Solving a new type of quadratic optimization problem having a joint numerical range constraint

Huu-Quang Nguyen · Ruey-Lin Sheu · Yong Xia

Received: date / Accepted: date

Abstract We propose a new formulation about quadratic optimization problems. The objective function \( F(f(x), g(x)) \) is given as composition of a quadratic function \( F(z) \) with two \( n \)-variate quadratic functions \( z_1 = f(x) \) and \( z_2 = g(x) \). In addition, it incorporates with a set of linear inequality constraints in \( z = (z_1, z_2)^T \), while having an implicit constraint that \( z \) belongs to the joint numerical range of \( (f, g) \). The formulation is very general in the sense that it covers quadratic programming with a single quadratic constraint of all types, including the inequality-type, the equality-type, and the interval-type. Even more, the composition of “quadratic with quadratics” as well as the joint numerical range constraint all together allow us to formulate existing unsolved (or not solved efficiently) problems into the new model. In this paper, we solve the quadratic hypersurfaces intersection problem (QSIC) proposed by Pólik and Terlaky; and the problem (AQP) to minimize the absolute value of a quadratic function over a quadratic constraint proposed by Ye and Zhang. We show that, when \( F(z) \) and the joint numerical range constraint are both convex, the optimal value of the convex optimization problem can be obtained by solving an SDP followed from a new development of the \( S \)-procedure. The optimal solution can be approximated by conducting a bisection method on \([0, 2\pi]\). On the other hand, if the joint numerical range of \( f(x) \) and \( g(x) \) is non-convex, the respective quadratic matrices of \( f(x) \) and \( g(x) \) must be linearly dependent. The linear dependence property enables us to solve (QSIC) and (AQP) accordingly by elementary analysis.
Keywords QCQP, Joint numerical range, S-procedure, Hidden convexity, Intersection of quadratic hypersurfaces, Absolute value of a quadratic function.

1 Introduction

The minimization problem considered in this paper takes the following form:

\[
\begin{align*}
v(\text{Po4}) &= \inf_{(x,z) \in \mathbb{R}^n \times \mathbb{R}^2} F(z) \\
&\text{s.t. } z_1 a + z_2 b - c \leq 0, \\
&\quad z \in C = \{(f(x), g(x)) | x \in \mathbb{R}^n\},
\end{align*}
\]

where \(a, b, c \in \mathbb{R}^m; f(x) = x^T P x + p^T x + p_0, g(x) = x^T Q x + q^T x + q_0\) are real-valued quadratic functions of \(n\)-variables with symmetric matrices \(P, Q; p, q \in \mathbb{R}^n\) and \(p_0, q_0 \in \mathbb{R}\). The objective function is a two-variate polynomial in \(z = (z_1, z_2)^T:\)

\[
F(z) = z^T \Theta z + \eta^T z
= \theta_1 z_1^2 + 2 \theta_2 z_1 z_2 + \theta_3 z_2^2 + \eta_1 z_1 + \eta_2 z_2,
\]

with the coefficients \(\theta_1, \theta_2, \theta_3, \eta_1, \eta_2 \in \mathbb{R}\).

The model (Po4) has a very special implicit constraint set

\[
C = \{(f(x), g(x)) | x \in \mathbb{R}^n\}.
\]

It is the joint numerical range of two quadratics \(f\) and \(g\), which, in general, is a non-convex set. Even if \(C\) is convex, since it is not described specifically by convex functions (note that both \(f, g\) could be non-convex), the problem (Po4) is still not easy to tackle. One of the major contributions in this paper is to show that, when \(F\) and the joint numerical range constraint \(C\) are both convex, the (abstract\(^1\)) convex optimization problem can be solved by an SDP followed from a new development of the S-procedure. In this case, (Po4) belongs to the hidden convex optimization \([19]\).

The model (Po4) also has a nice extensibility so that it can be tailored to describe many existing quadratic optimization problems and creates more new ones. By writing the implicit constraint \(z \in C\) explicitly into a kind of quadratically constrained quadratic problems (abbreviated as (QCQP) in literature) in \(\mathbb{R}^{n+2}\):

\[
\begin{align*}
v(\text{Po4}) &= \inf_{(x,z) \in \mathbb{R}^n \times \mathbb{R}^2} F(z) \\
&\text{s.t. } f(x)a + g(x)b - c \leq 0; \\
&\quad f(x) - z_1 = 0; \\
&\quad g(x) - z_2 = 0,
\end{align*}
\]

we see, in the following examples, that \(af(x) + bg(x) - c \leq 0\) can be used to model quadratic inequality/equality constraints with suitable choices of \(a, b, c\), while \(f(x) - z_1 = 0, g(x) - z_2 = 0\) used for perturbations of the values in \(f(x)\) and \(g(x)\) to give meaningful applications. Interesting examples include

---

\(^1\) Some researchers use the term “abstract” to describe the convex problem which minimizes a convex function over a convex set, while the convex set might not be described by convex functions with inequalities. The case we face here falls into the category.
A new type of quadratic optimization problem

(i) When \( \theta_1 = \theta_2 = \theta_3 = 0, \eta_1 = 1, \eta_2 = 0, a = 0, b = 1, c = 0 \), the model \( \text{(Po4)} \) is reduced to a quadratic program with a single quadratic inequality constraint (abbreviated as \( \text{(QP1QC)} \) in literature):

\[
\inf_{x \in \mathbb{R}^n} f(x) \quad \text{s.t. } g(x) \leq 0,
\]

which has been well studied in [7,9,23].

(ii) When \( \theta_1 = \theta_2 = \theta_3 = 0, \eta_1 = 1, \eta_2 = 0, a = (0,0)^T, b = (1,-1)^T, c = (0,0)^T \), \( \text{(Po4)} \) becomes a quadratic program with a single quadratic equality constraint (abbreviated as \( \text{(QP1EQC)} \) in literature):

\[
\inf_{x \in \mathbb{R}^n} f(x) \quad \text{s.t. } g(x) = 0.
\]

See [11,21] for a complete solution to \( \text{(QP1EQC)} \).

(iii) When \( \theta_1 = \theta_2 = \theta_3 = 0, \eta_1 = 1, \eta_2 = 0, a = (0,0)^T, b = (1,-1)^T, c = (\beta,-\alpha)^T \) and \( m = 2 \), \( \text{(Po4)} \) becomes generalized trust region subproblem (abbreviated as \( \text{(GTRS)} \) in literature):

\[
\inf_{x \in \mathbb{R}^n} f(x) \quad \text{s.t. } \alpha \leq g(x) \leq \beta.
\]

For discussions on \( \text{(GTRS)} \), please refer to [7,11,16].

(iv) The double well potential problems (DWP) in [1,20]:

\[
\min_{x \in \mathbb{R}^n} \frac{1}{2} \left( \frac{1}{2} \|Bx - c\|^2 - d \right)^2 + \frac{1}{2} x^T Ax - u^T x
\]

can be represented by \( \text{(Po4)} \), too. Just choose \( \theta_1 = 1, \theta_2 = \theta_3 = 0, \eta_1 = 0, \eta_2 = 1, a = b = c = 0 \), and let \( f(x) = \frac{1}{2\sqrt{2}} \|Bx - c\|^2, g(x) = -\frac{1}{2} \|Bx - c\|^2 + \frac{1}{2} x^T Ax - u^T x + d^2 \).

(v) (Not solved before) In computer graphics, it is fundamental to check whether two quadratic surfaces in 3D Euclidean space intersect and to compute the intersection curve if they do. See [6,17,18]. Pólik and Terlaky [9] then posed the general quadratic intersections problem in \( \mathbb{R}^n \). They ask “Let \( f(x) = x^T P x + p^T x + p_0 = 0, g(x) = x^T Q x + q^T x + q_0 = 0 \) be two quadric hypersurfaces. Can we determine whether or not the two hypersurfaces intersect without actually computing the intersections?” In fact, we may further assume that the input data \((P,p,p_0)\) and \((Q,q,q_0)\) are contaminated by various noises and model the general quadratic intersections problem by the following nonlinear least squares model:

\[
\inf_{(x,z) \in \mathbb{R}^n \times \mathbb{R}^2} (z_1)^2 + (z_2)^2 \quad \text{s.t. } \begin{cases} 
- x^T P x + p^T x + p_0 - z_1 = 0, \\
x^T Q x + q^T x + q_0 - z_2 = 0.
\end{cases}
\]

Obviously, \( \text{(QSIC)} \) is a special form of \([3]\) with \( a = b = c = 0 \). To our best knowledge, \( \text{(QSIC)} \) has not been solved before.
(vi) (Not solved efficiently before) Ye and Zhang in [23] proposed to minimize the absolute value of a quadratic function over a quadratic constraint as follows:

\[
(AQP) \quad \inf_{x \in \mathbb{R}^n} |x^T P x + p^T x + p_0|
\]

s.t. \(x^T Q x + q^T x + q_0 \leq 0\).

Note that (AQP) can be equivalently modeled as

\[
\inf_{x \in \mathbb{R}^n} (x^T P x + p^T x + p_0)^2
\]

s.t. \(x^T Q x + q^T x + q_0 \leq 0\),

which is again a special type of (Po4) with \(\theta_1 = 1, \theta_2 = \theta_3 = 0, \eta_1 = \eta_2 = 0, a = c = 0, b = 1\).

In [23], (AQP) was formulated as a quadratic optimization problem subject to two quadratic constraints as follows:

\[
\min t
\]

s.t. \(x^T Q x + q^T x + q_0 \leq 0\),

\[-t \leq x^T P x + p^T x + p_0 \leq t,\]

and it was suggested to find the optimal value \(t^*\) with the bisection method.

The procedure requires to conduct the following feasibility check for (many) fixed \(t \geq 0\):

\[
(AQP-Feas) \quad \inf_{x \in \mathbb{R}^n} x^T Q x + q^T x + q_0
\]

s.t. \(-t \leq x^T P x + p^T x + p_0 \leq t,\)

which is a type of (GTRS) (see case (iii) above). The entire procedure in [23], though polynomially implementable, is very cumbersome due to having to execute an excessive number of SDP’s. Fortunately, our new method in this paper can resolve (AQP) by at most two SDP’s without any assumption (v.s. both primal and dual Slater conditions were assumed in [23]).

The above examples confirm that (Po4) can be widely used in modelling many quadratic optimization problems, but, to solve it in the most general case is difficult. Interestingly, we show that, under the convexity assumptions:

- \(\Theta = \begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & \theta_3 \end{pmatrix} \succeq 0;\) \(\Theta\) (5)

- the joint numerical range \(C = \{(f(x), g(x)) | x \in \mathbb{R}^n\}\) is convex, \(\Theta\) (6)

optimal value (Po4) can be solved by solving an (SDP). Given the convexity (5)-(6), the key is to use the separation theorem for developing a new type of \(S\)-procedure (Theorem 1 in Sect. 2).

The \(S\)-procedure raises the question “when a quadratic function (think it as an objective function) restricted to a set described by quadratic functions (constraint) can be non-negative?” Since it can transform a quadratic optimization problem equivalently to a family of feasibility problems and relates to Lagrange duality, it has become one of the fundamental tools in control theory and optimization. To survey the results before 2006, please refer to [4,9]. Historically, if the constraint
consists of just a single quadratic function, people refer it as $S$-Lemma. If there are at least two quadratics in the constraint set, people call it the $S$-procedure\footnote{Some people might prefer not to distinguishing the two terms between $S$-lemma and $S$-procedure though.}. In literature, only the $S$-lemma was studied completely with a necessary and sufficient condition. It includes three complete versions: the classical $S$-Lemma proved in 1971 by Yakubovich \cite{22}; the $S$-lemma with equality proved in 2016 by Xia et. al \cite{21}; and the $S$-Lemma with interval bounds proved in 2015 by Wang et. al \cite{16}. On the other hand, a complete version of the $S$-procedure for two or more quadratic functions remains open even for two quadratic forms. Our result in this paper (Theorem 1) is an incomplete version of the $S$-procedure with $m+2$ quadratic constraints under convexity assumptions \cite{20} and \cite{19}, but it suffices to solve (QSIC) and (AQP) completely.

To apply the new $S$-procedure, we need to know in advance whether the joint numerical range $C$ is convex. By Theorem 4.16 in \cite{2}, it is known that $C$ can be non-convex only when the two matrices $P$, $Q$ are linearly dependent. We thus solve the two problems (QSIC) and (AQP) by dividing them into two cases:

- Suppose \{ $P, Q$ \} are linearly independent; $C$ is convex. In this case, (QSIC) and (AQP) can be viewed as a kind of (Po4) satisfying conditions \cite{5} and \cite{6}. The complete solution procedure is stated in Section 3.

- If $P = t^*Q$ (or $Q = t^*P$), by elementary analysis, we show that (QSIC) can be reduced to (QP1EQC) (see (ii) above), while (AQP) can be reduced to finding all the solutions to a KKT system.

2 Preliminary: A new S-procedure

Let $f(x) = x^T P x + 2p^T x + p_0$, $g(x) = x^T Q x + 2q^T x + q_0$ and $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ be defined as in \cite{1}. Given $\gamma \in \mathbb{R}$, we have the following new type of $S$-procedure.

\textbf{Theorem 1} Under conditions \cite{13} and \cite{14}, the following two statements are equivalent:

\begin{enumerate}
  \item[(G1)] $(\forall x \in \mathbb{R}^n, z \in \mathbb{R}^2) \ f(x) - z_1 = 0, g(x) - z_2 = 0, z_1 a + z_2 b \leq c \Rightarrow F(z) - \gamma \geq 0.$
  \item[(G2)] $(\exists \alpha, \beta, \mu \in \mathbb{R}, \mu \in \mathbb{R}^n)$ such that $F(z) - \gamma + \alpha (f(x) - z_1) + \beta (g(x) - z_2) + \mu^T (z_1 a + z_2 b - c) \geq 0, \ \forall (x, z) \in \mathbb{R}^n \times \mathbb{R}^2.$
\end{enumerate}

\textbf{Proof} Note that (G2) $\Rightarrow$ (G1) is trivial. We only prove that (G1) $\Rightarrow$ (G2). By condition \cite{13}, $C = D_1 = \{ z \in \mathbb{R}^2 | z_1 = f(x), z_2 = g(x), x \in \mathbb{R}^n \}$ is convex. Obviously, $D_2 = \{ z \in \mathbb{R}^2 | z_1 a + z_2 b \leq c \}$ is convex so that $D_1 \cap D_2$ is convex, too. By (G1),

$$F(z) - \gamma \geq 0 \ \forall z \in D_1 \cap D_2,$$

which implies that

$$D_1 \cap D_2 \cap D_3 = \emptyset,$$

where

$$D_3 := \{ z \in \mathbb{R}^2 | F(z) - \gamma < 0 \}.$$

By condition \cite{14}, $F$ is convex, so $D_3$ is an open convex set.
Let \( \{z : v^Tz = \bar{\gamma}\} \), with \( v = (\bar{\alpha}, \bar{\beta})^T \), separate \( D_1 \cap D_2 \) from \( D_3 \). Since \( D_3 \) is open, we assume, without loss the generality, that
\[
\bar{\alpha}_1 + \bar{\beta}_2 + \bar{\gamma} \geq 0, \quad \forall \ z \in D_1 \cap D_2, \tag{7}
\]
\[
\bar{\alpha}_1 + \bar{\beta}_2 + \bar{\gamma} < 0, \quad \forall \ z \in D_3. \tag{8}
\]

From (5), \( \bar{\alpha}_1 + \bar{\beta}_2 + \bar{\gamma} \geq 0 \Rightarrow F(z) - \gamma \geq 0 \). By S-lemma, there exists \( t \geq 0 \) such that
\[
F(z) - \gamma - t(\bar{\alpha}_1 + \bar{\beta}_2 + \bar{\gamma}) \geq 0, \quad \forall \ z \in \mathbb{R}^2. \tag{9}
\]

If \( t = 0 \), choose \( \alpha = \beta = 0, \mu = 0 \). Then, (G_2) holds.

If \( t > 0 \), by (9), the following system is unsolvable:
\[
t\bar{\alpha}_1 + t\bar{\beta}_2 + t\bar{\gamma} < 0,
\]
\[
z_1a + z_2b - c \leq 0, \quad z \in D_1.
\]

By the Farkas theorem (see [12, Theorem 21.1], [14, Section 6.10 21.1, [8, Theorem 2.1]), there exists \( \mu \in \mathbb{R}_+^m \) such that
\[
t\bar{\alpha}_1 + t\bar{\beta}_2 + t\bar{\gamma} + \mu^T(z_1a + z_2b - c) \geq 0, \quad \forall \ z \in D_1.
\]

Equivalently, there is \( \mu \geq 0 \) such that, \( \forall \ x \in \mathbb{R}^n \),
\[
t\bar{\alpha}f(x) + t\bar{\beta}g(x) + t\bar{\gamma} + \mu^T(f(x)a + g(x)b - c) \geq 0.
\]

Let
\[
\alpha = \mu^Ta + t\bar{\alpha}, \quad \beta = \mu^Tb + t\bar{\beta}. \tag{10}
\]

Then, \( \forall \ x \in \mathbb{R}^n \), one has:
\[
t\bar{\alpha}f(x) + t\bar{\beta}g(x) + t\bar{\gamma} + \mu^T(f(x)a + g(x)b - c) \geq 0
\]
\[
\Leftrightarrow (\mu^Ta + t\bar{\alpha})f(x) + (\mu^Tb + t\bar{\beta})g(x) + t\bar{\gamma} - \mu^Tc \geq 0
\]
\[
\Leftrightarrow \alpha f(x) + \beta g(x) + (\mu^Ta + t\bar{\alpha} - \alpha)z_1 + (\mu^Tb + t\bar{\beta} - \beta)z_2 + t\bar{\gamma} - \mu^Tc \geq 0 \quad \text{(by } (10))
\]
\[
\Leftrightarrow \alpha(f(x) - z_1) + \beta(g(x) - z_2) + \mu^T(z_1a + z_2b - c) \geq -t\bar{\alpha}z_1 - t\bar{\beta}z_2 - t\bar{\gamma}. \tag{11}
\]

Finally, we combine (6) with (11) to obtain (G_2).

As for condition (5), there is an easy-to-verify sufficient condition as follows.

**Theorem 2 (Theorem 4.16 in [2])** If \( \{P, Q\} \) are linearly independent, the joint numerical range \( C \) defined by (2) is a convex set in \( \mathbb{R}^2 \).

In the following, we give two examples to show that conditions (5) and (6) cannot be omitted from Theorem 1.

**Example 1** Let \( f(x) = x_1 + x_2, \ g(x) = 2x_1^2 - x_2^2, \ F(z) = 4z_1^2 + z_2 \) and \( a = b = c = \gamma = 0 \). Note that \( F(z) \) is convex and condition (5) holds. However, condition (6) is violated since the joint numerical range
\[
C = \{z \in \mathbb{R}^2 \mid z_2 \geq -2z_1^2\}
\]
is not convex. See [10, Example 3.1].
It can be verified that $z_1 = x_1 + x_2$ and $z_2 = 2x_1^2 - x_2^2$ imply that
\[ F(z) - 0 = 4(x_1 + x_2)^2 + (2x_1^2 - x_2^2) = 6x_1^2 + 8x_1x_2 + 3x_2^2 \geq 0. \]
Therefore, $\text{(G}_1\text{)}$ holds. On the other hand,
\[
F(z) - 0 + \alpha(f(x) - z_1) + \beta(g(x) - z_2) \\
= 4z_1^2 + z_2 + \alpha(x_1 + x_2 - z_1) + \beta(2x_1^2 - x_2^2 - z_2) \geq 0
\]
holds if and only if
\[
M = \begin{pmatrix}
2\beta & 0 & 0 & 0 & \alpha/2 \\
0 & -\beta & 0 & 0 & \alpha/2 \\
0 & 0 & 4 & 0 & -\alpha/2 \\
0 & 0 & 0 & 0 & (1 - \beta)/2 \\
\alpha/2 & \alpha/2 & -\alpha/2 & (1 - \beta)/2 & 0
\end{pmatrix} \succeq 0.
\]
To make $M \succeq 0$, we need $\beta = 0$. However, it leads to
\[
\begin{pmatrix}
1 & 0 \\
(1 - \beta)/2 & 0
\end{pmatrix} = \begin{pmatrix} 0 & 1/2 \end{pmatrix} \not\succeq 0
\]
and thus $\text{(G}_2\text{)}$ fails.

Example 2 Let $f(x) = x_1^2, g(x) = x_2^2$, and $F(z) = 2z_1z_2; a = b = c = \gamma = 0$. By Theorem 2, condition $\dagger$ holds. However, condition $\ddagger$ is clearly violated.

Note that $z_1 = x_1^2$ and $z_2 = x_2^2$ imply that $F(z) - 0 = x_1^2x_2^2 \geq 0$. Therefore $\text{(G}_1\text{)}$ holds. On the other hand,
\[
F(z) - 0 + \alpha(f(x) - z_1) + \beta(g(x) - z_2) \\
= 2z_1z_2 + \alpha(x_1^2 - z_1) + \beta(x_2^2 - z_2) \geq 0
\]
holds if and only if
\[
M = \begin{pmatrix}
\alpha & 0 & 0 & 0 & 0 \\
0 & \beta & 0 & 0 & 0 \\
0 & 0 & 1 & 1/2 & \alpha/2 \\
0 & 0 & 1 & 0 & \beta/2 \\
0 & 0 & -\alpha/2 & -\beta/2 & 0
\end{pmatrix} \succeq 0.
\]
However, this is impossible since \(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \not\succeq 0\). Then, $\text{(G}_2\text{)}$ fails.

3 Solving (Po4) under conditions $\dagger$ and $\ddagger$

3.1 Computing the optimal value $v(\text{Po4})$

Applying the new $S$-procedure in Theorem 1, we show that the optimal value $v(\text{Po4})$ under conditions $\dagger$ and $\ddagger$ can be obtained by solving an SDP.
The optimal value (cvx_optval): +43.7102.
A new type of quadratic optimization problem

Remark 1 The computational experiment was conducted in Matlab version R2016a running on a PC with Core i5 CPU and 8G memory. The SDP program (12) is modeled by CVX 1.21 [3] and solved by SDPT3 within CVX.

3.2 Solving for an optimal solution of (Po4)

In general, problem (Po4) under conditions (5) and (6) may not have an optimal solution. First, it can be unbounded from below. For example, let \( f(x) = x_1^2, g(x) = x_2^2, F(z) = z_1^2 - z_2, a = b = c = 0 \). Then both conditions (5) and (6) hold. It is easy to see that 
\[
v(\text{Po4}) = \inf_{x \in \mathbb{R}^2} \left\{ x_1^4 - x_2^2 \right\} = -\infty.
\]

Secondly, (Po4) may be bounded but not attainable. For example, let \( f(x) = x_1^2, \quad g(x) = x_1 x_2 - 1, \quad F(z) = z_1^2, \quad a = c = (0,0)^T, \quad b = (1, -1) \). Then (Po4) becomes
\[
\inf_{x \in \mathbb{R}^2} x_1^4 \\
\text{s.t.} \quad x_1 x_2 - 1 = 0.
\]
The optimal value is zero, but there exists no \( x \in \{ x | x_1 x_2 = 1 \} \) such that \( x_1^4 = 0 \).

Even if the optimal value \( v(\text{Po4}) \) is attained at some \( x^* \), we emphasize that it is generally not easy to solve \( x^* \) from the related SDP optimal solution. Ideally, we can use the following procedure for obtaining \( x^* \).

Algorithm 1 Solve an optimal solution \( x^* \) to (Po4) under conditions (5)-(6).

**Step 1:** Compute the optimal value \( v(\text{Po4}) \) by solving the (SDP) in (12).

**Step 2:** Solve \( z^* \in \mathbb{R}^2 \) that satisfies
\[
\begin{align*}
F(z^*) &= v(\text{Po4}), \\
z_1^2 a + z_2^2 b &\leq c, \\
z^* &\in C,
\end{align*}
\]
where \( C \) is the joint numerical range defined by (2).

**Step 3:** Solve \( x^* \in \mathbb{R}^n \) that satisfies
\[
\begin{align*}
f(x^*) &= z_1^*, \\
g(x^*) &= z_2^*.
\end{align*}
\]

Note that Step 3 is to solve a simultaneous system of two quadratic equations. Theoretically, it is known that a solution to a system of \( k \) quadratic equations, \( k \) fixed, can be found in polynomial time. See, e.g., the paper by Grigoriev and Pasechnik in 2005 [3]. However, their procedure appears to be difficult for implementation. In practice, we suggest to use the Newton method for finding the root \( x^* \) in (16).

Let us focus on Step 2 in Algorithm 1. Under condition (5), \( F(z) \) is a convex function. Then, \( \{ z \in \mathbb{R}^2 | F(z) = v(\text{Po4}) \} \) is either an ellipse, a parabola, or a pair of parallel straight lines in \( \mathbb{R}^2 \), whose forms are either \( F(z) = \theta_1 z_1^2 + \theta_2 z_2^2 \), \( F(z) = \theta_1 z_1^2 + \eta_2 z_2, \eta_2 \neq 0 \) or \( F(z) = \theta_1 z_1^2 \), respectively. Figure 1 depicts the case
Step 2.4: Set for F

Graphic representation for the subproblem (15). Fig. 1

\[ \theta_1 z_1^2 + \theta_3 z_2^2 = v(Po4) \]

for \( F(z) = \theta_1 z_1^2 + \theta_3 z_2^2 \) (in black dash), while the set (in stripes bounded by solid curves)

\[ \Omega_0 = C \cap \{ z \in \mathbb{R}^2 : z_1 a + z_2 b \leq c \} \]  \hspace{1cm} (17)

is also convex under condition (6). The intersection

\[ z^* = \Omega_0 \cap \{ z \in \mathbb{R}^2 : F(z) = v(Po4) \} \]

is the required solution for the subproblem (15). Below we propose a type of bisection method on \([0, 2\pi]\) for computing an approximate solution \( \tilde{z} \) of \( z^* \).

For simplicity, let us assume that \( F(z) = z_1^2 + z_2^2 \) for Algorithm 2 below. Given \( \epsilon > 0 \) and suppose in Step 1 of Algorithm 1 we obtain \( \bar{\epsilon} > 0 \) as an approximate value of \( v(Po4) \) by solving the SDP in (12) such that

\[ \bar{v} \in [v(Po4), v(Po4) + \epsilon/2]. \]  \hspace{1cm} (18)

Algorithm 2  Implementation of Step 2 in Algorithm 1, assuming \( F(z) = z_1^2 + z_2^2 \).

Step 2.1: Let \( \Omega_0 \) as in (17) and \( \epsilon, \bar{v} \) as in (18). Set \( l_0 = 0, u_0 = 2\pi, k = 0 \).

Step 2.2: Set \( \varphi_{k+1} : = \frac{l_k + u_k}{2} \) and

\[ \Omega_{k+1} : = \Omega_k \cap \{ z \in \mathbb{R}^2 : (\sin \varphi_{k+1}, -\cos \varphi_{k+1}) z \geq 0 \}. \]

Solve the following problem by formulating it as an SDP in the form of (12):

\[ v_{k+1} = \inf \{ F(z) : z \in \Omega_{k+1} \}. \]

Step 2.3: Test whether \( z^* \) belongs to \( \Omega_{k+1} : \)

- If \( v_{k+1} \leq \bar{v} \), set \( u_{k+1} := \varphi_{k+1}, l_{k+1} := l_k \).
- Otherwise, set \( u_{k+1} := u_k, l_{k+1} := \varphi_{k+1} \) and \( \Omega_{k+1} := \Omega_k \cap \{ z \in \mathbb{R}^2 : (\sin \varphi_{k+1}, -\cos \varphi_{k+1}) z \leq 0 \}. \)

Step 2.4: Set \( k : = k + 1 \). If \( u_k - l_k > \arccos \frac{\bar{v}}{\sqrt{\bar{v} + l_k/2}} \), go to Step 2.2. Otherwise, set \( \tilde{z} = (\sqrt{\bar{v}} \cos u_k, \sqrt{\bar{v}} \sin u_k) \), \( \tilde{z} = (\sqrt{\bar{v}} \sec \frac{\pi}{2\pi} \cos l_k, \sqrt{\bar{v}} \sec \frac{\pi}{2\pi} \sin l_k) \).

Step 2.5: Test whether \( [O, \tilde{z}] \cap \Omega_0 \neq \emptyset \) or \( [O, \tilde{z}] \cap \Omega_0 \neq \emptyset \) : \)

- if \( [O, \tilde{z}] \cap \Omega_0 \neq \emptyset \) (Fig. 2), find a point \( \tilde{z} \) in \( \{ z : \tilde{z}_1 z_1 - \tilde{z}_1 z_2 = 0 \} \cap \Omega_0 \) nearest to \( O \) (if \( [O, \tilde{z}] \cap \Omega_0 \neq \emptyset \), find a point \( \tilde{z} \) in \( \{ z : \tilde{z}_2 z_1 - \tilde{z}_2 z_2 = 0 \} \cap \Omega_0 \) nearest to \( O \)). Report \( \tilde{z} \) as an approximate solution of \( z^* \).
- Otherwise (Fig. 3): if \( [O, \tilde{z}] \cap \Omega_0 = \emptyset \) and \( [O, \tilde{z}] \cap \Omega_0 = \emptyset \), find a point \( \tilde{z} \) in \( \{ z : \tilde{z}_1 z_1 + \tilde{z}_2 z_2 = \bar{v} \} \cap \Omega_0 \) nearest to \( \tilde{z} \).
3.2.1 Analysis for Algorithm 2

At the first iteration of Algorithm 2, we use the midpoint \( \varphi_1 = \pi \in [0, 2\pi] \) to divide the feasible set \( \Omega_0 \) of (Po4) into two parts:

\[
\Omega_0 \cap \{ z \in \mathbb{R}^2 | z_2 \geq 0 \} \quad \text{and} \quad \Omega_0 \cap \{ z \in \mathbb{R}^2 | z_2 \leq 0 \}
\]

and then solve

\[
v_1 = \inf \{ F(z) | z \in \Omega_0, z_2 \geq 0 \},
\]

which is certainly a type of (Po4).

If \( v_1 \leq \bar{v} \), we set \( \Omega_1 = \Omega_0 \cap \{ z \in \mathbb{R}^2 | z_2 \geq 0 \} \) and then update \( u_1 = \varphi_1 = \pi \) with \( l_1 = l_0 \) unchanged. Otherwise, set \( \Omega_1 = \Omega_0 \cap \{ z \in \mathbb{R}^2 | z_2 \leq 0 \} \). Set \( l_1 = \varphi_1 = \pi \) whereas keeping \( u_1 = u_0 \) as the same. At the next iteration, \( \Omega_1 \) is further divided into two parts by either of the midpoints \( \varphi_2 = \pi/2 \in [0, \pi] \) or \( \varphi_2 = 3\pi/2 \in [\pi, 2\pi] \). This is equivalent to adding a new cut \( z_1 \geq 0 \) to \( \Omega_1 \). We solve

\[
\inf \left\{ F(z) | z \in \Omega_1 \cap \{ z \in \mathbb{R}^2 | z_1 \geq 0 \} \right\}
\]  (19)
to determine whether \((z_1^*, z_2^*)^T\) lies in \(\Omega_1 \cap \{z \in \mathbb{R}^2 \mid z_1 \geq 0\}\) or not. If affirmative, we set \(\Omega_2 = \Omega_1 \cap \{z \in \mathbb{R}^2 \mid z_1 \geq 0\}\). Otherwise, set \(\Omega_2 = \Omega_1 \cap \{z \in \mathbb{R}^2 \mid z_1 \leq 0\}\). Update the search interval \([l_k, u_k]\) accordingly. At the third iteration, the set \(\Omega_2\) is divided into two parts by the line \((\sin(\varphi_3), -\cos(\varphi_3))z = 0\) where the midpoint \(\varphi_3\) is either \(\pi/4\), or \(3\pi/4\), or \(5\pi/4\), or \(7\pi/4\).

In general, at the \(k^{th}\) iteration, Algorithm \(\mathcal{A}\) divides \([0, 2\pi]\) into \(2^k\) parts. Moreover, an optimal solution \(z^*\) at step 2 of Algorithm \(\mathcal{A}\) is located in \(\Omega_k\), which is a very narrow sector of arc angle \(2\pi/2^k\) intersecting with \(\Omega_0\). The stopping criterion requires the angle to be smaller than a preset constant, say \(2\pi/2^k \leq \arccos \frac{\pi}{\sqrt{v + \varepsilon/2}}\).

Therefore the number of steps necessarily to achieve this is

\[
k^*(v, \varepsilon) = \left\lceil \log_2 \frac{2\pi}{\arccos \frac{\sqrt{v}}{\sqrt{v + \varepsilon/2}}} \right\rceil + 1,
\]

where \([t]\) is the integer part of the real number \(t\).

When the number of the bisection steps \(k\) arrives \(k^*(v, \varepsilon)\), we choose the following two points from the boundary of the sector (see Step 2.4):

\[
\hat{z} = (\sqrt{v} \cos u_k, \sqrt{v} \sin u_k) \in \{z \in \mathbb{R}^2 \mid F(z) = z_1^2 + z_2^2 = \bar{v}\},
\]

\[
\hat{z} = (\sqrt{v} \sec (2\pi/2^k) \cos u_k, \sqrt{v} \sec (2\pi/2^k) \sin u_k).
\]

Notice that both points \(\hat{z}, \hat{\bar{z}}\) lie on the line \(\{z : \hat{z}_1 z_1 + \hat{z}_2 z_2 = \bar{v}\}\), which is tangent to the circle \(\{z \in \mathbb{R}^2 \mid F(z) = z_1^2 + z_2^2 = \bar{v}\}\) at \(\hat{\bar{z}}\). Moreover, for all \(\xi \in [\hat{z}, \hat{\bar{z}}]\),

\[
\sqrt{v} = \|\hat{z}\| \leq \|F(\xi)\| \leq \|\hat{\bar{z}}\| = \sqrt{v} \sec \frac{2\pi}{2^k} \leq \sqrt{v} \cdot \frac{\sqrt{v} + \varepsilon/2}{\sqrt{ar{v}}} \leq \sqrt{v \cdot (Po4) + \varepsilon}.
\]

That is,

\[
F(\xi) \leq v(Po4) + \varepsilon, \quad \xi \in [\hat{z}, \hat{\bar{z}}].
\]

Then, depending on the configuration of \(\Omega_0\), the two segments \([O \hat{z}]\) and \([O \hat{\bar{z}}]\) may (Fig. 2) or may not (Fig. 3) intersect \(\Omega_0\).

- If either \([O, \hat{z}] \cap \Omega_0 \neq \emptyset\) or \([O, \hat{\bar{z}}] \cap \Omega_0 \neq \emptyset\), say \([O, \hat{z}] \cap \Omega_0 \neq \emptyset\), Algorithm \(\mathcal{A}\) finds \(\hat{z}\) in \(\{z : \hat{z}_1 z_1 - \hat{z}_2 z_2 = 0\}\) and \(\Omega_0\) nearest to \(O\). See Fig. 2. Since the line \(\{z : \hat{z}_1 z_1 - \hat{z}_2 z_2 = 0\}\) passes both \(O\) and \(\hat{z}\), there is

\[
v(Po4) = F(z^*) \leq F(\hat{z}) \leq F(\bar{z}) = \bar{v} \leq v(Po4) + \frac{\varepsilon}{2}.
\]

Similarly, if \([O, \hat{\bar{z}}] \cap \Omega_0 \neq \emptyset\), Algorithm \(\mathcal{A}\) finds \(\bar{z}\) such that, by (21),

\[
v(Po4) = F(z^*) \leq F(\hat{\bar{z}}) \leq F(\bar{\bar{z}}) \leq v(Po4) + \varepsilon.
\]

In any case, we have

\[
v(Po4) \leq F(\bar{z}) \leq v(Po4) + \varepsilon.
\]

- On the other hand, if both \([O, \hat{z}] \cap \Omega_0\) and \([O, \hat{\bar{z}}] \cap \Omega_0\) are empty, Algorithm \(\mathcal{A}\) requires to find a point \(\hat{z}\) in \(\{z : \hat{z}_1 z_1 + \hat{z}_2 z_2 = \bar{v}\} \cap \Omega_0\) nearest to \(\hat{\bar{z}}\) (see Fig. 3). The following proposition shows that, when \(\Omega_0\) is either unbounded or contains an interior point, such a nearest point \(\hat{z}\) always exists.
Proposition 1 Let \( \tilde{z}, \hat{z} \) be defined as in Step 2.4 of Algorithm \([2]\). Suppose, at Step 2.5 of Algorithm \([2]\) there happen that \([O, \tilde{z}] \cap \Omega_0 = \emptyset \) and \([O, \hat{z}] \cap \Omega_0 = \emptyset \).

- If \( \Omega_0 \) is unbounded, then \( [\tilde{z}, \hat{z}] \cap \Omega_0 \neq \emptyset \).
- If \( \Omega_0 \) contains an interior point, then \( [\tilde{z}, \hat{z}] \cap \Omega_0 \neq \emptyset \) for all sufficiently small \( \epsilon > 0 \) where \( \epsilon \) is defined as in \([15]\).

Proof In considering the triangle \( \Delta(O\tilde{z}\hat{z}) \), one has \( z^* \in \Omega_0 \cap \Delta(O\tilde{z}\hat{z}) \) by Algorithm \([2]\). Since \( \Omega_0 \) is convex and unbounded and the two edges \([O, \tilde{z}], [O, \hat{z}] \) of \( \Delta(O\tilde{z}\hat{z}) \) do not intersect with \( \Omega_0 \), the edge \([\tilde{z}, \hat{z}] \) must intersect with \( \Omega_0 \). Namely, \( [\tilde{z}, \hat{z}] \cap \Omega_0 \neq \emptyset \).

Now, if \( \Omega_0 \) contains an interior point \( y' \in \Omega_0 \) such that the ball \( B(y', r) \subset \Omega_0 \) for some \( r > 0 \). Let \( 0 < \epsilon < r^2 \) with \( \epsilon \) defined as in \([15]\). From \([21]\), we know \( \|\tilde{z}\|^2 \leq v'(Po4) + \epsilon \) and \( \|\hat{z}\|^2 \leq v'(Po4) + \epsilon \) so that the triangle \( \Delta(O\tilde{z}\hat{z}) \) is contained in the ball of center \( O \) with radius \( \sqrt{v'(Po4) + \epsilon} \leq \sqrt{v'(Po4) + r} < \sqrt{v'(Po4)} + r \). On the other hand, for each \( y \in B(y', r) \subset \Omega_0 \), one has \( \|y\| \geq \|z^*\| = \sqrt{v'(Po4)} \), which implies that \( \|y\|^2 \geq \sqrt{v'(Po4)} + r \). Hence, \( y' \in \Omega_0 \) lies outside the triangle \( \Delta(O\tilde{z}\hat{z}) \) whereas \( z^* \in \Delta(O\tilde{z}\hat{z}) \). Due to the convexity of \( \Omega_0 \), the segment \([z^*, y'] \subset \Omega_0 \).

However, our assumption assumes that \([O, \tilde{z}] \cap \Omega_0 = \emptyset \) and \([O, \hat{z}] \cap \Omega_0 = \emptyset \) so that \([\tilde{z}, \hat{z}] \cap [z^*, y'] \neq \emptyset \). The result of this proposition follows immediately. \( \square \)

Hence, with mild assumptions, see Proposition \([11]\), \( \hat{z} \) always exists and \( \hat{z} \) belongs to the segment \([\tilde{z}, \hat{z}] \), so by \([21]\) we have \( F(\hat{z}) \leq v'(Po4) + \epsilon \).

In summary, we have the following result.

Theorem 4 Suppose \( F(z) = z_1^2 + z_2^2 \) and that Step 1 in Algorithm \([2]\) returns an approximate optimal value \( \bar{v} \in [v(Po4), v(Po4) + \epsilon/2] \) for a sufficiently small \( \epsilon > 0 \). Assume that the feasible domain of problem \([Po4] \), \( \Omega_0 \), is either unbounded or contains an interior point. Then, Algorithm \([2]\) terminates after \( k^*(\bar{v}, \epsilon) \) (defined as in \([20]\)) number of bisection steps with an approximate solution \( \tilde{z} \in \Omega_0 \) such that \( v(Po4) \leq F(\tilde{z}) \leq v(Po4) + \epsilon \).

3.2.2 Discussion on Step 2.5 of Algorithm \([2]\)

To determine whether \( \tilde{z} = (\sqrt{v} \cos u_k, \sqrt{v} \sin u_k) \) (the same for \( \hat{z} \)) obtained in Step 2.4 of Algorithm \([2]\) satisfies the condition \([O\tilde{z}] \cap \Omega_0 \neq \emptyset \), we observe that (see Fig. 2 and Fig. 3)

\[
(O\tilde{z}) \cap \Omega_0 = \emptyset \iff \inf \{\tilde{z}_1 z_1 + \tilde{z}_2 z_2 : \tilde{z}_1 z_1 - \tilde{z}_2 z_2 = 0, \ z \in \Omega_0\} > \bar{v} \\
\iff \bar{v} < \inf_{x \in \mathbb{R}^n} \begin{cases} \inf_{x \in \mathbb{R}^n} \tilde{z}_1 f(x) + \tilde{z}_2 g(x) \\ \text{s.t. } f(x) a + g(x) b - c \leq 0; \\ \tilde{z}_2 f(x) - \tilde{z}_1 g(x) = 0. \end{cases} \tag{25}
\]

Problem \([25]\) is again a type of \([Po4] \) so that whether or not \( \tilde{z} \in \Omega_0 \) can be checked by computing the optimal value \( v'(25) \) of a related SDP as specified in Theorem \([4]\). However, if \( \tilde{z} \notin \Omega_0 \), finding a point \( \tilde{z} \) in \( \{z : \tilde{z}_1 z_1 + \tilde{z}_2 z_2 = \bar{v}\} \cap \Omega_0 \) nearest to \( \tilde{z} \) is generally difficult. It amounts to solving a point \( \bar{x} \in \mathbb{R}^n \) such that \( \bar{z}_1 = f(\bar{x}), \bar{z}_2 = g(\bar{x}) \) and \( \bar{z} = (\bar{z}_1, \bar{z}_2) \) is a minimum solution to

\[
\inf \{\tilde{z}_1 z_1 - \tilde{z}_2 z_2 + \tilde{z}_1 z_1 + \tilde{z}_2 z_2 : \tilde{z}_1 z_1 + \tilde{z}_2 z_2 = \bar{v}, \ z \in \Omega_0\}. \tag{26}
\]
Nevertheless, there is one special case of (26) which we know how to circumvent the difficulty. When \( a = b = c = 0 \), (26) becomes
\[
\inf \{ \tilde{z}_2 f(x) - \tilde{z}_1 g(x) : \tilde{z}_1 f(x) + \tilde{z}_2 g(x) = \tilde{v}, \ x \in \mathbb{R}^n \}.
\] (27)
It is a (QP1EQC). See Introduction. By applying the S-lemma with equality (24) and the standard rank-one decomposition procedure [15], we can obtain an optimal solution \( \bar{x} \) to (27) such that \( \bar{x} = (f(\bar{x}), g(\bar{x})) \in \Omega_0 = \mathbb{C} \) is nearest to \( \tilde{z} \). According to Theorem 4, there indeed is \( v(\text{Po4}) \leq F(f(\bar{x}), g(\bar{x})) \leq v(\text{Po4}) + \epsilon \).

Notice that, in this special case \( a = b = c = 0 \), Step 3 in Algorithm 1 for solving a simultaneous quadratic system is not necessary. In the following section, we shall see that the special case suffices to solve the quadratic surfaces intersection problem (QSIC). We thus summarize the discussion into the following theorem.

**Theorem 5** Let \( a = b = c = 0 \); \( \tilde{v} \) be an approximate value of \( v(\text{Po4}) \) as specified in (18) and \( \tilde{z} \) be obtained in Step 2.4 of Algorithm 3.

- If \( [O] \cap \Omega_0 = \emptyset \) and \( [O] \cap \Omega_0 = \emptyset \), we solve \( \bar{x} \in \text{argmin}\{ \tilde{z}_2 f(x) - \tilde{z}_1 g(x) : \tilde{z}_1 f(x) + \tilde{z}_2 g(x) = \tilde{v} \}; \)

- If \( [O] \cap \Omega_0 \neq \emptyset \), we solve \( \bar{x} \in \text{argmin}\{ \tilde{z}_1 f(x) + \tilde{z}_2 g(x) : \tilde{z}_2 f(x) - \tilde{z}_1 g(x) = 0 \}; \)

- If \( [O] \cap \Omega_0 \neq \emptyset \), we solve \( \bar{x} \in \text{argmin}\{ \tilde{z}_1 f(x) + \tilde{z}_2 g(x) : \tilde{z}_2 f(x) - \tilde{z}_1 g(x) = 0 \}; \)

where \( [O] \cap \Omega_0 = \emptyset \iff \tilde{v} \in \text{inf}\{ \tilde{z}_1 f(x) + \tilde{z}_2 g(x) : \tilde{z}_2 f(x) - \tilde{z}_1 g(x) = 0, \ x \in \mathbb{R}^n \}. \)

Then, \( \bar{x} \) is an approximate optimal solution to (Po4) with \( F(z) = z_1^2 + z_2^2 \) such that \( v(\text{Po4}) \leq F(f(\bar{x}), g(\bar{x})) \leq v(\text{Po4}) + \epsilon \).

### 4 Applications

#### 4.1 Problem (QSIC)

The problem (QSIC) defined in (11) can be recast in the following form:

\[
\text{(QSIC)} \quad \inf_{x \in \mathbb{R}^n} (x^T P x + p^T x + p_0)^2 + (x^T Q x + q^T x + q_0)^2.
\]

It is to determine whether or not the two hypersurfaces, \( f(x) = x^T P x + p^T x + p_0 = 0 \) and \( g(x) = x^T Q x + q^T x + q_0 = 0 \), intersect. If \( v(\text{QSIC}) = 0 \), or \( v(\text{QSIC}) < \rho \) for some user-defined small \( \rho > 0 \) to accommodate possible noises arising from the problem data, the two hypersurfaces are thought to intersect with each other. Otherwise, \( v(\text{QSIC}) \) gives some sort of measurement as to how far the two hypersurfaces deviate from each other. For this purpose, the optimal value \( v(\text{QSIC}) \) suffices to decide the intersection problem. However, problem (QSIC) belongs to the special case that \( F(z) = z_1^2 + z_2^2 \) and \( a = b = c = 0 \). By Theorem 5, we are able to find an approximate solution \( \bar{x} \) which is either a common root of the two quadratic equations \( f(x) = g(x) = 0 \); or can be viewed as a point whose total distance to the two hypersurfaces \( f(x) = 0 \) and \( g(x) = 0 \) is the least.

We solve (QSIC) by two cases.

**Case 1**: If \( \{P, Q\} \) are linearly independent, by Theorem 2 and by taking \( F(z) = z_1^2 + z_2^2, a = b = c = 0 \), we see that (QSIC) is a (Po4) satisfying conditions (9) - (10). In this case, the optimal value \( v(\text{QSIC}) \) can be computed by the following (SDP):
\[
\gamma^* = \sup_{\gamma, \alpha, \beta \in \mathbb{R}} \left\{ \gamma \mid M \succeq 0 \right\},
\]
where $M$ is
\[
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & \alpha P + \beta Q & \alpha p + \beta q \\
0 & 0 & \alpha P^T + \beta q^T & \alpha p_0 + \beta q_0 - \gamma
\end{pmatrix}.
\]

The optimal solution of (QSIC), if attainable, can be computed approximately via
Algorithm 1, Algorithm 2 and Theorem 5.

**Case 2**: If \{P, Q\} are linearly dependent, say $Q = t^* P$. If $P = Q = 0$, then (QSIC) is an unconstrained convex quadratic optimization problem, which can be solved directly. Hence, let us assume that $P \neq 0$. Then, we multiply the first equation in (4) by $t^*$ and subtract it from the second equation to obtain
\[
\begin{align*}
\begin{bmatrix}
 x^T P x + p^T x + p_0 - z_1 = 0, \\
 (q^T - t^* p^T) x + (q_0 - t^* p_0) + t^* z_1 - z_2 = 0.
\end{bmatrix}
\end{align*}
\]

Define $y^T = [x^T, z^T]$; $h^T = [q^T - t^* p^T, t^*, -1]$; $h_0 = q_0 - t^* p_0$; and
\[
\bar{A} = \begin{pmatrix}
(0)_{n \times n} & (0)_{n \times 2} \\
(0)_{2 \times n} & I_{2 \times 2}
\end{pmatrix}, \quad \bar{P} = \begin{pmatrix}
P & (0)_{n \times 2} \\
(0)_{2 \times n} & (0)_{2 \times 2}
\end{pmatrix} \neq 0.
\]

Then, (QSIC) becomes
\[
\inf_{y \in \mathbb{R}^{n+2}} y^T \bar{A} y \\
\text{s.t.} \quad y^T \bar{P} y + \bar{p}^T y + p_0 = 0, \quad h^T y + h_0 = 0,
\]

where $\bar{p}^T = [p^T, -1, 0]$. By the null space representation, the solution set for the hyperplane $h^T y + h_0 = 0$ can be written as $y = y_0 + V z$ for some $z \in \mathbb{R}^{n+1}$, where $y_0 = -\frac{h_0}{h^T h}$ and $V \in \mathbb{R}^{(n+2) \times (n+1)}$ is the matrix basis of $\mathcal{N}(h)$. It implies that
(QSIC) is reduced to the following (QP1EQC) problem:
\[
\inf_{z \in \mathbb{R}^{n+1}} (y_0 + V z)^T \bar{A} (y_0 + V z) \\
\text{s.t.} \quad (y_0 + V z)^T P (y_0 + V z) + \bar{p}^T (y_0 + V z) + p_0 = 0,
\]

which can be solved by applying the S-lemma with equality. See [21].

### 4.2 Problem (AQP)

Notice that (AQP) can be equivalently formulated as follows:
\[
\inf_{(x, z_1) \in \mathbb{R}^n \times \mathbb{R}} z_1^2 \\
\text{s.t.} \quad \begin{cases}
x^T P x + p^T x + p_0 - z_1 = 0, \\
x^T Q x + q^T x + q_0 \leq 0.
\end{cases}
\]

We again split the discussion into two cases: Case 1: \{P, Q\} are linearly independent; Case 2: \{P, Q\} are linearly dependent.
• Case 1: If \( \{P, Q\} \) are linearly independent, by writing (30) as

\[
\inf_{(x, z_1) \in \mathbb{R}^n \times \mathbb{R}} z_1^2 \\
\quad \text{s.t. } \begin{cases}
  z_2 \leq 0, \\
  f(x) = x^T P x + p^T x + p_0 = z_1, \\
  g(x) = x^T Q x + q^T x + q_0 = z_2,
\end{cases}
\]  

(31)

with \( F(z) = z_1^2, a = 0, b = 1, c = 0 \), we see that (AQP) is a (Po4) satisfying conditions (5)-(6). Then, the optimal value \( v((31)) \) can be computed by:

\[
\gamma^* = v(\text{AQP}) = \sup_{\gamma, \alpha, \beta \in \mathbb{R}} \{ \gamma \mid M \succeq 0 \},
\]

where \( M \) is

\[
\begin{pmatrix}
  1 & 0 & [0] & \mu - \beta \\
  0 & 0 & \alpha P + \beta Q & \alpha p + \beta q \\
  -\alpha & 0 & 0 & \alpha p_0 + \beta q_0 - \gamma
\end{pmatrix}
\]

Although the optimal solution can be computed approximately via Algorithm 1 and Algorithm 2, yet due to \( F(z) = z_1^2 \) and \( z_2 \leq 0 \) being simple functions, we can solve it in a better way without resorting to the bisection method.

- First, if \( \gamma^* = v(\text{AQP}) = 0 \), then any optimal solution to the following (QP1EQC) is an optimal solution to (31).

(QP1EQC) \( \inf_{x \in \mathbb{R}^n} g(x) \) s.t. \( f(x) = 0. \)

The solution of (QP1EQC) can be obtained by solving an SDP and a standard matrix rank one decomposition. See [21,15].

- Secondly, if \( \gamma^* = v(\text{AQP}) > 0 \), then one has \( \{ z \in \mathbb{R}^2 \mid z_1^2 < \gamma^* \} \neq \emptyset \) and

\[
\{ z \in \mathbb{R}^2 : -\sqrt{\gamma^*} < z_1 < \sqrt{\gamma^*} \cap (C \cap \{ z \in \mathbb{R}^2 \mid z_2 \leq 0 \}) \} = \emptyset.
\]
Since $C \cap \{ z \in \mathbb{R}^2 : z_2 \leq 0 \}$ is convex, there must be either
\[
(C \cap \{ z \in \mathbb{R}^2 | z_2 \leq 0 \}) \subset \{ z \in \mathbb{R}^2 | z_1 \geq \sqrt{\gamma} \};
\]
or
\[
(C \cap \{ z \in \mathbb{R}^2 | z_2 \leq 0 \}) \subset \{ z \in \mathbb{R}^2 | z_1 \leq -\sqrt{\gamma} \}.
\]
In the former case, (AQP) is solved by $\inf \{ f(x) : g(x) \leq 0 \}$ (see Fig. 4 and 4), while the latter case can be solved by $\inf \{ -f(x) : g(x) \leq 0 \}$. Both problems are (QP1QC).

- Case 2: If $\{ P, Q \}$ are linearly dependent such that $Q = t^* P$, we assume that there exists a minimizer where KTCQ is satisfied. Now, we can use a linear combination to eliminate the matrix $Q$ in (31) and get
\[
\inf_{(x,z_1) \in \mathbb{R}^n \times \mathbb{R}} z_1^2
\]
\[
\text{s.t.} \quad \begin{cases}
-2z_1 = -\lambda_1 + \lambda_2 t^*; \\
0 = \lambda_1 (2Px + p) + \lambda_2 (qT - t^* pT); \\
x^T P x + p^T x + p_0 - z_1 = 0; \\
(q^T - t^* p^T) x + (q_0 - t^* p_0) + t^* z_1 = 0; \\
\lambda_2 (q^T - t^* p^T) x + (q_0 - t^* p_0) + t^* z_1 = 0; \\
\lambda_2 \geq 0.
\end{cases}
\]

- Case 1: $\lambda_2 > 0$. By the complementary slackness, $(q^T - t^* p^T) x + (q_0 - t^* p_0) + t^* z_1 = 0$. That is, the second inequality constraint in (32) is active. Therefore, to check whether there are KKT candidates $(z_1^*, x^*)$ associated with $\lambda_2 > 0$ satisfying (33), we can instead find optimal solutions $(z_1^*, x^*)$ to the following problem
\[
\inf_{(x,z_1) \in \mathbb{R}^n \times \mathbb{R}} z_1^2
\]
\[
\text{s.t.} \quad \begin{cases}
-2z_1 = -\lambda_1; \\
0 = \lambda_1 (2Px + p); \\
x^T P x + p^T x + p_0 - z_1 = 0; \\
(q^T - t^* p^T) x + (q_0 - t^* p_0) + t^* z_1 = 0; \\
\lambda_2 (q^T - t^* p^T) x + (q_0 - t^* p_0) + t^* z_1 = 0; \\
\lambda_2 \geq 0.
\end{cases}
\]

- Case 2: $\lambda_2 = 0$, then (33) becomes
\[
\begin{cases}
-2z_1 = -\lambda_1; \\
0 = \lambda_1 (2Px + p); \\
x^T P x + p^T x + p_0 - z_1 = 0; \\
(q^T - t^* p^T) x + (q_0 - t^* p_0) + t^* z_1 \leq 0.
\end{cases}
\]

We now further divide this case to two sub-cases: $\lambda_1 = 0$ and $\lambda_1 \neq 0$. 

With \( \lambda_1 = \lambda_2 = 0 \), we have \( z_1 = 0 \) and \((35)\) becomes
\[
\begin{align*}
& x^T P x + p^T x + p_0 = z_1 = 0; \\
& -(q^T - t^* p^T) x + (q_0 - t^* p_0) \leq 0.
\end{align*}
\]
(36)

Since the objective function is \( z_1^2 \), any solution of \((36)\) is also a solution of \((32)\). It can be decided by solving the following (QP1EQC):
\[
\begin{align*}
\inf_{x \in \mathbb{R}^n} & - (q^T - t^* p^T) x + (q_0 - t^* p_0) \\
\text{s.t.} & \quad x^T P x + p^T x + p_0 = 0.
\end{align*}
\]
(37)

If the optimal value \( v((37)) > 0 \), then \((36)\) does not have a solution. In this case, \( \lambda_1 \neq 0, \lambda_2 = 0 \) such that \((35)\) becomes
\[
\begin{align*}
\begin{cases}
2z_1 = \lambda_1 \neq 0; \\
0 = \lambda_1 (2 P x + p); \\
x^T P x + p^T x + p_0 - z_1 = 0; \\
(q^T - t^* p^T) x + (q_0 - t^* p_0) + t^* z_1 \leq 0.
\end{cases}
\end{align*}
\]
(38)

Then, \( 2 P x + p = 0 \) and \( z_1 = x^T P x + p^T x + p_0 = \frac{1}{2} p^T x + p_0 \).

Hence, \((38)\) is reduced to the following linear system in variables \((x, z_1, \lambda_1) \in \mathbb{R}^{n+2}\):
\[
\begin{align*}
\begin{cases}
2z_1 = \lambda_1 \neq 0; \\
2 P x + p = 0; \\
\frac{1}{2} p^T x + p_0 - z_1 = 0; \\
(q^T - t^* p^T) x + (q_0 - t^* p_0) + t^* z_1 \leq 0.
\end{cases}
\end{align*}
\]
(39)

In summary, when \( \{P, Q\} \) are linearly dependent, we can solve (i) \( \lambda_2 > 0 \) and \((34)\); (ii) \( \lambda_1 = \lambda_2 = 0 \) and \((37)\) with \( v((37)) \leq 0 \); and (iii) \( \lambda_1 \neq 0, \lambda_2 = 0 \) and the linear system \((39)\). The optimal solution of \((AQP)\) will be the valid ones among (i), (ii) and (iii) with the smallest objective value \( z_1^2 \) of \((AQP)\).

**Example 4**

\[
\begin{align*}
\inf_{x \in \mathbb{R}^2} & \quad - x_1^2 + x_2^2 + x_1 \\
\text{s.t.} & \quad - x_1^2 + x_2^2 + 1 \leq 0.
\end{align*}
\]
(40)

In this example, \( \{P, Q\} \) are linearly dependent with \( t^* = 1 \). Using the linear combination with \( t^* = 1 \) to eliminate \( Q \), we can write
\[
\begin{align*}
\inf_{(x, z_1) \in \mathbb{R}^2 \times \mathbb{R}} & \quad z_1^2 \\
\text{s.t.} & \quad - x_1^2 + x_2^2 + x_1 - z_1 = 0, \\
& \quad - x_1 + z_1 + 1 \leq 0.
\end{align*}
\]
(41)

Since we are going to utilize the KKT system of \((41)\), we first check the constraint qualification. Notice that,
\[
\nabla(- x_1^2 + x_2^2 + x_1 - z_1) = (-2x_1 + 1, 2x_2, -1)^T, \quad \nabla(- x_1 + z_1 + 1) = (-1, 0, 1)^T.
\]
The two gradients are linearly dependent if and only if $x_1 = x_2 = 0$. When $x_1 = x_2 = 0$, (11) has an empty feasible set $\{z_1 = 0, z_1 \leq -1\} = \emptyset$. In other words, every optimal solution to problem (10) must be a regular point of the KKT system.

Then, check all the following three cases:

(i) $\lambda_2 > 0$. Then, the inequality $-x_1 + z_1 + 1 \leq 0$ is active so that $x_1 = z_1 + 1$. We first solve the $(QP1EQC)$:

$$\inf_{(z_1, x_2) \in \mathbb{R} \times \mathbb{R}} z_1^2$$

st.

$$-(z_1 + 1)^2 + x_2^2 + 1 = 0$$

(42)

To obtain the solution $z_1^* = 0, x_2^* = 1, x_2^* = 0$ to (42). From the first equation $-2z_1 = -\lambda_1 + \lambda_2x_2^*$ in the KKT system (33), since $z_1^* = 0$, $t^* = 1$, we obtain $\lambda_1 = \lambda_2 > 0$.

From the second equation of (33), $(x_1^*, x_2^*)^T$ must satisfy

$$0 = (2Px + p) + (q^T - t^*p^T)$$

which is not true as $(2Px^* + p) + (q^T - t^*p^T) = (-3, 0)^T$. We conclude that there is no KKT point of (12) with $\lambda_2 > 0$.

(ii) $\lambda_1 = \lambda_2 = 0$. In this case, the KKT system (33) is reduced to (36), which is

$$\left\{ \begin{array}{l} z_1 = 0; \\ -x_1 + z_1 + 1 \leq 0. \end{array} \right.$$ (43)

Since (33) guarantees that $z_1 = 0$, any solution to (13), if exists, is an optimal solution to (AQP). It is not difficult to see that (13) has infinitely many solutions among which we can choose, for example, $z_1^* = 0, x_1^* = 1, x_2^* = 0$.

(iii) $\lambda_1 \neq 0, \lambda_2 = 0$. Dealing with this case might help to find other solutions to (10). In this case, we need to check the linear system in (39), which is:

$$\left\{ \begin{array}{l} 2z_1 = \lambda_1 \neq 0; \\ (-2x_1, 2x_2) + (1, 0) = (0, 0); \\ z_1 = \frac{z_2}{2}; \\ -x_1 + z_1 + 1 \leq 0. \end{array} \right.$$ (44)

From the second and the third equation of (14), we get $x_2 = 0.5, x_2 = 0, z_1 = 0.25$. They do not satisfy $-x_1 + z_1 + 1 \leq 0$ though. The system (14) does not have a solution, either.

5 Conclusion and Discussion

In this paper, we propose a new type of quadratic optimization problems involving a joint numerical range constraint. There are many natural applications arising from such a formulation. Some applications like the double well potential problem (DWP) [1,20] were already solved independently, while others like (QSIC) and (AQP) can now be resolved through our new approach. Interestingly, due to the composition of “quadratic with quadratics,” the objective function $F(f(x), g(x))$ is
indeed a polynomial of degree 4. We hope that our method can be later extended
to solve optimization problems involving quartic polynomials.

Notice that, in Theorem 1, we can replace \( \{f, g\} \) by \( k \) quadratic functions \( \{f_1, \ldots, f_k\} \). The theoretical results in this paper are still valid (the implementation can be more complex though), provided the following conditions are satisfied:

\[
\{(f_1(x), \ldots, f_k(x)) | x \in \mathbb{R}^n\} \text{ is convex,} \\
\Theta_{k \times k} \succeq 0,
\]

where \( \Theta_{k \times k} \) defines a quadratic function \( f(z) = z^T \Theta_{k \times k} z + \eta^T z \) of \( k \) variables. It suggests that the convexity of the joint numerical range of \( \{f_1, \ldots, f_k\} \) is the central feature for many optimization problems and should be studied carefully in the future.

**Acknowledgements**

Huu-Quang, Nguyen’s research work was sponsored partially by Taiwan MOST 107-2811-M-006-535 and Ruey-Lin Sheu’s research work was sponsored partially by Taiwan MOST 107-2115-M-006-011-MY2.

Xia’s research was supported by National Natural Science Foundation of China under grants 11822103, 11571029, 11771056, and Beijing Natural Science Foundation Z180005.

**References**

1. Fang, S.C. and Gao, D.Y. and Lin, G.X. and Sheu, R.L. and Xing, W., 2017. Double well potential function and its optimization in the n-dimensional real space- Part 1. Journal of Industrial and Management Optimization, 13, pp. 1291-1305.
2. Flores-Bazán, F. and Opazo, F., 2016. Characterizing the convexity of joint-range for a pair of inhomogeneous quadratic functions and strong duality. Minimax Theory and its Applications, 1, pp. 257-290.
3. Grant, M. and Boyd, S., 2010. CVX: Matlab software for disciplined convex programming. Version 1. 21 Web.
4. Derinkuyu, K. and Pinar, M.C., 2006. On the S-procedure and some variants. Mathematical Methods of Operations Research, 64, pp. 55-77.
5. Grigoriev, D. and Pasechnik, D.V., 2005. Polynomial-time computing over quadratic maps i: sampling in real algebraic sets. Computational complexity, 14(1), pp. 20-52.
6. Levin, J., 1979. Mathematical models for determining the intersections of quadratic surfaces. Computer Graphics and Image Processing, 11, pp. 73-87.
7. Moré, J.J., 1993. Generalizations of the trust region problem. Optimization methods and Software, 2(3-4), pp. 189-209.
8. Nguyen H.Q. and Sheu, R. L., 2020. Separation properties of quadratic functions. Available from: https://doi.org/10.13140/RG.2.2.18518.88647.
9. Polik, I. and Terlaky, T., 2006. A survey of the S-lemma. SIAM Review, 49, pp. 371-418.
10. Polyak, B. T., 1998. Convexity of quadratic transformations and its use in control and optimization. Journal of Optimization Theory and Applications, 99, pp. 553-583.
11. Pong, TK. and Wolkowicz, H., 2014. The generalized trust region subproblem. Journal of Optimization Theory and Applications, 58, pp. 273-322.
12. Rockafellar, R. T., 1970. Conex analysis. Princeton University Press.
13. Stern, R. and Wolkowicz, H., 1995. Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations. SIAM Journal on Optimization, 5, pp. 286-313.
14. Stoer, J. and Witzgall, C., 1970. *Convexity and Optimization in Finite Dimensions.* Springer-Verlag, Heidelberg.

15. Sturm, J. F., Zhang, S., 2003. "On cones of nonnegative quadratic functions." Mathematics of Operations Research, 28(2), pp. 246–267.

16. Wang, S. and Xia, Y., 2015. "Strong Duality for Generalized Trust Region Subproblem: S-Lemma with Interval Bounds." Optimization Letters, 9, pp. 1063-1073.

17. Wang, W. and Joe, B. and Goldman, R., 2002. "Computing quadric surface intersections based on an analysis of plane cubic curves." Graphic models, 64, pp. 335-367.

18. Will, I. and Manor, Y., 1993. "Quadric-surface intersection curves: shape and structure." Computer-Aided Design, 25, pp. 633-643.

19. Xia, Y. 2020. "A survey of hidden convex optimization." Journal of the Operations Research Society of China, 8(1), pp. 1-28.

20. Xia, Y. and Sheu, R.L. and Fang, S.C. and Xing, W, 2017. "Double well potential function and its optimization in the n-dimensional real space - Part 2." Journal of Industrial and Management Optimization, 13, pp. 1307-1328.

21. Xia, Y. and Wang, S. and Sheu, R.L., 2016. "S-lemma with equality and its applications." Mathematical Programming Series A, 156(1), pp. 513-547.

22. Yakubovich, V.A., 1971. "S-procedure in nonlinear control theory." Vestnik Leningrad University (in Russian), 1, pp. 62-77.

23. Ye, Y. and Zhang, S., 2003. "New results on quadratic minimization." SIAM Journal on Optimization, 14, pp. 245-267.