous nonconvex problems, vol 31. Springer Science & Business Media
58. Sojoudi S, Lavaei J (2013) On the exactness of semidefinite relaxation for nonlinear optimization over graphs: Part I. In: Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on, IEEE, pp 1043–1050
59. Tawarmalani M, Sahinidis NV (2005) A polyhedral branch-and-cut approach to global optimization. Mathematical Programming 103:225–249
60. Wang YS, Matni N, Doyle JC (2018) Separable and localized system-level synthesis for large-scale systems. IEEE Transactions on Automatic Control 63(12):4234–4249
61. Wang YS, Matni N, Doyle JC (2019) A system level approach to controller synthesis. IEEE Transactions on Automatic Control
62. Ye Y (1999) Approximating global quadratic optimization with convex quadratic constraints. Journal of Global Optimization 15:1–17
63. Ye Y (1999) Approximating quadratic programming with bound and quadratic constraints. Mathematical Programming 84:219–226
64. Zhang S (2000) Quadratic maximization and semidefinite relaxation. Mathematical Programming 87:453–465
65. Zhang S, Huang Y (2006) Complex quadratic optimization and semidefinite programming. SIAM Journal on Optimization 87:871–890
66. Zohrizadeh F, Kheirandishfard M, Nasir A, Madani R (2018) Sequential relaxation of unit commitment with AC transmission constraints. In: IEEE 57th Annual Conference on Decision and Control (CDC)
67. Zohrizadeh F, Kheirandishfard M, Quarm E, Madani R (2018) Penalized parabolic relaxation for optimal power flow problem. In: IEEE 57th Annual Conference on Decision and Control (CDC)

A Application to polynomial optimization

In this section, we show that the proposed penalized SDP approach can be used for polynomial optimization as well. A polynomial optimization problem is formulated as

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad u_0(\boldsymbol{x}) \tag{49a}$$

$$\text{s.t.} \quad u_k(\boldsymbol{x}) \leq 0, \qquad k \in \mathcal{I} \tag{49b}$$

$$u_k(\boldsymbol{x}) = 0, \qquad k \in \mathcal{E}, \tag{49c}$$

for every $k \in \{0\} \cup \mathcal{I} \cup \mathcal{E}$, where each function $u_k : \mathbb{R}^n \to \mathbb{R}$ is a polynomial of arbitrary degree. Problem (49a)–(49c) can be reformulated as a QCQP of the form:

$$\underset{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{y}\in\mathbb{R}^o}{\text{minimize}} \quad w_0(\boldsymbol{x},\boldsymbol{y}) \tag{50a}$$

$$\text{s.t.} \quad w_k(\boldsymbol{x},\boldsymbol{y}) \leq 0, \qquad k \in \mathcal{I} \tag{50b}$$

$$w_k(\boldsymbol{x},\boldsymbol{y}) = 0, \qquad k \in \mathcal{E} \tag{50c}$$

$$v_i(\boldsymbol{x},\boldsymbol{y}) = 0, \qquad i \in \mathcal{O}, \tag{50d}$$

where $\boldsymbol{y} \in \mathbb{R}^{|\mathcal{O}|}$ is an auxiliary variable, and $v_1, \ldots, v_{|\mathcal{O}|}$ and $w_0, w_1, \ldots, w_{|\{0\}\cup\mathcal{I}\cup\mathcal{E}|}$ are quadratic functions with the following properties:

– For every $\boldsymbol{x} \in \mathbb{R}^n$, the function $\boldsymbol{v}(\boldsymbol{x}, \cdot) : \mathbb{R}^{|\mathcal{O}|} \to \mathbb{R}^{|\mathcal{O}|}$ is invertible,
– If $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{0}_n$, then $w_k(\boldsymbol{x}, \boldsymbol{y}) = u_k(\boldsymbol{x})$ for every $k \in \{0\} \cup \mathcal{I} \cup \mathcal{E}$.

Based on the above properties, there is a one-to-one correspondence between the feasible sets of (49a)–(49c) and (50a)–(50d). Moreover, a feasible point $(\overset{*}{\boldsymbol{x}}, \overset{*}{\boldsymbol{y}})$ is an optimal solution to the QCQP (50a)–(50d) if and only if $\overset{*}{\boldsymbol{x}}$ is an optimal solution to the polynomial optimization problem (49a)–(49c).

Theorem 3 [41]) *Suppose that $\{u_k\}_{k \in \{0\} \cup \mathcal{I} \cup \mathcal{E}}$ are polynomials of degree at most d, consisting of m monomials in total. There exists a QCQP reformulation of the polynomial optimization* (49a)–(49c) *in the form of* (50a)–(50d)*, where $|\mathcal{O}| \leq mn\,(\lfloor \log_2(d) \rfloor + 1)$.*

The next proposition shows that the LICQ regularity of a point $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ is inherited by the corresponding point $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \in \mathbb{R}^n \times \mathbb{R}^o$ of the QCQP reformulation (50a)-(50d).

Proposition 1 *Consider a pair of vectors $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ and $\hat{\boldsymbol{y}} \in \mathbb{R}^{|\mathcal{O}|}$ satisfying $\boldsymbol{v}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = \mathbf{0}_n$. The following two statements are equivalent:*

1. *$\hat{\boldsymbol{x}}$ is feasible and satisfies the LICQ condition for the polynomial optimization problem* (49a)–(49b)*.*
2. *$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is feasible and satisfies the LICQ condition for the QCQP* (50a)–(50d)*.*

Proof From $\boldsymbol{u}(\hat{\boldsymbol{x}}) = \boldsymbol{w}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ and the invertiblity assumption for $\boldsymbol{v}(\hat{\boldsymbol{x}}, \cdot)$, we have

$$\begin{aligned} \frac{\partial \boldsymbol{u}(\hat{\boldsymbol{x}})}{\partial \boldsymbol{x}} &= \begin{bmatrix} \frac{\partial \boldsymbol{w}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{x}} & \frac{\partial \boldsymbol{w}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{y}} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & -\left(\frac{\partial \boldsymbol{v}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{y}}\right)^{-1} \frac{\partial \boldsymbol{v}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{x}} \end{bmatrix}^\top \\ &= \frac{\partial \boldsymbol{w}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{x}} - \frac{\partial \boldsymbol{w}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{y}} \left(\frac{\partial \boldsymbol{v}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{y}}\right)^{-1} \frac{\partial \boldsymbol{v}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{x}}. \end{aligned} \tag{51}$$

Therefore, $\mathcal{J}_{\mathrm{PO}}(\hat{\boldsymbol{x}}) = \frac{\partial \boldsymbol{u}(\hat{\boldsymbol{x}})}{\partial \boldsymbol{x}}$ is equal to the Schur complement of

$$\mathcal{J}_{\mathrm{QCQP}}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = \begin{bmatrix} \frac{\partial \boldsymbol{w}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{x}} & \frac{\partial \boldsymbol{w}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{y}} \\ \frac{\partial \boldsymbol{v}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{x}} & \frac{\partial \boldsymbol{v}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{y}} \end{bmatrix}, \tag{52}$$

which is the Jacobian matrix of the QCQP (50a)–(50d) at the point $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$. As a result, the matrix $\mathcal{J}_{\mathrm{PO}}(\hat{\boldsymbol{x}})$ is singular if and only if $\mathcal{J}_{\mathrm{QCQP}}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is singular.

B Additional strengthening by RLT

This appendix presents the reformulation-linearization technique (RLT) [57] inequalities used to strengthen convex relaxations (4a)–(4d) in our computations in the presence of affine constraints. Define $\mathcal{L}$ as the set of affine constrains in the QCQP (1a)–(1c), i.e., $\mathcal{L} \triangleq \{k \in \mathcal{I} \cup \mathcal{E} \mid \boldsymbol{A}_k = \mathbf{0}_{n \times n}\}$. Define also

$$\boldsymbol{H} \triangleq [\, \boldsymbol{B}\{\mathcal{L} \cap \mathcal{I}\}^\top, \boldsymbol{B}\{\mathcal{L} \cap \mathcal{E}\}^\top, -\boldsymbol{B}\{\mathcal{L} \cap \mathcal{E}\}^\top]^\top, \tag{53a}$$

$$\boldsymbol{h} \triangleq [\ \boldsymbol{c}\{\mathcal{L} \cap \mathcal{I}\}^\top,\ \boldsymbol{c}\{\mathcal{L} \cap \mathcal{E}\}^\top, -\ \boldsymbol{c}\{\mathcal{L} \cap \mathcal{E}\}^\top\ ]^\top, \tag{53b}$$

where $\boldsymbol{B} \triangleq [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{|\mathcal{I} \cap \mathcal{E}|}]^\top$ and $\boldsymbol{c} \triangleq [c_1, \ldots, c_{|\mathcal{I} \cap \mathcal{E}|}]^\top$. Every $\boldsymbol{x} \in \mathcal{F}$ satisfies

$$\boldsymbol{H}\boldsymbol{x} + \boldsymbol{h} \leq 0, \tag{54}$$

and, as a result, all elements of the matrix

$$\boldsymbol{H}\boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{H}^\top + \boldsymbol{h}\boldsymbol{x}^\top \boldsymbol{H}^\top + \boldsymbol{H}\boldsymbol{x}\boldsymbol{h}^\top + \boldsymbol{h}\boldsymbol{h}^\top \tag{55}$$

are nonnegative if $\boldsymbol{x}$ is feasible. Hence, the inequality

$$\boldsymbol{e}_i^\top \boldsymbol{V}(\boldsymbol{x}, \boldsymbol{x}\boldsymbol{x}^\top)\boldsymbol{e}_j \geq 0 \tag{56}$$

holds true for every $\boldsymbol{x} \in \mathcal{F}$ and $(i,j) \in \mathcal{H} \times \mathcal{H}$, where $\boldsymbol{V} : \mathbb{R}^n \times \mathbb{S}_n \to \mathbb{S}_{|\mathcal{H}|}$ is defined as

$$\boldsymbol{V}(\boldsymbol{x}, \boldsymbol{X}) \triangleq \boldsymbol{H}\boldsymbol{X}\boldsymbol{H}^\top + \boldsymbol{h}\boldsymbol{x}^\top\boldsymbol{H}^\top + \boldsymbol{H}\boldsymbol{x}\boldsymbol{h}^\top + \boldsymbol{h}\boldsymbol{h}^\top, \tag{57}$$

$\mathcal{H} \triangleq \{1, \ldots, |\mathcal{L} \cap \mathcal{I}| + 2|\mathcal{L} \cap \mathcal{E}|\}$, and $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{|\mathcal{H}|}$ denote the standard bases in $\mathbb{R}^{|\mathcal{H}|}$.

This leads to a strengthened relaxation of QCQP (1a)–(1c):

$$\begin{aligned}
\underset{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{X}\in\mathbb{S}_n}{\text{minimize}} \quad & \bar{q}_0(\boldsymbol{x}, \boldsymbol{X}) && \text{(58a)}\\
\text{s.t.} \quad & \bar{q}_k(\boldsymbol{x}, \boldsymbol{X}) \leq 0, \qquad k \in \mathcal{I} && \text{(58b)}\\
& \bar{q}_k(\boldsymbol{x}, \boldsymbol{X}) = 0, \qquad k \in \mathcal{E} && \text{(58c)}\\
& \boldsymbol{X} - \boldsymbol{x}\boldsymbol{x}^\top \succeq_{\mathcal{C}_r} 0 && \text{(58d)}\\
& \boldsymbol{e}_i^\top \boldsymbol{V}(\boldsymbol{x}, \boldsymbol{X})\boldsymbol{e}_j \geq 0, \qquad (i,j) \in \mathcal{V} && \text{(58e)}
\end{aligned}$$

where $\mathcal{V} \subseteq \mathcal{H} \times \mathcal{H}$ is a selection of RLT inequalities.