Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 9, No. 2, pp. 345–372, April 2013

A SMOOTHING SQP METHOD FOR MATHEMATICAL PROGRAMS WITH LINEAR SECOND-ORDER CONE COMPLEMENTARITY CONSTRAINTS*

Hiroshi Yamamura, Takayuki Okuno, Shunsuke Hayashi and Masao Fukushima

Abstract: In this paper, we focus on the mathematical program with second-order cone (SOC) complementarity constraints, which contains the well-known mathematical program with nonnegative complementarity constraints as a subclass. For solving such a problem, we propose a smoothing-based sequential quadratic programming (SQP) method. We first replace the SOC complementarity constraints with equality constraints using the smoothing natural residual function, and apply the SQP method to the smoothed problem with decreasing the smoothing parameter. We show that the proposed algorithm possesses the global convergence property under the Cartesian P_0 property and the nondegeneracy assumptions. We finally observe the effectiveness of the algorithm by means of numerical experiments.

Key words: *mathematical programs with equilibrium constraints, sequential quadratic programming, second-order cone complementarity*

Mathematics Subject Classification: *65K05, 90C30, 90C33*

---

1 Introduction

In this paper, we focus on the following mathematical program with second-order cone (SOC) complementarity constraints, abbreviated as MPSOCC:

$$\begin{array}{ll} \underset{x,y,z}{\text{Minimize}} & f(x,y) \\ \text{subject to} & Ax \leq b, \\ & z = Nx + My + q, \\ & \mathcal{K} \ni y \perp z \in \mathcal{K}, \end{array} \tag{1.1}$$

where $f\colon \mathbb{R}^{n+m} \to \mathbb{R}$ is a continuously differentiable function, $A \in \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^p$, $N \in \mathbb{R}^{m\times n}$, $M \in \mathbb{R}^{m\times m}$ and $q \in \mathbb{R}^m$ are given matrices and vectors, $\perp$ denotes the perpendicularity, and $\mathcal{K}$ is the Cartesian product of second-order cones, that is, $\mathcal{K} := \mathcal{K}^{m_1} \times \mathcal{K}^{m_2} \times \cdots \times \mathcal{K}^{m_\ell} \subseteq \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \cdots \times \mathbb{R}^{m_\ell} = \mathbb{R}^m$ with

$$\mathcal{K}^{m_i} = \begin{cases} \{u = (u_1,\ u_2) \in \mathbb{R} \times \mathbb{R}^{m_i-1} \,|\, \|u_2\| \leq u_1\} & (m_i \geq 2), \\ \mathbb{R}_+ = \{u \in \mathbb{R} \mid u \geq 0\} & (m_i = 1). \end{cases}$$

*This research was supported in part by Grant-in-Aid for Scientific Research (C) and Young Scientists (B) from Japan Society for the Promotion of Science.

Throughout the paper, we suppose that $m_i \geq 2$ for each i.

Mathematical program with equilibrium constraints (MPEC) [15] has been studied extensively, since it finds wide application such as design problems in engineering, the equilibrium problems in economics, and game-theoretic multi-level optimization problems. Particularly, equilibrium constraints in MPECs are often written as linear or nonlinear complementarity constraints. Such an MPEC is also called a mathematical program with complementarity constraints (MPCC). When $m_i = 1$ for all i, i.e., $\mathcal{K} = \mathbb{R}^m_+$, MPSOCC (1.1) reduces to the MPCC, for which there have been proposed many algorithms. For example, Fukushima and Tseng [12] proposed an active set algorithm, and proved that any accumulation point of the sequence generated by the algorithm is a B-stationary point under the uniform linear independence constraint qualification (LICQ) on the ε-feasible set. Luo, Pang, and Ralph [15] proposed a piece-wise sequential quadratic programming (SQP) algorithm, and showed that the generated sequence converges to a B-stationary point locally superlinearly or quadratically under the LICQ and the second order sufficient conditions. Fukushima, Luo, and Pang [10] proposed an SQP-type algorithm, and showed that the sequence generated by the algorithm globally converg es to a B-stationary point under the nondegeneracy condition at the limit point.

Problems with SOC constraints also attract much attention of many researchers. One of the typical problems is the second-order cone program (SOCP) [1]. The SOCP has a lot of applications such as the antenna array weight design, the finite response impulse (FIR) filter design, the portfolio optimization, and the magnetic shield design optimization. Moreover, SOCP includes many classes of problems such as linear program (LP), convex quadratic program (QP), etc. The second-order cone complementarity problem (SOCCP) [5, 7, 11] is another type of problems involving SOC constraints. Fukushima, Luo and Tseng [11] studied smoothing functions for the Fischer-Burmeister function and the natural residual function with respect to the SOC complementarity condition. Using those smoothing functions, Hayashi, Yamashita and Fukus hima [14] proposed a globally and quadratically convergent algorithm based on the smoothing and regularization methods. As an application of the SOCCP, Nishimura, Hayashi and Fukushima [16] studied the SOCCP reformulation of the robust Nash equilibrium problem in an N-person non-cooperative game.

As mentioned in the last two paragraphs, there have been many researches on the MPECs with "nonnegative" complementarity constraints and the optimization/complementarity problems with SOC constraints. However, there are only a few studies on MPECs with SOC complementarity constraints. For example, Yan and Fukushima [20] proposed a smoothing method for solving such problems. To show convergence of the algorithm, they assume that smoothed subproblems are solved exactly. However, it can hardly be expected in practice. To overcome such a difficulty, we propose to combine an SQP-type method with the smoothing method. The proposed method replaces the SOC complementarity condition of MPSOCC (1.1) with a certain vector equation by using a smoothed natural residual function, thereby yielding convex quadratic programming subproblems which can be solved efficiently by any state-of-the-art method such as the active method and interior point method. Although our method may be viewed as an extension of the SQP method in [10], the convergence analysis is quite different since it exploits the particular properties of the natural residual associated with the SOC complementarity condition.

This paper is organized as follows. In Section 2, we give some preliminaries. In Section 3, we reformulate MPSOCC (1.1) as a nonlinear programming problem by replacing the second-order cone complementarity constraints by equivalent nonsmooth equality constraints. In Section 4, we introduce a smoothing technique to deal with the nonsmooth constraints. In Section 5, we propose an SQP-type algorithm for solving problem (1.1) and show that the proposed method is well-defined under the Cartesian P_0 property. In Section 6, we show that the proposed algorithm possesses the global convergence property under the nondegeneracy assumptions. In Section 7, we give some numerical examples. In Section 8, we conclude the paper with some remarks.

Throughout the paper, we use the following notations. For a given vector $z \in \mathbb{R}^m$, z_i denotes the i-th element of $z \in \mathbb{R}^m$, while $z^i \in \mathbb{R}^{m_i}$ denotes the i-th column subvector conforming to the given Cartesian structure of $\mathcal{K}$. For subvectors $z^1 \in \mathbb{R}^{m_1}, z^2 \in \mathbb{R}^{m_2}, \dots, z^\ell \in \mathbb{R}^{m_\ell}$, we often let $(z^1, z^2, \dots, z^\ell)$ denote the column vector $((z^1)^\top, \cdots, (z^\ell)^\top) \in \mathbb{R}^{m_1+\cdots+m_\ell}$. For a vector $z \in \mathbb{R}^m$, we denote $\|z\|_1 := \sum_{i=1}^m |z_i|$, $\|z\| := \sqrt{z^\top z}$ and $\|z\|_\infty := \max_{1\le i\le m} |z_i|$. For a matrix $M \in \mathbb{R}^{m\times m}$, $\|M\|$ denotes the operator norm defined by $\|M\| = \max_{\|x\|=1} \|Mx\|$. For matrices $M, N \in \mathbb{R}^{m\times m}$, $M \succ (\succeq) N$ denotes that $M - N$ is a positive (semi)-definite matrix. We denote the interior and the boundary of $\mathcal{K}$ by $\operatorname{int}\mathcal{K}$ and $\operatorname{bd}\mathcal{K}$, respectively.[†] For vectors $y, z \in \mathbb{R}^m$, $y \succeq_{\mathcal{K}} z$ and $y \succ_{\mathcal{K}} z$ mean $y - z \in \mathcal{K}$ and $y - z \in \operatorname{int}\mathcal{K}$, respectively. We denote the nonnegative cone in $\mathbb{R}^m$ and its interior by $\mathbb{R}^m_+ := \{z \in \mathbb{R}^m \mid z_i \ge 0 \ (i = 1, 2, \dots, m)\}$ and $\mathbb{R}^m_{++} := \{z \in \mathbb{R}^m \mid z_i > 0 \ (i = 1, 2, \dots, m)\}$, respectively. Finally, for a set $C \subseteq \mathbb{R}^m$ and a vector $\bar{z} \in C$, we denote the normal cone [18] of C at $\bar{z}$ by $\mathcal{N}_C(z)$.

2 Preliminaries

2.1 Spectral Factorization and Natural Residual

The proposed algorithm relies on the fact that the SOC complementarity condition $\mathcal{K} \ni y \perp z \in \mathcal{K}$ can be rewritten as a system of equations by means of the natural residual. To be specific, we first recall the spectral factorization of a vector with respect to the SOC, $\mathcal{K}^m$.

Definition 2.1. For any vector $z := (z_1, z_2) \in \mathbb{R}\times\mathbb{R}^{m-1}$, we define the spectral factorization with respect to $\mathcal{K}^m$ as

$$z = \lambda_1 c^1 + \lambda_2 c^2,$$

where λ_1 and λ_2 are the spectral values given by

$$\lambda_j = z_1 + (-1)^j \|z_2\|, \quad j = 1,\ 2,$$

and c^1 and c^2 are the spectral vectors given by

$$c^j = \begin{cases} \dfrac{1}{2}\left(1,\ (-1)^j \dfrac{z_2}{\|z_2\|}\right) & \text{if } z_2 \ne 0 \\ \dfrac{1}{2}(1,\ (-1)^j v) & \text{if } z_2 = 0 \end{cases} \qquad j = 1,\ 2,$$

respectively, where $v \in \mathbb{R}^{m-1}$ is an arbitrary vector such that $\|v\| = 1$.

By using the spectral factorization, we can write the Euclidean projection onto $\mathcal{K}^m$ explicitly as follows [11]:

$$\begin{aligned} P_{\mathcal{K}^m}(z) &:= \operatorname{argmin}_{z'\in\mathcal{K}^m} \|z' - z\| \\ &= \max\{0,\ \lambda_1\}c^1 + \max\{0,\ \lambda_2\}c^2, \end{aligned}$$

where λ_j and c^j $(j = 1, 2)$ are the spectral values and the spectral vectors of z, respectively. Now, let us define the natural residual for the SOC complementarity condition by using the Euclidean projection.

[†] Note that $\operatorname{int}\mathcal{K} = \{(z_1, z_2) \in \mathbb{R}\times\mathbb{R}^{m-1} \mid z_1 > \|z_2\|\}$ and $\operatorname{bd}\mathcal{K}^m = \{(z_1, z_2) \in \mathbb{R}\times\mathbb{R}^{m-1} \mid z_1 = \|z_2\|\}$. In addition, for $\mathcal{K} = \mathcal{K}^{m_1} \times \cdots \times \mathcal{K}^{m_\ell}$, we have $\operatorname{int}\mathcal{K} = \operatorname{int}\mathcal{K}^{m_1} \times \cdots \times \operatorname{int}\mathcal{K}^{m_\ell}$ and $\operatorname{bd}\mathcal{K} = \mathcal{K} \setminus \operatorname{int}\mathcal{K}$.

Definition 2.2. Let $y := (y^1, y^2, \cdots, y^\ell)$ and $z := (z^1, z^2, \cdots, z^\ell) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \cdots \times \mathbb{R}^{m_\ell} = \mathbb{R}^m$ be arbitrary vectors. Then, the natural residual function $\Phi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ with respect to $\mathcal{K} = \mathcal{K}^{m_1} \times \mathcal{K}^{m_2} \times \cdots \times \mathcal{K}^{m_\ell}$ is defined as

$$\begin{aligned} \Phi(y,z) &:= y - P_{\mathcal{K}}(y-z) \\ &= \begin{pmatrix} \varphi^1(y^1,z^1) \\ \vdots \\ \varphi^\ell(y^\ell,z^\ell) \end{pmatrix} \in \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_\ell}, \end{aligned} \tag{2.1}$$

where

$$\varphi^i(y^i,z^i) = y^i - P_{\mathcal{K}^{m_i}}(y^i - z^i),\ i = 1,2,\dots,\ell.$$

It can be shown [11] that

$$\begin{aligned} \varphi^i(y^i,z^i) = 0 &\iff \mathcal{K}^{m_i} \ni y^i \perp z^i \in \mathcal{K}^{m_i}, \\ \Phi(y,z) = 0 &\iff \mathcal{K} \ni y \perp z \in \mathcal{K}. \end{aligned}$$

The following proposition states the property of function $\Phi(y,z)$ when $y - z \notin \mathrm{bd}\,(\mathcal{K} \cup -\mathcal{K})$.

Proposition 2.3. *Let $y, z \in \mathbb{R}^m$ be chosen so that $y - z \notin \mathrm{bd}\,(\mathcal{K} \cup -\mathcal{K})$. Then, the function $\Phi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ defined by* (2.1) *is continuously differentiable at (y,z), and the following equality holds*:

$$\nabla_y \Phi(y,z) + \nabla_z \Phi(y,z) = I_m,$$

where $I_m \in \mathbb{R}^{m\times m}$ denotes the identity matrix.

Proof. It suffices to consider the case where $\mathcal{K} = \mathcal{K}^m$. Let $\lambda_1, \lambda_2 \in \mathbb{R}$ be the spectral values of $y - z$ defined as in Definition 2.1. Note that, from $y - z \notin \mathrm{bd}\,(\mathcal{K}^m \cup -\mathcal{K}^m)$, we have $\lambda_1, \lambda_2 \neq 0$. Then, from [14, Proposition 4.8], the Clarke subdifferential $\partial P_{\mathcal{K}^m}(y-z)$ is explicitly given as

$$\partial P_{\mathcal{K}^m}(y-z) = \begin{cases} I_m & (\lambda_1 > 0,\ \lambda_2 > 0), \\ \dfrac{\lambda_2}{\lambda_2 - \lambda_1} I_m + W & (\lambda_1 < 0,\ \lambda_2 > 0), \\ O & (\lambda_1 < 0,\ \lambda_2 < 0), \end{cases}$$

where

$$W := \frac{1}{2}\begin{pmatrix} -r_1 & r_2^\top \\ r_2 & -r_1 r_2 r_2^\top \end{pmatrix}, \qquad (r_1, r_2) := \frac{(y_1 - z_1, y_2 - z_2)}{\|y_2 - z_2\|}.$$

Thus $P_{\mathcal{K}^m}$ is differentiable at $y - z$. This fact readily implies the continuous differentiability of Φ at (y,z) since $\Phi(y,z) = y - P_{\mathcal{K}^m}(y-z)$. We next show the second half of the proposition. By an easy calculation, we have

$$\nabla_y \Phi(y,z) \quad = \quad I_m - \nabla P_{\mathcal{K}^m}(y-z).$$

Similarly, we have $\nabla_z \Phi(y,z) = \nabla P_{\mathcal{K}^m}(y-z)$. Hence we obtain the desired equality. □

2.2 Cartesian P_0 and P Matrices

In this subsection, we introduce the Cartesian P_0 and the Cartesian P matrices. The concept of the Cartesian $P_0\,(P)$ matrix is a natural extension of the well-known $P_0\,(P)$ matrix [8]. Although the Cartesian $P_0\,(P)$ matrix can be defined not only for the SOC but also for the semidefinite cone [6] and the symmetric cone [13], we restrict ourselves to the case of the SOCs.

Definition 2.4. Suppose that the Cartesian structure of $\mathcal{K} \subseteq \mathbb{R}^m$ is given as $\mathcal{K} := \mathcal{K}^{m_1} \times \mathcal{K}^{m_2} \times \cdots \times \mathcal{K}^{m_\ell}$. Then, $M \in \mathbb{R}^{m\times m}$ is called

(a) a Cartesian P_0 matrix if, for every nonzero $z = (z^1, \dots, z^\ell) \in \mathbb{R}^m = \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_\ell}$, there exists an index $i \in \{1, \dots, \ell\}$ such that $(z^i)^\top (Mz)^i \geq 0$;

(b) a Cartesian P matrix if, for every nonzero $z = (z^1, \dots, z^\ell) \in \mathbb{R}^m = \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_\ell}$, there exists an index $i \in \{1, \dots, \ell\}$ such that $(z^i)^\top (Mz)^i > 0$.

Here, $(Mz)^i \in \mathbb{R}^{m_i}$ denotes the i-th subvector of $Mz \in \mathbb{R}^m$ conforming to the Cartesian structure of $\mathcal{K}$.

Notice that the definition of the Cartesian P_0 (P) property depends on the Cartesian structure of $\mathcal{K}$. In what follows, we assume that the Cartesian structure of $\mathcal{K}$ is always given as $\mathcal{K} = \mathcal{K}^{m_1} \times \mathcal{K}^{m_2} \times \cdots \times \mathcal{K}^{m_\ell}$. The definition of the "classical" $P_0\,(P)$ matrix corresponds to the case where $\mathcal{K} = \mathbb{R}^m_+$. It is easily seen that every Cartesian $P_0\,(P)$ matrix is a P_0 (P) matrix [17].

The following proposition implies that the Cartesian $P_0\,(P)$ property is preserved under a nonsingular block-diagonal transformation.

Proposition 2.5. *Let $M \in \mathbb{R}^{m\times m}$ be any matrix, and $H_i \in \mathbb{R}^{m_i \times m_i} (i = 1, 2, \dots, \ell)$ be arbitrary nonsingular matrices. Let the matrix $M' \in \mathbb{R}^{m\times m}$ be defined by*

$$M' := H^\top M H, \quad H := \begin{pmatrix} H_1 & & 0 \\ & \ddots & \\ 0 & & H_\ell \end{pmatrix}.$$

Then, the following statements hold.

(a) *If M is a Cartesian P_0 matrix, then M' is a Cartesian P_0 matrix.*

(b) *If M is a Cartesian P matrix, then M' is a Cartesian P matrix.*

Proof. We first show (b). Let $z = (z^1, \dots, z^\ell) \in \mathbb{R}^m = \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_\ell}$ be an arbitrary nonzero vector. We show that there exists an $i \in \{1, 2, \dots, \ell\}$ such that $(z^i)^\top (M'z)^i > 0$.

Note that

$$
\begin{aligned}
(M'z)^i &= \left(\begin{pmatrix} H_1 & & 0 \\ & \ddots & \\ 0 & & H_\ell \end{pmatrix}^\top M \begin{pmatrix} H_1 & & 0 \\ & \ddots & \\ 0 & & H_\ell \end{pmatrix} z \right)^i \\
&= \left(\begin{pmatrix} H_1 & & 0 \\ & \ddots & \\ 0 & & H_\ell \end{pmatrix}^\top M \begin{pmatrix} H_1 z^1 \\ \vdots \\ H_\ell z^\ell \end{pmatrix} \right)^i \\
&= \begin{pmatrix} H_1^\top \sum_{k=1}^{\ell} M_{1k} H_k z^k \\ \vdots \\ H_\ell^\top \sum_{k=1}^{\ell} M_{\ell k} H_k z^k \end{pmatrix}^i \\
&= \left(H_i^\top \sum_{k=1}^{\ell} M_{ik} H_k z^k \right).
\end{aligned}
$$

where $(\cdot)^i$ and $(\cdot)_{ik}$ denote the i-th subvector and the (i,k)-th block entry, respectively, conforming to the Cartesian structure of $\mathcal{K}$. Hence, we have

$$
(z^i)^\top (M'z)^i = (z^i)^\top H_i^\top \sum_{k=1}^{\ell} M_{ik} H_k z^k = \left(H_i z^i\right)^\top \sum_{k=1}^{\ell} M_{ik} H_k z^k = ((Hz)^i)^\top (MHz)^i.
$$

Since M is a Cartesian P matrix and $Hz \neq 0$ from the nonsingularity of H, we have $(z^i)^\top (M'z)^i = ((Hz)^i)^\top (MHz)^i > 0$ for some i. Hence, M' is a Cartesian P matrix.

We omit the proof of (a) since it can be shown in a similar manner to (b). □

3 Reformulation of MPSOCC and Relationship between the KKT Conditions and B-Stationary Points

In the previous section, we introduced the natural residual function Φ and observed that second-order cone complementarity condition $\mathcal{K} \ni y \perp z \in \mathcal{K}$ can be represented as $\Phi(y,z) = 0$ equivalently. In this section, we rewrite MPSOCC (1.1) as the following problem where the SOC complementarity constraint is replaced by the equivalent equality constraint involving the natural residual function Φ:

$$
\begin{array}{ll}
\underset{x,y,z}{\text{Minimize}} & f(x,y) \\
\text{subject to} & Ax \leq b, \\
& z = Nx + My + q, \\
& \Phi(y,z) = 0.
\end{array} \tag{3.1}
$$

We also call this problem MPSOCC. MPSOCC (3.1) is a nonsmooth optimization problem since Φ is not differentiable everywhere. However, as is claimed by the next proposition, Φ is continuously differentiable at any (y,z) satisfying the following nondegeneracy condition:

Definition 3.1 (Nondegeneracy)**.** Suppose that $(y,z)\in\mathbb{R}^m\times\mathbb{R}^m$ satisfies the SOC complementarity condition $\mathcal{K}\ni y\perp z\in\mathcal{K}$. Moreover, decompose y and z as $y=(y^1,y^2,\dots,y^\ell)$ and $z=(z^1,z^2,\dots,z^\ell)\in\mathbb{R}^{m_1}\times\mathbb{R}^{m_2}\times\cdots\times\mathbb{R}^{m_\ell}=\mathbb{R}^m$ conforming to the Cartesian structure of $\mathcal{K}$. Then, (y,z) is said to be nondegenerate if, for every $i=1,2,\dots,\ell$, one of the following three conditions holds:

(i) $y^i\in\operatorname{int}\mathcal{K}^{m_i},\ z^i=0$;

(ii) $y^i=0,\ z^i\in\operatorname{int}\mathcal{K}^{m_i}$;

(iii) $y^i\in\operatorname{bd}\mathcal{K}^{m_i}\setminus\{0\},\ z^i\in\operatorname{bd}\mathcal{K}^{m_i}\setminus\{0\},\ (y^i)^\top z^i=0$.

Proposition 3.2. *Let $(\overline{y},\overline{z})\in\mathbb{R}^m\times\mathbb{R}^m$ satisfy the nondegeneracy condition. Then, Φ is continuously differentiable at $(\overline{y},\overline{z})$.*

Proof. The nondegeneracy condition readily yields $\overline{y}-\overline{z}\notin\operatorname{bd}(\mathcal{K}\cup-\mathcal{K})$. Hence, Φ is continuously differentiable at $(\overline{y},\overline{z})$ by Proposition 2.3. ☐

Now, let $X:=\{x\in\mathbb{R}^n\mid Ax\le b\}$, and let $\overline{w}:=(\overline{x},\overline{y},\overline{z})\in\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^m$ be a feasible point of MPSOCC (1.1) that satisfies the nondegeneracy condition. By Proposition 3.2, MPSOCC (3.1) can be viewed as a smooth optimization problem in a small neighborhood of $\overline{w}$. Then, the Karush-Kuhn-Tucker (KKT) conditions on $\overline{w}$ are represented as

$$\begin{aligned}&\begin{pmatrix}\nabla_x f(\overline{x},\overline{y})\\ \nabla_y f(\overline{x},\overline{y})\\ 0\end{pmatrix}+\begin{pmatrix}N^T\\ M^T\\ -I\end{pmatrix}u+\begin{pmatrix}0\\ \nabla_y\Phi(\overline{y},\overline{z})\\ \nabla_z\Phi(\overline{y},\overline{z})\end{pmatrix}v\in-\mathcal{N}_X(\overline{x})\times\{0\}^{2m},\\ &\Phi(\overline{y},\overline{z})=0,\ A\overline{x}\le b,\ \overline{z}=N\overline{x}+M\overline{y}+q,\end{aligned}\tag{3.2}$$

where $\{0\}^{2m}:=\{0\}\times\{0\}\times\cdots\times\{0\}\subseteq\mathbb{R}^{2m}$, and $u\in\mathbb{R}^\ell, v\in\mathbb{R}^m$ and $\eta\in\mathbb{R}^m$ are Lagrange multipliers.

We next consider the stationarity of MPSOCC (1.1) or (3.1). So far, several kinds of stationary points have been studied in the literature of MPECs, e.g., see [19]. Among them, a Bouligand- or B-stationary point is the most desirable, since it is directly related to the first order optimality condition. Specifically, a B-stationary point for MPSOCC (1.1) is defined as follows:

Definition 3.3 (B-stationarity)**.** Let $\mathcal{F}\subseteq\mathbb{R}^{n+2m}$ denote the feasible set of MPSOCC (1.1). We say that $\overline{w}:=(\overline{x},\overline{y},\overline{z})\in\mathcal{F}$ is a B-stationary point of MPSOCC (1.1) if $(-\nabla f(\overline{x},\overline{y}),0)\in\mathcal{N}_{\mathcal{F}}(\overline{w})$ holds.

In what follows, we show that a point satisfying KKT conditions (3.2) is a B-stationary point. For this purpose, we give three useful lemmas.

Lemma 3.4 ([18, Proposition 6.41])**.** *Let $C:=C_1\times C_2\times\cdots\times C_s$ for nonempty closed sets $C_i\subseteq\mathbb{R}^{n_i}$. Choose $\overline{\zeta}:=(\overline{\zeta}_1,\overline{\zeta}_2,\dots,\overline{\zeta}_s)\in C_1\times C_2\times\cdots\times C_s$. Then, we have $\mathcal{N}_C(\overline{\zeta})=\mathcal{N}_{C_1}(\overline{\zeta}_1)\times\mathcal{N}_{C_2}(\overline{\zeta}_2)\times\cdots\times\mathcal{N}_{C_s}(\overline{\zeta}_s)$.*

Lemma 3.5 ([18, Chapter 6-C])**.** *For a continuously differentiable function $F:\mathbb{R}^p\to\mathbb{R}^q$, let $D:=\{\zeta\in\mathbb{R}^p\mid F(\zeta)=0\}$. Choose $\overline{\zeta}\in D$ arbitrarily. If $\nabla F(\overline{\zeta})$ has full column rank, then we have*

$$\mathcal{N}_D(\overline{\zeta})=\nabla F(\overline{\zeta})\mathbb{R}^q:=\{\zeta\in\mathbb{R}^p\mid\zeta=\nabla F(\overline{\zeta})v,\ v\in\mathbb{R}^q\}.$$

Lemma 3.6 ([18, Theorem 6.14])**.** *For a continuously differentiable function* $F : \mathbb{R}^p \to \mathbb{R}^q$ *and a closed set* $C \subseteq \mathbb{R}^p$*, let* $D := \{\zeta \in C \mid F(\zeta) = 0\}$*. Choose* $\overline{\zeta} \in D$ *arbitrarily. Then we have*

$$\mathcal{N}_D(\overline{\zeta}) \supseteq \nabla F(\overline{\zeta})\mathbb{R}^q + \mathcal{N}_C(\overline{\zeta}).$$

Now, we show that the KKT conditions (3.2) are sufficient conditions for $\overline{w} = (\overline{x}, \overline{y}, \overline{z})$ to be B-stationary.

Proposition 3.7. *Let* $\overline{w} := (\overline{x}, \overline{y}, \overline{z})$ *be a feasible point of MPSOCC* (3.1)*. Suppose that the nondegeneracy condition holds at* $(\overline{y}, \overline{z})$*. If* $\overline{w}$ *satisfies the KKT conditions* (3.2)*, then* $\overline{w}$ *is a B-stationary point of MPSOCC* (1.1)*.*

Proof. Let $Y := \{(y, z) \in \mathbb{R}^m \times \mathbb{R}^m \mid \Phi(y, z) = 0\}$. We first note that the nondegeneracy at $(\overline{y}, \overline{z})$ implies the continuous differentiability of Φ at $(\overline{y}, \overline{z})$ from Proposition 2.3. Choose $v \in \mathbb{R}^m$ such that $\nabla\Phi(\overline{y}, \overline{z})v = 0$. From Proposition 2.3, we then have $\nabla_y\Phi(\overline{y}, \overline{z})v = 0$ and $\nabla_z\Phi(\overline{y}, \overline{z})v = (I - \nabla_y\Phi(\overline{y}, \overline{z}))v = 0$, which readily imply $v = 0$, and thus $\nabla\Phi(\overline{y}, \overline{z})$ has full column rank. Therefore, by Lemma 3.5 with $p = q := m$, $D := Y$ and $F := \Phi$, we have

$$\mathcal{N}_Y(\overline{y}, \overline{z}) = \nabla\Phi(\overline{y}, \overline{z})\mathbb{R}^m. \tag{3.3}$$

Then, it holds that

$$\mathcal{N}_{X\times Y}(\overline{w}) = \mathcal{N}_X(\overline{x}) \times \mathcal{N}_Y(\overline{y}, \overline{z}) = \mathcal{N}_X(\overline{x}) \times \nabla\Phi(\overline{y}, \overline{z})\mathbb{R}^m, \tag{3.4}$$

where the first equality follows from Lemma 3.4 and the second equality follows from (3.3). Now, let $\mathcal{F} \subseteq \mathbb{R}^{n+2m}$ denote the feasible set of MPSOCC (3.1), i.e., $\mathcal{F} = \{(x, y, z) \in X \times Y \mid Nx + My - z + q = 0\}$. Then, from Lemma 3.6, we have

$$\mathcal{N}_{\mathcal{F}}(\overline{w}) \supseteq \left(N^\top, M^\top, -I\right)^\top \mathbb{R}^m + \mathcal{N}_{X\times Y}(\overline{w}). \tag{3.5}$$

Combining (3.4) with (3.5), we obtain

$$\mathcal{N}_{\mathcal{F}}(\overline{w}) \supseteq \left(N^\top, M^\top, -I\right)^\top \mathbb{R}^m + \mathcal{N}_X(\overline{x}) \times \nabla\Phi(\overline{y}, \overline{z})\mathbb{R}^m,$$

which together with the KKT conditions (3.2) implies $(-\nabla f(\overline{x}, \overline{y}), 0) \in \mathcal{N}_{\mathcal{F}}(\overline{w})$. Thus, $\overline{w}$ is a B-stationary point. □

4 Smoothing Function of Natural Residual

The natural residual function Φ given in Definition 2.2 is not differentiable everywhere, and therefore, we cannot employ a derivative-based algorithm such as Newton's method to solve MPSOCC (3.1). To overcome such a difficulty, we will utilize a smoothing technique.

Definition 4.1. Let $\Psi : \mathbb{R}^m \to \mathbb{R}^m$ be a nondifferentiable function. Then, the function $\Psi_\mu : \mathbb{R}^m \to \mathbb{R}^m$ parametrized by $\mu > 0$ is called a smoothing function of Ψ if it satisfies the following properties: For any $\mu > 0$, Ψ_μ is differentiable on $\mathbb{R}^m$; for any $z \in \mathbb{R}^m$, it holds that $\lim_{\mu\to 0+} \Psi_\mu(z) = \Psi(z)$.

A smoothing function of the natural residual function can be constructed by means of the Chen-Mangasarian (CM) function $\hat{g} : \mathbb{R} \to \mathbb{R}$ [11].

Definition 4.2. A differentiable convex function $\hat{g} : \mathbb{R} \to \mathbb{R}_+$ is called a CM function if

$$\lim_{\alpha\to-\infty} \hat{g}(\alpha) = 0, \qquad \lim_{\alpha\to\infty} (\hat{g}(\alpha) - \alpha) = 0, \qquad 0 < \hat{g}'(\alpha) < 1 \qquad (\alpha \in \mathbb{R}). \tag{4.1}$$

Notice that, if function $p_\mu : \mathbb{R} \to \mathbb{R}$ is defined by $p_\mu(\alpha) := \mu\hat{g}(\alpha/\mu)$ with a CM function $\hat{g}$ and a positive parameter μ, then it becomes a smoothing function for $p(\alpha) := \max\{0, \alpha\}$. Thanks to this fact, we can next provide a smoothing function P_μ for the projection operator $P_\mathcal{K}$.

Definition 4.3. Let $z \in \mathbb{R}^m$ be an arbitrary vector decomposed as $z = (z^1, z^2, \cdots, z^\ell) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \cdots \times \mathbb{R}^{m_\ell} = \mathbb{R}^m$ conforming to the given Cartesian structure of $\mathcal{K}$. For an arbitrary CM function $\hat{g} : \mathbb{R} \to \mathbb{R}$, let $g : \mathbb{R}^m \to \mathbb{R}^m$ be defined as

$$\begin{aligned} g(z) &:= \begin{pmatrix} g^1(z^1) \\ \vdots \\ g^\ell(z^\ell) \end{pmatrix}, \qquad (4.2) \\ g^i(z) &:= \hat{g}(\lambda_{i1})c^{i1} + \hat{g}(\lambda_{i2})c^{i2}, \end{aligned}$$

where $\lambda_{ij} \in \mathbb{R}$ and $c^{ij} \in \mathbb{R}^{m_i}$ $((i,j) \in \{1, 2, \cdots, \ell\} \times \{1, 2\})$ are the spectral values and the spectral vectors of subvectors z^i with respect to $\mathcal{K}^{m_i}$, respectively. Then, the smoothing function P_μ: $\mathbb{R}^m \to \mathbb{R}^m$ of $P_\mathcal{K}$ is given as

$$P_\mu(z) := \mu g(z/\mu).$$

Now, by using the above smoothing function P_μ, we can define the smoothing function for the natural residual Φ.

Definition 4.4. Let $\mu > 0$ be arbitrary. Let $g : \mathbb{R}^m \to \mathbb{R}^m, g^i : \mathbb{R}^{m_i} \to \mathbb{R}^{m_i}$ $(i = 1, 2, \ldots, \ell)$, and $P_\mu : \mathbb{R}^m \to \mathbb{R}^m$ be defined as in Definition 4.3. Then, the smoothing function $\Phi_\mu : \mathbb{R}^m \to \mathbb{R}^m$ for the natural residual Φ is given as

$$\begin{aligned} \Phi_\mu(y, z) &:= y - P_\mu(y - z) \\ &= y - \mu g\left(\frac{y - z}{\mu}\right) \\ &= \begin{pmatrix} y^1 - \mu g^1\left(\dfrac{y^1 - z^1}{\mu}\right) \\ \vdots \\ y^\ell - \mu g^\ell\left(\dfrac{y^\ell - z^\ell}{\mu}\right) \end{pmatrix}. \qquad (4.3) \end{aligned}$$

Before closing this subsection, we provide the following propositions which will be used in the subsequent analyses.

Proposition 4.5. *Let $P_\mu : \mathbb{R}^m \to \mathbb{R}^m$ be defined as in Definition 4.3, and choose $\overline{z} \notin \mathrm{bd}\,(\mathcal{K} \cup -\mathcal{K})$ arbitrarily. Let $\{z^k\} \subseteq \mathbb{R}^m$ and $\{\mu_k\} \subseteq \mathbb{R}_{++}$ be arbitrary sequences such that $z^k \to \overline{z}$ and $\mu_k \to 0$ as $k \to \infty$. Then, we have*

$$\nabla P_\mathcal{K}(\overline{z}) = \lim_{k \to \infty} \nabla P_{\mu_k}(z^k). \qquad (4.4)$$

Proof. For simplicity, we consider the case where $\mathcal{K} = \mathcal{K}^m$. Let $\overline{z}$ and z^k be decomposed as $\overline{z} = \overline{\lambda}_1\overline{c}^1 + \overline{\lambda}_2\overline{c}^2$ and $z^k = \lambda_1^k c_k^1 + \lambda_2^k c_k^2$, where $\overline{\lambda}_i$, $\lambda_i^k \in \mathbb{R}$ are spectral values, and $\overline{c}^i$, $c_k^i \in \mathbb{R}^m$ $(i = 1, 2)$ are spectral vectors of $\overline{z}$ and z^k, respectively. Since $\overline{z} \notin \mathrm{bd}\,(\mathcal{K}^m \cup -\mathcal{K}^m)$,

$P_{\mathcal{K}^m}$ is differentiable at $\overline{z}$ and $\nabla P_{\mathcal{K}^m}(\overline{z})$ is given as

$$\nabla P_{\mathcal{K}^m}(\overline{z}) = \begin{cases} I_m & (\overline{\lambda}_1 > 0,\ \overline{\lambda}_2 > 0), \\ \dfrac{\overline{\lambda}_2}{\overline{\lambda}_2 - \overline{\lambda}_1} I_m + W & (\overline{\lambda}_1 < 0,\ \overline{\lambda}_2 > 0), \\ O & (\overline{\lambda}_1 < 0,\ \overline{\lambda}_2 < 0), \end{cases}$$

where

$$W := \frac{1}{2}\begin{pmatrix} -r_1 & r_2^\top \\ r_2 & -r_1 r_2 r_2^\top \end{pmatrix},\ (r_1, r_2) := \frac{(\overline{z}_1, \overline{z}_2)}{\|\overline{z}_2\|},\ \overline{z} := (\overline{z}_1, \overline{z}_2) \in \mathbb{R} \times \mathbb{R}^{m-1}.$$

On the other hand, by [11, Proposition 5.2], $\nabla P_{\mu_k}(z^k)$ is written as

$$\nabla P_{\mu_k}(z^k) = \begin{cases} \hat{g}'(z_1^k/\mu_k) I_m & (z_2^k = 0), \\ \begin{pmatrix} b_{\mu_k} & \dfrac{c_{\mu_k}(z_2^k)^\top}{\|z_2^k\|} \\ \dfrac{c_{\mu_k} z_2^k}{\|z_2^k\|} & a_{\mu_k} I_{m-1} + (b_{\mu_k} - a_{\mu_k}) \dfrac{z_2^k {z_2^k}^\top}{\|z_2^k\|^2} \end{pmatrix} & (z_2^k \neq 0), \end{cases}$$

where

$$a_{\mu_k} = \frac{\hat{g}(\lambda_2^k/\mu_k) - \hat{g}(\lambda_1^k/\mu_k)}{\lambda_2^k/\mu_k - \lambda_1^k/\mu_k},\quad b_{\mu_k} = \frac{1}{2}\left(\hat{g}'\left(\frac{\lambda_2^k}{\mu_k}\right) + \hat{g}'\left(\frac{\lambda_1^k}{\mu_k}\right)\right),$$
$$c_{\mu_k} = \frac{1}{2}\left(\hat{g}'\left(\frac{\lambda_2^k}{\mu_k}\right) - \hat{g}'\left(\frac{\lambda_1^k}{\mu_k}\right)\right),\quad z^k := (z_1^k, z_2^k) \in \mathbb{R} \times \mathbb{R}^{m-1},$$

and $\hat{g}$ is defined as in Definition 4.3. Note that, from the definition of $\hat{g}$, we have $\hat{g}(\alpha) - \alpha \to 0$, $\hat{g}(-\alpha) \to 0$, $\hat{g}'(\alpha) \to 1$ and $\hat{g}'(-\alpha) \to 0$ as $\alpha \to \infty$. Then, it follows that

$$\lim_{k\to\infty} a_{\mu_k} = \begin{cases} 1 & (0 < \overline{\lambda}_1 \leq \overline{\lambda}_2) \\ \overline{\lambda}_2/(\overline{\lambda}_2 - \overline{\lambda}_1) & (\overline{\lambda}_1 < 0 < \overline{\lambda}_2) \\ 0 & (\overline{\lambda}_1 \leq \overline{\lambda}_2 < 0) \end{cases},$$

$$\lim_{k\to\infty} b_{\mu_k} = \begin{cases} 1 & (0 < \overline{\lambda}_1 \leq \overline{\lambda}_2) \\ 1/2 & (\overline{\lambda}_1 < 0 < \overline{\lambda}_2) \\ 0 & (\overline{\lambda}_1 \leq \overline{\lambda}_2 < 0) \end{cases},$$

$$\lim_{k\to\infty} c_{\mu_k} = \begin{cases} 0 & (0 < \overline{\lambda}_1 \leq \overline{\lambda}_2) \\ 1/2 & (\overline{\lambda}_1 < 0 < \overline{\lambda}_2) \\ 0 & (\overline{\lambda}_1 \leq \overline{\lambda}_2 < 0) \end{cases}.$$

From this fact, it is not difficult to observe that $\nabla P_{\mathcal{K}^m}(\overline{z}) = \lim_{k\to\infty} \nabla P_{\mu_k}(z^k)$. □

Proposition 4.6 ([11, Proposition 5.1]). *Let $\Phi_\mu : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ and $\Phi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ be defined by* (2.1) *and* (4.3)*, respectively. Let $\rho := \hat{g}(0)$. Then, for any $y, z \in \mathbb{R}^n$ and $\mu > \nu > 0$, we have*

$$\rho(\mu - \nu)e \succeq_{\mathcal{K}} \Phi_\nu(y, z) - \Phi_\mu(y, z) \succ_{\mathcal{K}} 0,$$
$$\rho\mu e \succeq_{\mathcal{K}} \Phi(y, z) - \Phi_\mu(y, z) \succ_{\mathcal{K}} 0,$$

where $e := (e^1, e^2, \ldots, e^\ell) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \cdots \times \mathbb{R}^{m_\ell}$ with $e^i := (1, 0, 0, \ldots, 0)^\top \in \mathbb{R}^{m_i}$ for $i = 1, 2, \ldots, \ell$.

Proposition 4.7 ([11, Corollary 5.3 and Proposition 6.1])**.** *Let* $\Phi_\mu : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ and $\Phi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ *be defined by* (2.1) *and* (4.3)*, respectively. Let* $g : \mathbb{R}^m \to \mathbb{R}^m$ *be defined by* (4.2)*. Then, the following statements hold.*

(a) *Function g is continuously differentiable and* $\nabla g(z) = \mathrm{diag}(\nabla g^1(z^1), \dots, \nabla g^\ell(z^\ell)) \in \mathbb{R}^{m\times m}$ *is symmetric for any* $z \in \mathbb{R}^m$*, where the latter matrix denotes the block-diagonal matrix with block-diagonal elements* $\nabla g^i(z^i),\ i = 1, 2, \dots, \ell$.

(b) *For any* $y, z \in \mathbb{R}^m$*, we have*

$$\nabla_y \Phi_\mu(y, z) = I_m - \nabla g\left(\frac{y-z}{\mu}\right),\ \nabla_z \Phi_\mu(y, z) = \nabla g\left(\frac{y-z}{\mu}\right),$$

where $I_m \in \mathbb{R}^{m\times m}$ *denotes the identity matrix.*

(c) *For any* $y, z \in \mathbb{R}^m$*, we have*

$$O \prec \nabla_y \Phi_\mu(y, z) \prec I_m, \quad O \prec \nabla_z \Phi_\mu(y, z) \prec I_m, \quad O \prec \nabla g\left(\frac{y-z}{\mu}\right) \prec I_m.$$

Proposition 4.8. *Let* $\Phi_\mu : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ *and* $\Phi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ *be defined by* (2.1) *and* (4.3)*, respectively. Let* $\rho := \hat{g}(0)$*. Then,* $\|\Phi_\mu(y, z) - \Phi(y, z)\| \le \sqrt{2}\rho\mu$ *for any* $(\mu, y, z) \in \mathbb{R}_{++} \times \mathbb{R}^m \times \mathbb{R}^m$*.*

Proof. For simplicity, we only consider the case where $\mathcal{K} = \mathcal{K}^m$. Let $\Phi(y, z) - \Phi_\mu(y, z) = \lambda_1 c^1 + \lambda_2 c^2$, where $\lambda_i \in \mathbb{R}$ and $c^i \in \mathbb{R}^m$ $(i = 1, 2)$ are the spectral values and spectral vectors of $\Phi(y, z) - \Phi_\mu(y, z)$. Since $e = c^1 + c^2$ and $\rho\mu e \succeq_{\mathcal{K}^m} \Phi(y, z) - \Phi_\mu(y, z) \succ_{\mathcal{K}^m} 0$ from Proposition 4.6, we have $\rho\mu(c^1 + c^2) \succeq_{\mathcal{K}^m} \lambda_1 c^1 + \lambda_2 c^2 \succ_{\mathcal{K}^m} 0$, which implies $0 < \lambda_1 \le \lambda_2 \le \rho\mu$. Hence, we obtain

$$\|\Phi(y, z) - \Phi_\mu(y, z)\| = \|\lambda_1 c^1 + \lambda_2 c^2\| \le \lambda_1\|c^1\| + \lambda_2\|c^2\| \le \sqrt{2}\rho\mu,$$

where the first inequality is due to the triangle inequality and $0 < \lambda_1 \le \lambda_2$, and the last inequality follows from $\|c^1\| = \|c^2\| = 1/\sqrt{2}$ and $\lambda_1 \le \lambda_2 \le \rho\mu$. This completes the proof. ☐

Proposition 4.9. *Let* $\Phi_\mu : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ *be defined by* (4.3)*. Then, for any* $\mu > \nu > 0$ *and* $(y, z) \in \mathbb{R}^m \times \mathbb{R}^m$*, it holds that*

$$\|\Phi_\nu(y, z)\|_1 - \|\Phi_\mu(y, z)\|_1 \le m\rho(\mu - \nu),$$

where $\rho = \hat{g}(0)$*.*

Proof. We first assume $\mathcal{K} = \mathcal{K}^m$. From Proposition 4.6, we have

$$\rho(\mu - \nu)e - (\Phi_\nu(y, z) - \Phi_\mu(y, z)) \in \mathcal{K}^m, \tag{4.5}$$

$$\Phi_\nu(y, z) - \Phi_\mu(y, z) \in \mathcal{K}^m, \tag{4.6}$$

where $e = (1, 0, \dots, 0)^\top \in \mathbb{R}^m$. Moreover, for any $w = (w_1, w_2, \dots, w_m)^\top \in \mathcal{K}^m$, we have

$$w_1 \ge |w_i| \quad (i = 1, \dots, m), \tag{4.7}$$

since $w_1 \geq \sqrt{w_2^2 + \cdots + w_m^2}$. Therefore, for each $i = 1, 2, \ldots, m$, we have

$$\begin{aligned} \rho(\mu - \nu) &\geq \left(\Phi_\nu(y,z) - \Phi_\mu(y,z)\right)_1 \\ &\geq \left|\left(\Phi_\nu(y,z) - \Phi_\mu(y,z)\right)_i\right| \\ &\geq |\Phi_\nu(y,z)_i| - |\Phi_\mu(y,z)_i|, \end{aligned} \tag{4.8}$$

where the first inequality holds from (4.3) and (4.5), the second equality holds from (4.6) and (4.7), and the last equality holds from the triangle inequality. Summing up (4.8) for all i, we obtain the desired conclusion. When $\mathcal{K} = \mathcal{K}^{m_1} \times \cdots \times \mathcal{K}^{m_\ell}$, we can prove it in a similar way. □

5 Algorithm

In this section, we propose an SQP type algorithm for MPSOCC (1.1). The SQP method solves a quadratic programming (QP) problem in each iteration to determine the search direction. This method is known as one of the most efficient methods for solving nonlinear programming problems. In the remainder of the paper, to apply the SQP method, we mainly consider MPSOCC (3.1) equivalent to MPSOCC (1.1). We should notice, however, that the SQP method cannot be applied directly to MPSOCC (3.1), since $\Phi(y,z)$ is not differentiable everywhere. We thus consider the following problem where the smooth equality constraint $\Phi_\mu(y,z) = 0$ replaces $\Phi(y,z) = 0$ in each iteration

$$\begin{array}{ll} \underset{x,y,z}{\text{Minimize}} & f(x,y) \\ \text{subject to} & Ax \leq b, \\ & z = Nx + My + q, \\ & \Phi_\mu(y,z) = 0. \end{array} \tag{5.1}$$

Given a current iterate (x^k, y^k, z^k) satisfying $Ax^k \leq b^k$ and $z^k = Nx^k + My^k + q$, we then generate the search direction (dx^k, dy^k, dz^k) by solving the following QP subproblem, which consists of quadratic and linear approximations of the objective and constraint functions of problem (5.1) with $\mu = \mu_k$, respectively, at (x^k, y^k, z^k):

$$\begin{array}{ll} \underset{dx,dy,dz}{\text{Minimize}} & \nabla f(x^k,y^k)^\top \begin{pmatrix} dx \\ dy \end{pmatrix} + \dfrac{1}{2} \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix}^\top B_k \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} \\ \text{subject to} & Adx \leq b - Ax^k, \\ & \begin{pmatrix} N & M & -I_m \\ 0 & \nabla_y \Phi_{\mu_k}(y^k,z^k)^\top & \nabla_z \Phi_{\mu_k}(y^k,z^k)^\top \end{pmatrix} \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} = -\begin{pmatrix} 0 \\ \Phi_{\mu_k}(y^k,z^k) \end{pmatrix}, \end{array} \tag{5.2}$$

where $B_k \in \mathbb{R}^{(n+2m)\times(n+2m)}$ is a positive definite symmetric matrix. In the numerical experiments in Section 7, B_k will be updated by using the modified Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula. Note that the Karush-Kuhn-Tucker (KKT) conditions of QP (5.2) can be written as

$$\begin{aligned}
&\begin{pmatrix} \nabla_x f(x^k, y^k) \\ \nabla_y f(x^k, y^k) \\ 0 \end{pmatrix} + B_k \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} + \begin{pmatrix} N^\top \\ M^\top \\ -I_m \end{pmatrix} u \\
&\qquad\qquad + \begin{pmatrix} 0 \\ \nabla_y \Phi_{\mu_k}(y^k, z^k) \\ \nabla_z \Phi_{\mu_k}(y^k, z^k) \end{pmatrix} v + \begin{pmatrix} A^\top \\ 0 \\ 0 \end{pmatrix} \eta = 0, \\
&\begin{pmatrix} N & M & -I_m \\ 0 & \nabla_y \Phi_{\mu_k}(y^k, z^k)^\top & \nabla_z \Phi_{\mu_k}(y^k, z^k)^\top \end{pmatrix} \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} = - \begin{pmatrix} 0 \\ \Phi_{\mu_k}(y^k, z^k) \end{pmatrix}, \\
&0 \le (b - Ax^k - Adx) \perp \eta \ge 0,
\end{aligned} \tag{5.3}$$

where $(\eta,\, u,\, v) \in \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^m$ denotes the Lagrange multipliers.

For simplicity of notation, we denote

$$w := (x, y, z) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m, \;\; dw := (dx, dy, dz) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m.$$

Also, we define the ℓ_1 penalty function by

$$\theta_{\mu,\alpha}(w) := f(x, y) + \alpha \|\Phi_\mu(y, z)\|_1, \tag{5.4}$$

where $\alpha > 0$ is the penalty parameter. Note that this function has the directional derivative $\theta'_{\mu,\alpha}(w; dw)$ for any w and dw.

Algorithm 1.

Step 0: Choose parameters $\delta \in (0, \infty)$, $\beta \in (0, 1)$, $\rho \in (0, 1)$, $\sigma \in (0, 1)$, $\mu_0 \in (0, \infty)$, $\alpha_{-1} \in (0, \infty)$ and a symmetric positive definite matrix $B_0 \in \mathbb{R}^{(n+2m)\times(n+2m)}$. Choose $w^0 = (x^0, y^0, z^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ such that $Nx^0 + My^0 + q = z^0$ and $Ax^0 \le b$. Set $k := 0$.

Step 1: Solve QP subproblem (5.2) to obtain the optimum $dw^k = (dx^k, dy^k, dz^k)$ and the Lagrange multipliers (η^k, u^k, v^k).

Step 2: If $dw^k = 0$, then let $w^{k+1} := w^k$, $\alpha_k := \alpha_{k-1}$ and go to Step 3. Otherwise, update the penalty parameter by

$$\alpha_k := \begin{cases} \alpha_{k-1} & \text{if } \; \alpha_{k-1} \ge \|v^k\|_\infty + \delta, \\ \max\{\|v^k\|_\infty + \delta, \;\; \alpha_{k-1} + 2\delta\} & \text{otherwise.} \end{cases} \tag{5.5}$$

Then, set the step size $\tau_k := \rho^L$, where L is the smallest nonnegative integer satisfying the Armijo condition

$$\theta_{\mu_k,\alpha_k}(w^k + \rho^L dw^k) \le \theta_{\mu_k,\alpha_k}(w^k) + \sigma \rho^L \theta'_{\mu_k,\alpha_k}(w^k; dw^k). \tag{5.6}$$

Let $w^{k+1} := w^k + \tau_k dw^k$, and go to Step 3.

Step 3: Terminate if a certain criterion is satisfied. Otherwise, let $\mu_{k+1} := \beta\mu_k$ and update B_k to determine a symmetric positive definite matrix B_{k+1}. Return to Step 1 with k replaced by $k + 1$.

In the remainder of this section, we establish the well-definedness of Algorithm 1. We first show the feasibility of QP subproblem (5.2). In general, a QP subproblem generated by the SQP method may not be feasible, even if the original nonlinear programming problem is feasible. However, in the present case, we can show that QP subproblem (5.2) is always feasible under the Cartesian P_0 property of the matrix M. To this end, the following lemma will be useful.

Lemma 5.1. *Let $M \in \mathbb{R}^{m\times m}$ be a Cartesian P_0 matrix. Let $H_i \in \mathbb{R}^{m_i \times m_i}$ $(i = 1, 2, \dots, \ell)$ be positive definite matrices with $m = \sum_{i=1}^{\ell} m_i$, and $H \in \mathbb{R}^{m\times m}$ be a block diagonal matrix with block diagonal elements H_i $(i = 1, \dots, \ell)$. Then, $H + M$ is nonsingular.*

Proof. The matrix $H + M$ can easily be shown to be a Cartesian P matrix, which is nonsingular. □

The next proposition shows the feasibility and solvability of QP subproblem (5.2). In the proof, the matrix

$$D_k := \begin{pmatrix} M & -I_m \\ \nabla_y \Phi_{\mu_k}(y^k, z^k)^\top & \nabla_z \Phi_{\mu_k}(y^k, z^k)^\top \end{pmatrix} \tag{5.7}$$

plays an important role.

Proposition 5.2 (Feasibility of QP subproblem)**.** *Let M be a Cartesian P_0 matrix, and $\{w^k\}$ be a sequence generated by Algorithm* 1*. Then,* (i) *$Ax^k \le b$ and $z^k = Nx^k + My^k + q$ hold for all k, and* (ii) *QP subproblem* (5.2) *is feasible and hence has a unique solution for all k.*

Proof. Since (i) can be shown easily, we only show (ii). Since the objective function of QP (5.2) is strongly convex, it suffices to show the feasibility. We first show that the matrix D_k defined by (5.7) is nonsingular. Note that, by Proposition 4.7(c), $\nabla_z \Phi_{\mu_k}(y^k, z^k)$ is nonsigular. Let $\tilde{D}_k$ be the Schur complement of the matrix $\nabla_z \Phi_{\mu_k}(y^k, z^k)^\top$ with respect to D_k, that is,

$$\begin{aligned} \tilde{D}_k &:= M + \left(\nabla_z \Phi_{\mu_k}(y^k, z^k)^\top\right)^{-1} \nabla_y \Phi_{\mu}(y^k, z^k)^\top \\ &= M + \left(\nabla g\left(\frac{y^k - z^k}{\mu_k}\right)^\top\right)^{-1} \left(I_m - \nabla g\left(\frac{y^k - z^k}{\mu_k}\right)^\top\right) \\ &= M + \operatorname{diag}\left(\left(\nabla g^i\left(\frac{y^{i,k} - z^{i,k}}{\mu_k}\right)^\top\right)^{-1} - I_{m_i}\right)_{i=1}^{\ell}, \end{aligned}$$

where $y^{i,k}$ and $z^{i,k}$ denote the i-th subvectors of y^k and z^k, respectively, conforming to the Cartesian structure of $\mathcal{K}$, and each equality follows from Proposition 4.7. Since M is a Cartesian P_0 matrix and $(\nabla g^i((y^{i,k} - z^{i,k})/\mu_k)^\top)^{-1} - I_{m_i} \in \mathbb{R}^{m_i \times m_i}$ is positive definite from Proposition 4.7 (c), Lemma 5.1 ensures that $\tilde{D}_k$ is nonsingular, and hence D_k is nonsingular. It then follows from (i) that

$$dx = 0, \quad \begin{pmatrix} dy \\ dz \end{pmatrix} = -D_k^{-1} \begin{pmatrix} 0 \\ \Phi_{\mu_k}(y^k, z^k) \end{pmatrix}$$

comprise a feasible solution to (5.2). This completes the proof. □

The following proposition shows that the search direction dw^k produced in Step 1 of Algorithm 1 is a descent direction of the penalty function θ_{μ_k,α_k} defined by (5.4). It guarantees the well-definedness of the line search in Step 2 in the sense that there exists a finite L satisfying the Armijo condition (5.6).

Proposition 5.3 (Descent direction)**.** *Let* $\{w^k\}$ *and* $\{dw^k\}$ *be sequences generated by Algorithm* 1*. Then, we have*

(a) $\theta'_{\mu_k,\alpha_k}(w^k; dw^k) = \nabla_x f(x^k, y^k)^\top dx^k + \nabla_y f(x^k, y^k)^\top dy^k - \alpha_k \|\Phi_{\mu_k}(y^k, z^k)\|_1,$

(b) $\theta'_{\mu_k,\alpha_k}(w^k; dw^k) \le -(dw^k)^\top B_k dw^k$

for each k*. Moreover, if* $\Phi_{\mu_k}(y^k, z^k) \neq 0$*, then the inequality in* (b) *holds strictly.*

Proof. We first show (a). Let J_+^k, J_0^k, and $J_-^k \subseteq \{1, 2, \ldots, m\}$ be the index sets defined by

$$\begin{aligned} J_+^k &= \{j \mid \Phi_{\mu_k}(y^k, z^k)_j > 0\}, \\ J_0^k &= \{j \mid \Phi_{\mu_k}(y^k, z^k)_j = 0\}, \\ J_-^k &= \{j \mid \Phi_{\mu_k}(y^k, z^k)_j < 0\}, \end{aligned}$$

where $\Phi_{\mu_k}(y^k, z^k)_j \in \mathbb{R}$ denotes the j-th component of $\Phi_{\mu_k}(y^k, z^k) \in \mathbb{R}^m$. Then, we have

$$\begin{aligned} \theta'_{\mu_k,\alpha_k}(w^k;\ dw^k) &= \nabla f(x^k, y^k)^\top \begin{pmatrix} dx^k \\ dy^k \end{pmatrix} + \alpha_k \sum_{j \in J_+^k} \left[\nabla \Phi_{\mu_k}(y^k, z^k)\right]_j^\top \begin{pmatrix} dy^k \\ dz^k \end{pmatrix} \\ &+ \alpha_k \sum_{j \in J_0^k} \left| \left[\nabla \Phi_{\mu_k}(y^k, z^k)\right]_j^\top \begin{pmatrix} dy^k \\ dz^k \end{pmatrix} \right| - \alpha_k \sum_{j \in J_-^k} \left[\nabla \Phi_{\mu_k}(y^k, z^k)\right]_j^\top \begin{pmatrix} dy^k \\ dz^k \end{pmatrix}, \end{aligned} \tag{5.8}$$

where $[\nabla \Phi_{\mu_k}(y^k, z^k)]_j$ denotes the j-th column vector of $\nabla \Phi_{\mu_k}(y^k, z^k)$. Since

$$\left[\nabla \Phi_{\mu_k}(y^k, z^k)\right]_j^\top \begin{pmatrix} dy^k \\ dz^k \end{pmatrix} = -\Phi_{\mu_k}(y^k, z^k)_j$$

from the constraint of QP subproblem (5.2), we have

$$\theta'_{\mu_k,\alpha_k}(w^k; dw^k) = \nabla_x f(x^k, y^k)^\top dx^k + \nabla_y f(x^k, y^k)^\top dy^k - \alpha_k \|\Phi_{\mu_k}(y^k, z^k)\|_1. \tag{5.9}$$

We next show (b). Taking the inner product of $dw^k = (dx^k, dy^k, dz^k)$ and both sides of the first equality in the KKT conditions (5.3) with $dw = dw^k$, $\eta = \eta^k$, $u = u^k$, $v = v^k$ for the subproblem (5.2), we obtain

$$\begin{aligned} &\nabla f(x^k, y^k)^\top \begin{pmatrix} dx^k \\ dy^k \end{pmatrix} + (dw^k)^\top B_k dw^k + (u^k)^\top (N dx^k + M dy^k - dz^k) \\ &+ (v^k)^\top \nabla \Phi_{\mu_k}(y^k, z^k) \begin{pmatrix} dy^k \\ dz^k \end{pmatrix} + (\eta^k)^\top A dx^k = 0. \end{aligned} \tag{5.10}$$

Moreover, from the constraints of the subproblem (5.2) and the KKT conditions (5.3), we have

$$N dx^k + M dy^k - dz^k = 0, \tag{5.11}$$

$$\nabla \Phi_{\mu_k}(y^k, z^k)^\top \begin{pmatrix} dy^k \\ dz^k \end{pmatrix} = -\Phi_{\mu_k}(y^k, z^k), \tag{5.12}$$

and

$$0 = (\eta^k)^\top (b - Ax^k - Adx^k) = -(\eta^k)^\top Adx^k + (\eta^k)^\top (b - Ax^k) \geq -(\eta^k)^\top Adx^k, \quad (5.13)$$

where the inequality is due to $\eta^k \geq 0$ and $b - Ax^k \geq 0$ from (5.3). Substituting (5.11)–(5.13) into (5.10), we have

$$\nabla f(x^k, y^k)^\top \begin{pmatrix} dx^k \\ dy^k \end{pmatrix} + (dw^k)^\top B_k dw^k - (v^k)^\top \Phi_{\mu_k}(y^k, z^k) \leq 0.$$

Furthermore, from (5.9), we obtain

$$\begin{aligned} \theta'_{\mu_k,\alpha_k}(w^k; dw^k) &\leq -(dw^k)^\top B_k dw^k + (v^k)^\top \Phi_{\mu_k}(y^k, z^k) - \alpha_k \|\Phi_{\mu_k}(y^k, z^k)\|_1 \\ &= -(dw^k)^\top B_k dw^k + \sum_{j \in J_+^k} (v_j^k - \alpha_k) \left[\Phi_{\mu_k}(y^k, z^k)\right]_j \\ &\quad + \sum_{j \in J_-^k} (v_j^k + \alpha_k) \left[\Phi_{\mu_k}(y^k, z^k)\right]_j \\ &\leq -(dw^k)^\top B_k dw^k, \end{aligned}$$

where the last inequality follows from $\alpha_k > \|v^k\|_\infty$ and the definitions of J_+^k and J_-^k. Moreover, if $\Phi_{\mu_k}(y^k, z^k) \neq 0$, then the last inequality holds strictly since $J_+^k \cup J_-^k \neq \emptyset$. This completes the proof of (b). □

6 Convergence Analysis

In this section, we study the convergence property of the proposed algorithm. To begin with, we make the following assumption.

Assumption 1. Let sequences $\{w^k\}$ and $\{B_k\}$ be produced by Algorithm 1.

(a) $\{w^k\}$ is bounded.

(b) There exist constants $\gamma_1, \gamma_2 > 0$ such that $\gamma_1 \|d\|^2 \leq d^\top B_k d \leq \gamma_2 \|d\|^2$ for any $d \in \mathbb{R}^{n+2m}$ and k.

(c) There exists a constant $c > 0$ such that $\|D_k^{-1}\| \leq c$ for any k, where D_k is the matrix defined by (5.7).

Assumption 1 (b) means that $\{B_k\}$ is bounded and uniformly positive definite. Assumption 1 (c) holds if and only if any accumulation point of $\{D_k\}$ is nonsingular. The next proposition provides a sufficient condition under which Assumption 1 (c) holds.

Proposition 6.1. *Suppose that M is a Cartesian P matrix and Assumption 1 (a) holds. Then, Assumption 1 (c) holds.*

Proof. Let $(\overline{E}_y, \overline{E}_z)$ be an arbitrary accumulation point of $\{(\nabla_y \Phi_{\mu_k}(y^k, z^k), \nabla_z \Phi_{\mu_k}(y^k, z^k))\}$. Then, it suffices to show that the matrix

$$D_\infty := \begin{pmatrix} M & -I_m \\ \overline{E}_y & \overline{E}_z \end{pmatrix}$$

is nonsingular. By Proposition 4.7, $\overline{E}_y$ and $\overline{E}_z$ satisfy the following three properties:

(a) $\overline{E}_y + \overline{E}_z = I_m$;

(b) $0 \preceq \overline{E}_y \preceq I_m$ and $0 \preceq \overline{E}_z \preceq I_m$;

(c) $\overline{E}_y$ and $\overline{E}_z$ are symmetric and have the block-diagonal structure conforming to the Cartesian structure of $\mathcal{K} = \mathcal{K}^{m_1} \times \cdots \times \mathcal{K}^{m_\ell}$.

From (a) and (b), there exists an orthogonal matrix $H \in \mathbb{R}^{m\times m}$,

$$H\overline{E}_y H^\top = \mathrm{diag}(\alpha_i)_{i=1}^m,\ H\overline{E}_z H^\top = \mathrm{diag}(1-\alpha_i)_{i=1}^m,\quad 0 \le \alpha_i \le 1\ (i=1,2,\dots,m), \tag{6.1}$$

where α_i $(i = 1, 2, \dots, m)$ are the eigenvalues of $\overline{E}_y$. Moreover, from (c), H has the same block-diagonal structure as $\overline{E}_y$ and $\overline{E}_z$. Hence, $\tilde{M} := HMH^\top$ is a Cartesian P matrix by Proposition 2.5.

Now, let $\tilde{D}_\infty \in \mathbb{R}^{2m\times 2m}$ be defined as

$$\tilde{D}_\infty := \begin{pmatrix} H & 0 \\ 0 & H \end{pmatrix} D_\infty \begin{pmatrix} H^\top & 0 \\ 0 & H^\top \end{pmatrix} = \begin{pmatrix} \tilde{M} & -I_m \\ \mathrm{diag}(\alpha_i)_{i=1}^m & \mathrm{diag}(1-\alpha_i)_{i=1}^m \end{pmatrix}, \tag{6.2}$$

and let $(\zeta, \eta) \in \mathbb{R}^m \times \mathbb{R}^m$ be an arbitrary vector such that $\tilde{D}_\infty \binom{\zeta}{\eta} = 0$. Then, we have

$$\tilde{M}\zeta = \eta, \tag{6.3}$$
$$\alpha_i\zeta_i + (1-\alpha_i)\eta_i = 0 \quad (i = 1, 2, \dots, m). \tag{6.4}$$

If $\alpha_i = 0$, then we have $(\tilde{M}\zeta)_i = \eta_i = 0$ from (6.3) and (6.4). If $\alpha_i = 1$, then we have $\zeta_i = 0$ from (6.4). If $0 < \alpha_i < 1$, then we have $\zeta_i(\tilde{M}\zeta)_i = \zeta_i\eta_i = -\alpha_i(1-\alpha_i)^{-1}\zeta_i^2 \le 0$. Thus, for all i, we have $\zeta_i(\tilde{M}\zeta)_i \le 0$. Since $\tilde{M}$ is a Cartesian P matrix and every Cartesian P matrix is a P matrix, we must have $\zeta = \eta = 0$. Hence, $\tilde{D}_\infty$ is nonsingular. From (6.2) and the nonsingularity of H, matrix D_∞ is also nonsingular. $\square$

The following three lemmas play crucial roles in establishing the convergence theorem for the algorithm.

Lemma 6.2. *Let $\{w^k\}$ be a sequence generated by Algorithm 1, and $dw^k := (dx^k, dy^k, dz^k) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ be the unique optimum of QP subproblem (5.2) for each k. Let $\{(\mu_k, \tau_k)\} \subseteq \mathbb{R}_{++} \times \mathbb{R}_{++}$ be a sequence converging to $(0, 0)$ and $\alpha > 0$ be a fixed scalar. Suppose that $\{w^k\}$ satisfies Assumption 1(a). In addition, assume that $\{dw^k\}$ has an accumulation point, and $\overline{w} := (\overline{x}, \overline{y}, \overline{z})$ and $\overline{dw} := (\overline{dx}, \overline{dy}, \overline{dz})$ are arbitrary accumulation points of $\{w^k\}$ and $\{dw^k\}$, respectively. Then we have*

$$\limsup_{k\to\infty} \left(\frac{\theta_{\mu_k,\alpha}(w^k + \tau_k dw^k) - \theta_{\mu_k,\alpha}(w^k)}{\tau_k} - \theta'_{\mu_k,\alpha}(w^k; dw^k) \right) \le 0$$

provided $\overline{y} - \overline{z} \notin \mathrm{bd}\,(\mathcal{K} \cup -\mathcal{K})$ and $\overline{dw} \ne 0$.

Proof. Taking subsequences if necessary, we may suppose $w^k \to \overline{w}$ and $dw^k \to \overline{dw}$. Moreover, we have $\theta'_{\mu_k,\alpha}(w^k; dw^k) = \nabla_x f(x^k, y^k)^\top dx^k + \nabla_y f(x^k, y^k)^\top dy^k - \alpha\|\Phi_{\mu_k}(y^k, z^k)\|_1$ as shown in (5.9). Thus, we have only to show

$$\limsup_{k\to\infty} \left(\frac{\theta_{\mu_k,\alpha}(w^k + \tau_k dw^k) - \theta_{\mu_k,\alpha}(w^k)}{\tau_k} \right) \le \nabla f(\overline{x}, \overline{y})^\top \begin{pmatrix} \overline{dx} \\ \overline{dy} \end{pmatrix} - \alpha\|\Phi(\overline{y}, \overline{z})\|_1.$$

From the mean-value theorem and the continuity of ∇f, we have

$$\lim_{k\to\infty} \frac{f(x^k+\tau_k dx^k, y^k+\tau_k dy^k) - f(x^k,y^k)}{\tau_k} = \nabla f(\overline{x},\overline{y})^\top \begin{pmatrix} \overline{dx} \\ \overline{dy} \end{pmatrix}.$$

Therefore, it suffices to show that

$$\limsup_{k\to\infty} \frac{|\Phi_{\mu_k}(y^k+\tau_k dy^k, z^k+\tau_k dz^k)_j| - |\Phi_{\mu_k}(y^k,z^k)_j|}{\tau_k} \le -|\Phi(\overline{y},\overline{z})_j| \tag{6.5}$$

for each j. By Definition 5 and Proposition 4.7, we have $\nabla_y \Phi_{\mu_k}(y^k,z^k) = I_m - \nabla P_{\mu_k}(y^k - z^k)$ and $\nabla_z \Phi_{\mu_k}(y^k,z^k) = \nabla P_{\mu_k}(y^k - z^k)$, which together with the constraints of QP subproblem (5.2) yield

$$\begin{aligned} -\Phi_{\mu_k}(y^k,z^k) &= \nabla_y \Phi_{\mu_k}(y^k,z^k)^\top dy^k + \nabla_z \Phi_{\mu_k}(y^k,z^k)^\top dz^k \\ &= (I - \nabla P_{\mu_k}(y^k-z^k)^\top) dy^k + \nabla P_{\mu_k}(y^k-z^k)^\top dz^k \\ &= dy^k - \nabla P_{\mu_k}(y^k-z^k)^\top (dy^k - dz^k). \end{aligned} \tag{6.6}$$

Hence, we have

$$\begin{aligned} & \Phi_{\mu_k}(y^k+\tau_k dy^k, z^k+\tau_k dz^k) - \Phi_{\mu_k}(y^k,z^k) \\ = \; & \tau_k dy^k - \big(P_{\mu_k}(y^k+\tau_k dy^k - (z^k+\tau_k dz^k)) - P_{\mu_k}(y^k-z^k)\big) \\ = \; & \tau_k \left(-\Phi_{\mu_k}(y^k,z^k) + \nabla P_{\mu_k}(y^k-z^k)^\top (dy^k - dz^k) \right) \\ & -P_{\mu_k}(y^k - z^k + \tau_k(dy^k - dz^k)) + P_{\mu_k}(y^k-z^k) \\ = \; & -\tau_k \Phi_{\mu_k}(y^k,z^k) + \tau_k \delta^k, \end{aligned} \tag{6.7}$$

where the first equality is due to (4.3), the second equality follows from (6.6), and $\delta^k := (\delta_1^k, \delta_2^k, \ldots, \delta_m^k) \in \mathbb{R}^m$ is given by

$$\delta_j^k := \nabla P_{\mu_k}(y^k-z^k)_j^\top (dy^k - dz^k) - \tau_k^{-1} \big(P_{\mu_k}(y^k - z^k + \tau_k(dy^k-dz^k))_j - P_{\mu_k}(y^k-z^k)_j\big). \tag{6.8}$$

To show (6.5), we consider three cases: (i) $\Phi(\overline{y},\overline{z})_j = 0$, (ii) $\Phi(\overline{y},\overline{z})_j > 0$ and (iii) $\Phi(\overline{y},\overline{z})_j < 0$. In case (i), we first notice that

$$\begin{aligned} & \frac{|\Phi_{\mu_k}(y^k+\tau_k dy^k, z^k+\tau_k dz_k)_j| - |\Phi_{\mu_k}(y^k,z^k)_j|}{\tau_k} \\ & \le \frac{|\Phi_{\mu_k}(y^k+\tau_k dy^k, z^k+\tau_k dz_k)_j - \Phi_{\mu_k}(y^k,z^k)_j|}{\tau_k} \\ & = |\Phi_{\mu_k}(y^k,z^k)_j - \delta_j^k|, \end{aligned} \tag{6.9}$$

where the equality follows from (6.7). By applying the mean-value theorem in (6.8), we can find $\zeta_{kj} \in (0,1)$ for each k such that

$$\delta_j^k = \big(\nabla P_{\mu_k}(y^k-z^k) - \nabla P_{\mu_k}\big(y^k - z^k + \zeta_{kj}\tau_k(dy^k-dz^k)\big)\big)_j^\top (dy^k - dz^k). \tag{6.10}$$

Since Proposition 4.5 and the boundedness of $\{dw^k\}$ imply

$$\lim_{k\to\infty} \nabla P_{\mu_k}(y^k-z^k) - \nabla P_{\mu_k}(y^k - z^k + \zeta_{kj}\tau_k(dy^k - dz^k)) = 0, \tag{6.11}$$

we obtain $\lim_{k\to\infty} \delta_j^k = 0$. Moreover, by Proposition 4.8, we have $\lim_{k\to\infty} \Phi_{\mu_k}(y^k, z^k)_j = \Phi(\overline{y}, \overline{z})_j = 0$. Then, by letting $k \to \infty$ in (6.9), we obtain (6.5). In case (ii) and case (iii), we have

$$|\Phi_{\mu_k}(y^k + \tau_k dy^k, z^k + \tau_k dz^k)_j| - |\Phi_{\mu_k}(y^k, z^k)_j| = -\Phi_{\mu_k}(y^j, z^k)_j + \delta_j^k$$

and

$$|\Phi_{\mu_k}(y^k + \tau_k dy^k, z^k + \tau_k dz^k)_j| - |\Phi_{\mu_k}(y^k, z^k)_j| = \Phi_{\mu_k}(y^j, z^k)_j - \delta_j^k,$$

respectively. Then, a similar argument to that in case (i) leads to the desired inequality (6.5). □

Lemma 6.3. *Let $\{w^k\}$ be a sequence generated by Algorithm 1. Suppose that Assumption 1 holds. Then, we have the following statements.*

(i) *$\{dw^k\}$ and $\{(u^k, v^k)\}$ are bounded.*

(ii) *There exists k_0 such that $\alpha_k = \alpha_{k_0}$ for all $k \geq k_0$.*

(iii) *The sequences $\{\theta_{\mu_k,\alpha_k}(w^k)\}$ and $\{\theta_{\mu_k,\alpha_k}(w^{k+1})\}$ converge to the same limit.*

Proof. We first prove (i). Let

$$\begin{pmatrix} d\tilde{y}^k \\ d\tilde{z}^k \end{pmatrix} := -D_k^{-1} \begin{pmatrix} 0 \\ \Phi_{\mu_k}(y^k, z^k) \end{pmatrix} \tag{6.12}$$

and $d\tilde{w}^k := (0,\ d\tilde{y}^k,\ d\tilde{z}^k)$, where D_k is defined by (5.7). Then, $d\tilde{w}^k$ is a feasible point of QP subproblem (5.2). Note that the objective function of QP subproblem (5.2) is rewritten as

$$\frac{1}{2}(dw - B_k^{-1}g^k)^\top B_k(dw - B_k^{-1}g^k) + \text{constant},$$

where $g^k := (\nabla f(x^k, y^k), 0)$. Since dw^k is the optimum of QP subproblem (5.2), we have

$$(d\tilde{w}^k - B_k^{-1}g^k)^\top B_k(d\tilde{w}^k - B_k^{-1}g^k) \geq (dw^k - B_k^{-1}g^k)^\top B_k(dw^k - B_k^{-1}g^k).$$

This together with Assumption 1 (b) implies

$$\gamma_2\|d\tilde{w}^k - B_k^{-1}g^k\|^2 \geq \gamma_1\|dw^k - B_k^{-1}g^k\|^2. \tag{6.13}$$

Now, notice that $\{d\tilde{w}^k\}$ is bounded from (6.12), Assumption 1 (a), (c) and Proposition 4.8. In addition, $\{B_k^{-1}\}$ and $\{g^k\}$ are also bounded from Assumption 1 (a), (b). Thus, we have the boundedness of $\{dw^k\}$ from (6.13). On the other hand, from the first equality of the KKT conditions (5.3), we have

$$\begin{pmatrix} u^k \\ v^k \end{pmatrix} = -(D_k^\top)^{-1}\left(\begin{pmatrix} \nabla_y f(x^k, y^k) \\ 0 \end{pmatrix} + \tilde{B}_k dw^k\right),$$

where $\tilde{B}_k$ is the $2m \times (n + 2m)$ matrix consisting of the last $2m$ rows of B_k. This equation together with Assumption 1 and the boundedness of $\{dw^k\}$ yields the boundedness of $\{(u^k, v^k)\}$.

We next prove (ii). From the update rule (5.5), we can easily see that $\{\alpha_k\}$ is nondecreasing. Moreover, if

$$\|v^k\|_\infty > \alpha_{k-1} - \delta, \tag{6.14}$$

then we have $\alpha_k = \max\{\|v^k\|_\infty + \delta, \alpha_{k-1} + 2\delta\} \geq \alpha_{k-1} + 2\delta$, that is, α_k increases at least by 2δ at a time. Let $\hat{K} := \{k | \|v^k\|_\infty > \alpha_{k-1} - \delta\}$. If $|\hat{K}| = \infty$, then $\alpha_k \to \infty$ as $k \to \infty$ and hence $\{\|v^k\|_\infty\}$ is unbounded from (6.14). However this contradicts (i). Thus we have (ii).

We finally show (iii). Since we have (ii), there exist $\overline{\alpha}$ and k_0 such that $\overline{\alpha} = \alpha_k$ for all $k \geq k_0$. In what follows, we suppose $k \geq k_0$. Since $\mu_{k+1} \leq \mu_k$, Proposition 4.9 together with (5.4) implies

$$\begin{aligned} \theta_{\mu_{k+1},\overline{\alpha}}(w^{k+1}) + \overline{\alpha} m\rho\mu_{k+1} &\leq \theta_{\mu_k,\overline{\alpha}}(w^{k+1}) + \overline{\alpha} m\rho\mu_k & (6.15)\\ &\leq \theta_{\mu_k,\overline{\alpha}}(w^k) + \overline{\alpha} m\rho\mu_k, & (6.16) \end{aligned}$$

where the last inequality follows from the Armijo condition (5.6) and Proposition 5.3 (b). From (6.16), $\{\theta_{\mu_k,\overline{\alpha}}(w^k) + \overline{\alpha} m\rho\mu_k\}$ is a monotonically nonincreasing sequence. In addition, $\{\theta_{\mu_k,\overline{\alpha}}(w^k) + \overline{\alpha} m\rho\mu_k\}$ is bounded, since $\{\theta_{\mu_k,\overline{\alpha}}(w^k)\}$ is bounded from Assumption 1 (a). Therefore, $\{\theta_{\mu_k,\overline{\alpha}}(w^k) + \overline{\alpha} m\rho\mu_k\}$ is convergent. This fact together with $\lim_{k\to\infty} \mu_k = 0$ and (6.16) yields that $\{\theta_{\mu_k,\overline{\alpha}}(w^{k+1})\}$ and $\{\theta_{\mu_k,\overline{\alpha}}(w^k)\}$ must converge to the same limit. □

Finally, we show that a sequence generated by the algorithm globally converges to B-stationary point of MPSOCC (3.1) under the assumption that any accumulation point $\overline{w} = (\overline{x}, \overline{y}, \overline{z})$ satisfies $\overline{y} - \overline{z} \notin \mathrm{bd}\,(\mathcal{K} \cup -\mathcal{K})$, which is equivalent to the nondegeneracy condition given in Definition 3.1 if $\overline{y}$ and $\overline{z}$ satisfy the SOC complementarity condition.

Theorem 6.1. Let $\{w^k\}$ be a sequence generated by Algorithm 1. Suppose that $\{w^k\}$ satisfies Assumption 1. Let $\overline{w} = (\overline{x}, \overline{y}, \overline{z})$ be an arbitrary accumulation point of $\{w^k\}$. If $(\overline{y}, \overline{z})$ satisfies $\overline{y} - \overline{z} \notin \mathrm{bd}\,(\mathcal{K} \cup -\mathcal{K})$, then $\overline{w}$ is a B-stationary point of MPSOCC (1.1).

Proof. By Lemma 6.3(ii), there exists some constant $\alpha > 0$ and k_0 such that $\alpha_k = \alpha$ for all $k \geq k_0$. In this proof, we suppose without loss of generality that $\alpha_k = \alpha$ holds for all k.

We first show that

$$\lim_{k\to\infty} \|dw^k\| = 0. \tag{6.17}$$

From Proposition 5.3 (b) and Assumption 1 (b), we have

$$\theta'_{\mu_k,\alpha}(w^k; dw^k) \leq -(dw^k)^\top B_k dw^k \leq -\gamma_1 \|dw^k\|^2, \tag{6.18}$$

which together with Armijo condition (5.6) yields

$$\begin{aligned} \theta_{\mu_k,\alpha}(w^{k+1}) &\leq \theta_{\mu_k,\alpha}(w^k) + \sigma\tau_k \theta'_{\mu_k,\alpha}(w^k; dw^k)\\ &\leq \theta_{\mu_k,\alpha}(w^k) - \gamma_1\sigma\tau_k \|dw^k\|^2. \end{aligned}$$

Hence, from Lemma 6.3 (iii), we obtain

$$\lim_{k\to\infty} \tau_k \|dw^k\|^2 = 0.$$

Now, assume for contradiction that (6.17) does not hold. Then, there exists an infinite index set $K \subseteq \{0, 1, \ldots\}$ such that

$$\lim_{\substack{k\to\infty\\ k\in K}} \|dw^k\| > 0, \tag{6.19}$$

and hence

$$\lim_{\substack{k\to\infty\\ k\in K}} \tau_k = 0.$$

Let ℓ_k be the smallest nonnegative integer L satisfying (5.6), i.e., $\rho^{\ell_k} = \tau_k$. Then, from the definition of ℓ_k, we have

$$\theta_{\mu_k,\alpha}(w^k + \rho^{\ell_k - 1} dw^k) > \theta_{\mu_k,\alpha}(w^k) + \sigma\rho^{\ell_k-1}\theta'_{\mu_k,\alpha}(w^k; dw^k),$$

which implies

$$\xi_k := \frac{\theta_{\mu_k,\alpha}(w^k + \rho^{\ell_k-1}dw^k) - \theta_{\mu_k,\alpha}(w^k)}{\rho^{\ell_k-1}} - \theta'_{\mu_k,\alpha}(w^k; dw^k) > -(1-\sigma)\theta'_{\mu_k,\alpha}(w^k; dw^k) \quad (6.20)$$

By Lemma 6.2 together with $\lim_{k\to\infty,\ k\in K} \rho^{\ell_k-1}=0$ and $\overline{y} - \overline{z} \notin \operatorname{bd}(\mathcal{K} \cup -\mathcal{K})$, we have $\limsup_{k\to\infty,\ k\in K} \xi_k \leq 0$. Moreover, we have from (6.18)

$$-(1-\sigma)\theta'_{\mu_k,\alpha}(w^k; dw^k) \geq (1-\sigma)\gamma_1 \|dw^k\|^2. \quad (6.21)$$

From (6.20) and (6.21), we must have $\lim_{k\to\infty,\ k\in K} \|dw^k\| = 0$. However, this contradicts (6.19), and hence we have (6.17).

Next, we show that $\overline{w}$ satisfies the KKT conditions (3.2) of MPSOCC (3.1). Let $\{(\eta^k, u^k, v^k)\}$ be the sequence of multipliers corresponding to $\{dw^k\}$. Then, $\{(u^k, v^k)\}$ is bounded from Lemma 6.3 (i). Hence, there exist vectors $\overline{u}$, $\overline{v}$ and an index set K' such that $\lim_{k\to\infty,k\in K'}(w^k, u^k, v^k) = (\overline{w}, \overline{u}, \overline{v})$. By (5.3) and letting $X := \{x \in \mathbb{R}^n \mid Ax \leq b\}$ we have

$$\zeta^k \in -\mathcal{N}_X(x^k) \times \{0\}^{2m}, \quad (6.22)$$

$$\begin{pmatrix} N & M & -I_m \\ 0 & \nabla_y \Phi_{\mu_k}(y^k, z^k)^\top & \nabla_z \Phi_{\mu_k}(y^k, z^k)^\top \end{pmatrix} \begin{pmatrix} dx^k \\ dy^k \\ dz^k \end{pmatrix} = -\begin{pmatrix} 0 \\ \Phi_{\mu_k}(y^k, z^k) \end{pmatrix}, \quad (6.23)$$

$$0 \leq b - Ax^k - Adx^k, \quad (6.24)$$

where

$$\zeta^k := \begin{pmatrix} \nabla_x f(x^k, y^k) \\ \nabla_y f(x^k, y^k) \\ 0 \end{pmatrix} + B_k \begin{pmatrix} dx^k \\ dy^k \\ dz^k \end{pmatrix} + \begin{pmatrix} N^\top \\ M^\top \\ -I_m \end{pmatrix} u^k + \begin{pmatrix} 0 \\ \nabla_y \Phi_{\mu_k}(y^k, z^k) \\ \nabla_z \Phi_{\mu_k}(y^k, z^k) \end{pmatrix} v^k, \quad (6.25)$$

and (6.22) follows from the first equation of (5.3) with $\mathcal{N}_X(x^k) = \{A^\top \eta \mid \eta \geq 0,\ \eta^\top(Ax^k - b) = 0\}$. By letting $k \in K'$ tend to ∞ in (6.23) and (6.24), we have

$$0 \leq b - A\overline{x}, \quad \Phi(\overline{y}, \overline{z}) = 0. \quad (6.26)$$

Note that (6.26) implies $\overline{x} \in X$. In addition, note that $\{\zeta^k\}_{k\in K'}$ is a convergent sequence satisfying (6.22), since $\{B_k\}$ is bounded from Assumption 1 and $\lim_{k\to\infty} \nabla\Phi_{\mu_k}(y^k, w^k) = \nabla\Phi(\overline{y}, \overline{w})$ from the nondegeneracy condition. Then, these facts together with (6.17) and the closedness of the point-to-set map $\mathcal{N}_X(\cdot)$ yield

$$\begin{pmatrix} \nabla_x f(\overline{x}, \overline{y}) \\ \nabla_y f(\overline{x}, \overline{y}) \\ 0 \end{pmatrix} + \begin{pmatrix} N^T \\ M^T \\ -I_m \end{pmatrix} \overline{u} + \begin{pmatrix} 0 \\ \nabla_y \Phi(\overline{y}, \overline{z}) \\ \nabla_z \Phi(\overline{y}, \overline{z}) \end{pmatrix} \overline{v} = \lim_{k\to\infty,k\in K'} \zeta^k \in -\mathcal{N}_X(\overline{x}) \times \{0\}^{2m}.$$

This together with (6.26) means that $\overline{w}$ satisfies the KKT conditions (3.2) of MPSOCC (3.1). By Proposition 3.7, $\overline{w}$ is a B-stationary point of MPSOCC (1.1). $\square$

7 Numerical Experiments

In this section, we implement Algorithm 1 for solving problem (1.1) and report some numerical results. The program is coded in Matlab 2008a and run on a machine with an Intel®Core2 Duo E6850 3.00GHz CPU and 4GB RAM. In Step 0 of the algorithm, we set the parameters as

$$\delta := 1, \quad \alpha_{-1} := 10, \quad \sigma := 10^{-3}, \quad \rho := 0.9.$$

The choice of smoothing parameters $\{\mu_k\}$, i.e., μ_0 and β, and a starting point w^0 vary with the experiment. We let B_0 be the identity matrix, and update B_k by the modified BFGS formula:

$$B_{k+1} \quad := \quad B_k - \frac{B_k s^k (B_k s^k)^\top}{(s^k)^\top B_k s^k} + \frac{\zeta^k (\zeta^k)^\top}{(s^k)^\top \zeta^k}$$

with $s^k = w^{k+1} - w^k$ and $\zeta^k = \theta_k \tilde{\zeta}^k + (1-\theta_k) B_k s^k$, where $\tilde{\zeta}^k := \nabla_w L_{\mu_{k+1}}(w^{k+1}, u^k, v^k, \eta^k) - \nabla_w L_{\mu_k}(w^k, u^k, v^k, \eta^k)$, L_μ denotes the Lagrangian function defined by $L_\mu(w, u, v, \eta) := f(x, y) + \Phi_\mu(y, z)v + (Nx + My + q - z)u + (Ax - b)\eta$, and θ_k is determined by

$$\theta_k := \begin{cases} 1 & \text{if } (s^k)^\top \tilde{\zeta}^k \geq 0.2 (s^k)^\top B_k s^k \\ \dfrac{0.8 (s^k)^\top B_k s^k}{(s^k)^\top (B_k s^k - \tilde{\zeta}^k)} & \text{otherwise.} \end{cases}$$

In Step 1, we use the *quadprog* solver in Matlab Optimization Toolbox for solving the QP subproblems. In Step 3, we terminate the algorithm if the following condition is satisfied:

$$\|\Phi(y^k, z^k)\|_\infty + \|dw^k\|_\infty \leq 10^{-7}. \tag{7.1}$$

The rationale for using (7.1) is as follows. If $\Phi(y^k, z^k) = 0$, i.e., $\mathcal{K} \ni y^k \perp z^k \in \mathcal{K}$ holds, then $w^k = (x^k, y^k, z^k)$ is feasible to MPSOCC (1.1), since the remaining constraints $Ax^k \leq b$ and $z^k = Nx^k + My^k + q$ always hold from Proposition 5.2. Moreover, if $\|dw^k\|_\infty = 0$, then w^k satisfies the KKT conditions (3.2) of MPSOCC (3.1). Thus, by Theorem 6.1, $\|\Phi(y^k, z^k)\|_\infty + \|dw^k\|_\infty = 0$ indicates that w^k is a B-stationary point under the assumption $y^k - z^k \notin \mathrm{bd}\,(\mathcal{K} \cup -\mathcal{K})$, which is equivalent to the nondegeneracy condition in Definition 3.1 if $\mathcal{K} \ni y^k \perp z^k \in \mathcal{K}$. Hence, (7.1) is appropriate for a stopping criterion of the algorithm. As the CM function, we choose $\hat{g}(\alpha) := ((\alpha^2 + 4)^{1/2} + \alpha)/2$.

Experiment 1

In the first experiment, we solve the following test problem of the form (1.1):

$$\begin{array}{ll} \text{Minimize} & \|x\|^2 + \|y\|^2 \\ \text{subject to} & Ax \leq b, \\ & z = Nx + My + q, \\ & \mathcal{K} \ni y \perp z \in \mathcal{K}, \end{array} \tag{7.2}$$

where $(x, y, z) \in \mathbb{R}^{10} \times \mathbb{R}^m \times \mathbb{R}^m$, and each element of $A \in \mathbb{R}^{10\times 10}$, $N \in \mathbb{R}^{m\times 10}$ is randomly chosen from $[-1, 1]$. Moreover, each element of $b \in \mathbb{R}^{10}$ is randomly chosen from $[0, 1]$. In addition, $M \in \mathbb{R}^{m\times m}$ is a positive semi-definite symmetric matrix generated by $M = M_1 M_1^\top + 0.01I$, and $M_1 \in \mathbb{R}^{m\times m}$ is a matrix whose entries are randomly chosen from

$[-1,1]$. The vector $q \in \mathbb{R}^m$ is set to be $q := \xi_z - M\xi_y$ with $\xi_y \in \mathbb{R}^m$ and $\xi_z \in \mathbb{R}^m$, whose components are randomly chosen from $[-1,1]$. We choose different Cartesian structures for $\mathcal{K}$, and generate 50 problems for each $\mathcal{K}$. In applying Algorithm 1, we set an initial point $w^0 = (x^0, y^0, z^0) := (0, \xi_y, \xi_z) \in \mathbb{R}^{10} \times \mathbb{R}^m \times \mathbb{R}^m$, so that $Ax^0 \leq b$ and $z^0 = Nx^0 + My^0 + q$ are satisfied. We choose smoothing parameters $\{\mu_k\}$ as $\mu_k := 100 \times 0.8^k$.

The obtained results are shown in Tables 1 and 2, where $(\mathcal{K}^\nu)^\kappa := \mathcal{K}^\nu \times \mathcal{K}^\nu \times \cdots \times \mathcal{K}^\nu \subseteq \mathbb{R}^{\nu\kappa}$ and each column represents the following:

- $\sharp$ite: the average number of iterations among 50 test problems for each $\mathcal{K}$;
- cpu(s): the average cpu-time in second among 50 test problems for each $\mathcal{K}$;
- non(%): percentage of test problems whose solutions obtained by Algorithm 1 satisfy the nondegeneracy condition in Definition 3.1.

Recall that convergence to a B-stationary point is proved under the nondegeneracy condition. Hence, the value of "non" represents the percentage of problems for which the algorithm successfully finds B-stationary points. From Table 1, we can observe that $\sharp$ite does not change so much although the values of cpu(s) tends to be larger as m increases. From Table 2, we can see that non(%) tends to be less than 100 if $\mathcal{K}$ includes $\mathcal{K}^1$ or $\mathcal{K}^2$. Indeed, when $\mathcal{K} = (\mathcal{K}^1)^{100}$ and $(\mathcal{K}^2)^{50}$, the values of "non(%)" is 74 and 86, respectively, whereas it becomes 100 when the dimension of all SOCs in $\mathcal{K}$ is larger than 10.

m	$\mathcal{K}$	$\sharp$ite	cpu(s)	non(%)
10	$\mathcal{K}^{10}$	57.42	0.400	100
20	$\mathcal{K}^{20}$	56.30	0.471	100
30	$\mathcal{K}^{30}$	55.78	0.568	100
40	$\mathcal{K}^{40}$	55.44	0.726	100
50	$\mathcal{K}^{50}$	55.06	0.987	100
60	$\mathcal{K}^{60}$	54.96	1.388	100
70	$\mathcal{K}^{70}$	54.74	1.797	100
80	$\mathcal{K}^{80}$	54.66	2.130	100
90	$\mathcal{K}^{90}$	54.40	2.437	100
100	$\mathcal{K}^{100}$	54.20	2.930	100

Table 1: Results for problems with a single SOC complementarity constraint (Experiment 1)

Experiment 2.

In the second experiment, we apply Algorithm 1 to a bilevel programming problem with a robust optimization problem in the *lower-level*. Bilevel programming has wide application such as network design and production planning [2, 9]. On the other hand, robust optimization is known to be a powerful methodology to treat optimization problems with uncertain

m	$\mathcal{K}$	♯ite	cpu(s)	non(%)
100	$\mathcal{K}^{100}$	54.20	2.930	100
100	$(\mathcal{K}^{50})^2$	55.18	3.037	100
100	$\mathcal{K}^{50} \times \mathcal{K}^{20} \times \mathcal{K}^{30}$	56.28	3.016	100
100	$(\mathcal{K}^{10})^{10}$	64.10	3.687	100
100	$\mathcal{K}^{50} \times \mathcal{K}^{20} \times (\mathcal{K}^{10})^2 \times \mathcal{K}^5 \times (\mathcal{K}^1)^5$	78.22	4.558	98
100	$(\mathcal{K}^2)^{50}$	78.68	6.012	86
100	$(\mathcal{K}^1)^{100}$	87.84	6.685	74

Table 2: Results for problems with multiple SOC complementarity constraints (Experiment 1)

data [3, 4]. In this experiment, we solve the following problem:

$$\begin{array}{ll} \underset{(x,y)\in\mathbb{R}^4\times\mathbb{R}^4}{\text{Minimize}} & \|x - Cy\|^2 + \sum_{i=1}^4 x_i \\ \text{subject to} & 0 \le x_i \le 5 \ (i = 1, 2, 3, 4), \\ & 1 \le -x_1 + 2x_2 + x_4 \le 3, \\ & 1 \le x_2 + x_3 - x_4 \le 2, \\ & y \text{ solves } P(x), \end{array} \tag{7.3}$$

with

$$P(x): \ \underset{y\in\mathbb{R}^4}{\text{Minimize}} \ \max_{\tilde{x}\in U_r(x)} \tilde{x}^\top y + \frac{1}{2} y^\top M y,$$

where $r \ge 0$ is an uncertainty parameter, $U_r(x) \subseteq \mathbb{R}^4$ is an uncertainty set defined by $U_r(x) := \{\tilde{x} \in \mathbb{R}^4 \mid \|\tilde{x} - x\| \le r\}$, and

$$M := \begin{pmatrix} 2 & 2 & 0 & -1 \\ 2 & 4 & -2 & 0 \\ 0 & -2 & 2 & 0 \\ -1 & 0 & 0 & 6 \end{pmatrix}, \ C := \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 2 & 2 & 3 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

For solving problem (7.3), we introduce an auxiliary variable $\gamma \in \mathbb{R}$ to reformulate the lower-level minimax problem $P(x)$ as the following SOCP:

$$\begin{array}{ll} \underset{(\gamma,y)\in\mathbb{R}\times\mathbb{R}^4}{\text{Minimize}} & \frac{1}{2} y^\top M y + x^\top y + r\gamma \\ \text{subject to} & \begin{pmatrix} \gamma \\ y \end{pmatrix} \in \mathcal{K}^5. \end{array}$$

Furthermore, the above SOCP can be rewritten as the following SOC complementarity problem:

$$\mathcal{K}^5 \ni \begin{pmatrix} \gamma \\ y \end{pmatrix} \perp \begin{pmatrix} r \\ My + x \end{pmatrix} \in \mathcal{K}^5.$$

Thus, we can convert problem (7.3) to the following problem:

$$\begin{array}{ll} \underset{(x,y,z,\gamma)\in\mathbb{R}^4\times\mathbb{R}^4\times\mathbb{R}^5\times\mathbb{R}}{\text{Minimize}} & \|x - Cy\|^2 + \sum_{i=1}^4 x_i \\ \text{subject to} & 0 \le x_i \le 5 \ (i = 1,2,3,4), \\ & 1 \le -x_1 + 2x_2 + x_4 \le 3, \\ & 1 \le x_2 + x_3 - x_4 \le 2, \\ & z = \begin{pmatrix} r \\ My + x \end{pmatrix},\ \mathcal{K}^5 \ni \begin{pmatrix} \gamma \\ y \end{pmatrix} \perp z \in \mathcal{K}^5, \end{array} \tag{7.4}$$

which is of the form (1.1). For the sake of comparison, problem (7.4) is solved not only by Algorithm 1, but also by the smoothing method [20], which is described as follows:

Smoothing method

Step 0. Choose a positive sequence $\{\tau_\ell\}$ such that $\tau_\ell \to 0$. Set $\ell := 0$.

Step 1. Find a stationary point $w^\ell = (x^\ell, y^\ell, z^\ell)$ of the smoothed problem (5.1) with $\mu = \tau_\ell$.

Step 2. If w^ℓ is feasible for MPSOCC (1.1), then stop. Otherwise, set $\ell := \ell + 1$ and go to Step 1.

Each algorithm is implemented in the following way: In Step 0 of Algorithm 1, we set smoothing parameters $\mu_k := 0.8^k$ $(k \ge 0)$ and a starting point $x^0 := (1,1,1,1), (\gamma_0, y^0, z^0) := (0,0,0) \in \mathbb{R} \times \mathbb{R}^4 \times \mathbb{R}^5$. The choice of the other parameters is the same as in Experiment 1. In Step 0 of the smoothing method, we set $\tau_\ell := 0.8^\ell$ for $\ell \ge 0$. In Step 1, for solving problem (5.1) with $\mu = \tau_\ell$, we apply Algorithm 1 with slight modification, where the smoothing parameter μ_k is fixed to τ_ℓ for all k and the termination criterion is replaced by $\|dw^k\|_\infty \le 10^{-7}$. In Step 2, we stop the smoothing method when $\|\Phi(y^\ell, z^\ell)\|_\infty \le 10^{-7}$ is satisfied.[‡]

We then test both the methods to problem (7.4) with $r = 0.02, 0.04, 0.06, 0.08$ and 0.10. The obtained results are shown in Tables 3 and 4, whose columns represent the following:

- (x^*, y^*, γ^*): the value of x, y, γ obtained by Algorithm 1;
- $(\lambda_1^*, \lambda_2^*)$: spectral values of $\binom{\gamma^*}{y^*} + z^*$ with respect to $\mathcal{K}^5$ defined as in Definition 2.1, where $z^* := (r, M\binom{\gamma^*}{y^*} + x^*)$;
- $\sharp\text{ite}_{\text{out}}$: the number of outer iterations;
- $\sharp$QP: the number of QP-subproblems (5.2) solved in each trial.

From Table 3, for all r, we can observe $(\lambda_1^*, \lambda_2^*) > 0$, which means $\binom{\gamma^*}{y^*} + z^* \in \operatorname{int} \mathcal{K}^5$, i.e., the nondegeneracy condition holds at the obtained solution. Hence, Algorithm 1 finds a B-stationary point of problem (7.4) successfully. From Table 4, we cannot find a significant difference between the values of $\sharp\text{ite}_{\text{out}}$ for the two methods. However, it is observed that the value of $\sharp$QP in the smoothing method tends to be much larger than that in Algorithm 1. Indeed, when $r = 0.02$, the smoothing method has $\sharp\text{QP} = 218$, which is almost five times larger than $\sharp\text{QP} = 44$ in Algorithm 1. This fact suggests that the smoothing method needs to solve a number of QP-subproblems in Step 1 for solving each smoothed problem (5.1)

[‡]The remaining constraints $Ax^\ell \le b$ and $z^\ell = Nx^\ell + My^\ell + q$ are automatically satisfied since w^ℓ is feasible to the smoothed problem (5.1).

with fixed μ, while Algorithm 1 only solves one QP subproblem (5.2) for each smoothed problem (5.1). As a result, the computational cost in the smoothing method tends to be larger than Algorithm 1.

r	x^*	y^*	γ^*	$(\lambda_1^*, \lambda_2^*)$
0.02	$(0.9236, 0.9618, 0.0382, 0.0000)$	$(-0.6152, 0.1120, 0.0915, 0.1020)$	0.6402	$(0.040, 1.280)$
0.04	$(0.9495, 0.9747, 0.0253, 0.0000)$	$(-0.6170, 0.1106, 0.0950, 0.1018)$	0.6421	$(0.080, 1284)$
0.06	$(0.9754, 0.9877, 0.0123, 0.0000)$	$(-0.6189, 0.1092, 0.0985, 0.1016)$	0.6442	$(0.120, 1.288)$
0.08	$(1.0021, 1.0007, 0.0000, 0.0007)$	$(-0.6218, 0.1084, 0.1021, 0.1014)$	0.6474	$(0.160, 1.295)$
0.10	$(1.0416, 1.0139, 0.0000, 0.0139)$	$(-0.6356, 0.1123, 0.1044, 0.1011)$	0.6616	$(0.200, 1.323)$

Table 3: Results for Algorithm 1 (Experiment 2)

	Algorithm 1			smoothing method		
r	cpu(s)	$\sharp$ite$_{\text{out}}$	$\sharp$QP	cpu(s)	$\sharp$ite$_{\text{out}}$	$\sharp$QP
0.02	0.207	45	44	1.005	43	218
0.04	0.213	43	42	0.941	42	205
0.06	0.213	43	41	0.908	41	199
0.08	0.214	42	40	0.873	40	188
0.10	0.209	41	40	0.875	40	188

Table 4: Comparison of Algorithm 1 and the smoothing method

8 Conclusion

In this paper, we have considered the mathematical program with SOC complementarity constraints. We have proposed an algorithm based on the smoothing and the sequential quadratic programming (SQP) methods, in which we replace the SOC complementarity constraints with smooth equality constraints by means of the natural residual and its smoothing function, and apply the SQP method while decreasing the smoothing parameter gradually. We have shown that the proposed algorithm possesses the global convergence property under the Cartesian P_0 and the nondegeneracy assumptions. We have further confirmed the efficiency of the algorithm through numerical experiments.

References

[1] F. Alizadeh and D. Goldfarb, Second-order cone programming, *Mathematical Programming* 95 (2003) 3–51.

[2] J.F. Bard, *Practical Bilevel Optimization*, Kluwer Academic Publishers, The Netherlands, 1998.

[3] A. Ben-Tal and A. Nemirovski, Robust convex optimization, *Mathematics of Operations Research* 23 (1998) 769–805.

[4] A. Ben-Tal and A. Nemirovski, Robust solutions of uncertain linear programs, *Operations Research Letters* 25 (1999) 1–13.

[5] J.-S. Chen, X. Chen and P. Tseng, Analysis of nonsmooth vector-valued functions associated with second-order cones, *Mathematical Programming* 101 (2004) 95–117.

[6] X. Chen and H.D. Qi, Cartesian P-property and its applications to the semidefinite linear complementarity problem, *Mathematical Programming* 106 (2006) 177–201.

[7] X.D. Chen, D. Sun and J. Sun, Complementarity functions and numerical experiments on some smoothing Newton methods for second-order-cone complementarity problems, *Computational Optimization and Applications* 25 (2003) 39–56.

[8] R.W. Cottle, J.-S. Pang and R.E. Stone, *The Linear Complementarity Problem*, Academic Press, New York, 1992.

[9] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, The Netherlands, 2002.

[10] M. Fukushima, Z.-Q. Luo and J.S. Pang, A globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints, *Computational Optimization and Applications* 10 (1998) 5–34.

[11] M. Fukushima, Z.-Q. Luo and P. Tseng, Smoothing functions for second-order cone complementarity problems, *SIAM Journal on Optimization* 12 (2001) 436–460.

[12] M. Fukushima and P. Tseng, An implementable active-set algorithm for computing a B-stationary point of a mathematical program with linear complementarity constraints, *SIAM Journal on Optimization* 12 (2002) 724–739; erattum, ibid.17(2007) 1253–1257.

[13] M. S. Gowda, R. Sznajder and J. Tao, Some P-properties for linear transformations on Euclidean Jordan algebras, *Linear Algebra and its Applications* 393 (2004) 203–232.

[14] S. Hayashi, N. Yamashita and M. Fukushima, A combined smoothing and regularization method for monotone second-order cone complementarity problems, *SIAM Journal on Optimization* 15 (2005) 593–615.

[15] Z.-Q. Luo, J.-S. Pang and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, 1996.

[16] R. Nishimura, S. Hayashi and M. Fukushima, Robust Nash equilibria in N-person noncooperative games: Uniqueness and reformulation, *Pacific Journal of Optimization* 5 (2009) 237–259.

[17] S.H. Pan and J.S. Chen, A regularization method for the second-order cone complementarity problem with the Cartesian P_0-property, *Nonlinear Analysis: Theory, Methods & Applications*, 70 (2009) 1475–1491.

[18] R.T. Rockafellar and R.J. -B. Wets, *Variational Analysis*, Springer, New York, 1998.

[19] H.S. Scheel and S. Scholtes, Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity, *Mathematics of Operations Research* 25 (2000). 1–22.

[20] T. Yan and M. Fukushima, Smoothing method for mathematical programs with symmetric cone complementarity constraints, *Optimization* 60 (2011) 113–128.

Manuscript received 12 June 2012
revised 17 July 2012
accepted for publication 17 July 2012

Hiroshi Yamamura
Department of Applied Mathematics and Physics, Graduate School of Informatics
Kyoto University, Kyoto 606-8501, Japan
E-mail address: yamamura@amp.i.kyoto-u.ac.jp

Takayuki Okuno
Department of Applied Mathematics and Physics, Graduate School of Informatics
Kyoto University, Kyoto 606-8501, Japan
E-mail address: t_okuno@amp.i.kyoto-u.ac.jp

Shunsuke Hayashi
Department of Applied Mathematics and Physics, Graduate School of Informatics
Kyoto University, Kyoto 606-8501, Japan
E-mail address: shunhaya@amp.i.kyoto-u.ac.jp

Masao Fukushima
Department of Applied Mathematics and Physics, Graduate School of Informatics
Kyoto University, Kyoto 606-8501, Japan
E-mail address: fuku@amp.i.kyoto-u.ac.jp