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 an algorithm based on the smoothing and the sequential quadratic programming (SQP) methods. 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 while decreasing the smoothing parameter. The SQP type method proposed in this paper has an advantage that the exact solution of each subproblem can be calculated easily since it is a convex quadratic programming problem. We further show that the proposed algorithm possesses the global convergence property under the Cartesian P_0 and the nonsingularity assumptions. We finally observe the effectiveness of the algorithm by means of numerical experiments.

Contents

1 Introduction

In this paper, we focus on a mathematical program with second-order cone (SOC) complementarity constraints of the form

$$\begin{aligned} \min_{x,y,w} \ & f(x,y) \\ \text{s.t.} \ & Ax \leq b, \\ & w = Nx + My + q, \\ & \mathcal{K} \ni y \perp w \in \mathcal{K}, \end{aligned} \tag{1.1}$$

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, $\mathcal{K} := \mathcal{K}^{m_1} \times \mathcal{K}^{m_2} \times \cdots \times \mathcal{K}^{m_l} \subset \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \cdots \times \mathbb{R}^{m_l} = \mathbb{R}^m$ with

$$\mathcal{K}^{m_i} = \begin{cases} \{u = (u_1,\ u_2) \in \mathbb{R} \times \mathbb{R}^{m_i - 1} | \|u_2\|_2 \leq u_1\} & (m_i \geq 2), \\ \mathbb{R}_+ = \{u \in \mathbb{R} \mid u \geq 0\} & (m_i = 1). \end{cases}$$

Throughout the paper, we suppose that $m_i \geq 2$ and the following assumption holds for problem (1.1).

Assumption 1. *M is a Cartesian P_0 matrix.*

The definition of the Cartesian P_0 matrix will be provided in section 2.

Mathematical program with equilibrium constraints (MPEC) [10] has been studied extensively so far, since it can be applied to several problems such as the design problems in engineering, the equilibrium problems in economics, and the 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 $m_1 = m_2 = \cdots = m_l = 1$, i.e., $\mathcal{K} = \mathbb{R}^m_+$, problem (1.1) is reduced to the MPCC, for which there have been proposed many algorithms. For example, Fukushima and Tseng [7] proposed an active set algorithm. They proved that any accumulation point of the sequence generated by the algorithm is a B-stationary point under the uniform linear inequality constraint qualification (LICQ) on the ε feasible set. Luo, Pang, and Ralph [10] proposed a piece-wise sequential quadratic programming (SQP) algorithm, and showed that the generated sequence converges to a B-stationary point superlinearly or quadratically under the LICQ and the second order sufficient conditions. Fukushima, Luo, and Pang [5] proposed an SQP-type algorithm, and showed that the sequence generated by the algorithm converges 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] [11]. 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) [2] [4] [6] is another type of problems involving SOC constraints. Fukushima, Luo and Tseng [6] studied smoothing functions for the Fischer-Burmeister function and the natural residual with respect to SOC complementarity conditions. Using this smoothing function, Hayashi, Yamashita and Fukushima [9] proposed a globally and quadratically convergent algorithm based on the smoothing and regularization methods. As an application of the SOCCP, Nishimura, Hayashi and Fukushima [12] studied the SOCCP reformulation of the robust Nash equilibrium problem in N-person non-cooperative game.

As mentioned in the last two paragraphs, there have been many researches on the MPECs with "nonnegative" complementarity and the SOC constrained optimization/complementarity problems. However, there are only a few studies on MPECs with SOC complementarity constraints. For example, Yan and Fukushima [15] proposed the smoothing method for solving such problems. To show the convergence of the algorithm, they assumed that each subproblem is solved exactly. However, since the subproblem is a nonlinear program, it is difficult in general to solve it exactly, and the total computational cost can be very high. To overcome such a disadvantage, we apply an SQP type method to the nonlinear programming problem in which the SOC complementarity condition of problem (1.1) is replaced by a certain vector equation by using the natural residual and its smoothing function. The SQP type method proposed in this paper has an advantage that the exact solution of each subproblem can be calculated easily since it is a convex QP.

This paper is organized as follows. In Section 2, we give some preliminaries. In Section 3, we propose an SQP type algorithm for solving problem (1.1). In Section 4, we show that the proposed algorithm possesses the global convergence property under the Cartesian P_0 and the nonsingularity assumptions. In Section 5, we give some numerical examples. In Section 6, 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 corresponding to the Cartesian structure of $\mathcal{K}$. For vectors $z^1 \in \mathbb{R}^{m_1}, z^2 \in \mathbb{R}^{m_2} \dots, z^l \in \mathbb{R}^{m_l}$, we often denote by $(z^1, z^2, \dots, z^l)$ the column vector $((z^1)^\top, \cdots, (z^l)^\top) \in \mathbb{R}^{m_1+\cdots+m_l}$. For a vector $x \in \mathbb{R}^m$, $\|x\|$ denotes the Euclidean norm defined by $\|x\| := (x^\top x)^{1/2}$. For a matrix $M \in \mathbb{R}^{m\times m}$, $\|M\|$ denotes the operator norm defined by $\|M\| = \max_{\|x\|=1} \|Mx\|$. For vectors $x, y \in \mathbb{R}^m$, $x \succeq_{\mathcal{K}} y$ and $x \succ_{\mathcal{K}} y$ mean $x - y \in \mathcal{K}$ and $x - y \in \mathrm{int}\mathcal{K}^m$, respectively. We denote the nonnegative cone and its interior set by $\mathbb{R}^m_+ := \{x \in \mathbb{R}^m \mid x_i \geq 0 \ (i = 1, 2, \dots, m)\}$ and $\mathbb{R}^m_{++} := \{x \in \mathbb{R}^m \mid x_i > 0 \ (i = 1, 2, \dots, m)\}$, respectively.

2 Preliminaries

2.1 Spectral factorization and natural residual

In constructing the algorithm to solve problem (1.1), it is difficult to handle the SOC complementarity condition $\mathcal{K} \ni y \perp w \in \mathcal{K}$ directly. We therefore transform the SOC complementarity condition into an equivalent equation by means of the natural residual. To this end, we first define the spectral factorization with respect to a single SOC $\mathcal{K}^m$.

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

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

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

By using this spectral factorization, we can describe the Euclidean projection onto $\mathcal{K}^m$ explicitly as follows [9]:

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

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 by using the Euclidean projection.

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

$$\begin{aligned} \Phi(y, w) &:= y - P_{\mathcal{K}}(y - w) \qquad (2.1) \\ &= \begin{pmatrix} \varphi^1(y^1, w^1) \\ \vdots \\ \varphi^l(y^l, w^l) \end{pmatrix} \in \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_l}, \end{aligned}$$

where

$$\varphi^i(y, w) = y^i - P_\mu(y^i - w^i),\ \ i = 1, 2, \ldots, l.$$

For these functions, it is known that

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

Moreover, if the SOC complementarity holds in a strict sense, we have the following proposition.

Proposition 2.1. *Let* $(y, w) \in \mathbb{R}^m \times \mathbb{R}^m$ *be chosen arbitrarily so that* $y^\top w = 0$, $y \in \mathcal{K}$, $w \in \mathcal{K}$ *and* $y + w \in \operatorname{int} \mathcal{K}$. *Then, the function* $\Phi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ *defined in Definition 2 is continuously differentiable at* (y, w) *and satisfies*

$$\nabla_y \Phi(y, w) + \nabla_w \Phi(y, w) = I_m,$$

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

Proof. By an easy calculation, we have

$$\begin{aligned}\nabla_y \Phi(y,w) &= \nabla_y \left(y - P_{\mathcal{K}}(y-w)\right) \\ &= I_m - \nabla_y (y-w) \nabla_z P_{\mathcal{K}}(z) \\ &= I_m - \nabla_z P_{\mathcal{K}}(z),\end{aligned}$$

where $z := y - w$. Similarly, we have $\nabla_w \Phi(y,w) = \nabla_z P_{\mathcal{K}}(z)$. Hence we obtain

$$\nabla_y \Phi(y,w) + \nabla_w \Phi(y,w) = I_m.$$

□

2.2 Smoothing function of natural residual

In the previous section, we introduced the natural residual Φ which equivalently reformulates the SOC complementarity condition $\mathcal{K} \ni y \perp w \in \mathcal{K}$ as $\Phi(y,w) = 0$. However, Φ is not differentiable everywhere, and therefore, we cannot apply an algorithm which requires the differentiation such as Newton method, etc. To overcome such a drawback, we introduce the following smoothing function.

Definition 3. *Let $\Phi : \mathbb{R}^m \to \mathbb{R}^{m'}$ be a nondifferentiable function. Then, the function $\Phi_\mu : \mathbb{R}^m \to \mathbb{R}^{m'}$ parameterized by $\mu > 0$ is called a smoothing function of Φ if it satisfies the following properties.*

1. For any $\mu > 0$, function Φ_μ is differentiable over $\mathbb{R}^m$,
2. For any $x \in \mathbb{R}^m$, it holds that $\lim_{\mu \to 0+} \Phi_\mu(x) = \Phi(x)$.

In order to define a smoothing function of the natural residual function, we first introduce the Chen-Mangasarian (CM) function $\hat{g} : \mathbb{R} \to \mathbb{R}$.

Definition 4. *A differentiable convex function $\hat{g} : \mathbb{R} \to \mathbb{R}_+$ is called a CM function if the following properties hold:*

$$\lim_{\alpha \to -\infty} \hat{g}(\alpha) = 0, \qquad \lim_{\alpha \to \infty} (\hat{g}(\alpha) - \alpha) = 0, \qquad 0 < \hat{g}'(\alpha) < 1 \qquad (\alpha \in \mathbb{R}). \tag{2.2}$$

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

Definition 5. *Let $z \in \mathbb{R}^m$ be an arbitrary vector decomposed as $z = (z^1, z^2, \cdots, z^l) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \cdots \times \mathbb{R}^{m_l} = \mathbb{R}^m$ according to the Cartesian structure of $\mathcal{K}$. For an arbitrary CM function $\hat{g} : \mathbb{R} \to \mathbb{R}$, let $g : \mathbb{R} \to \mathbb{R}$ be defined as*

$$g(z) := \begin{pmatrix} g^1(z^1) \\ \vdots \\ g^l(z^l) \end{pmatrix}, \tag{2.3}$$

$$g^i(z) := \hat{g}(\lambda_{i1}) c^{i1} + \hat{g}(\lambda_{i2}) c^{i2},$$

where λ_{ij} *and* c^{ij} $((i,j) \in \{1,2,\cdots,l\} \times \{1,2\})$ *are the spectral values and the spectral vectors of subvectors* z^i *with respect to* $\mathcal{K}^m$*, respectively. Then, the smoothing function* P_μ *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 6. *Let* $\mu > 0$ *be arbitrary. Let* $g : \mathbb{R}^m \to \mathbb{R}^m, g^i : \mathbb{R}^{m_i} \to \mathbb{R}^{m_i}$ $(i = 1, 2 \ldots, l)$*, and* $P_\mu : \mathbb{R}^m \to \mathbb{R}^m$ *be defined as in Definition* 5*. Then, the smoothing function* Φ_μ *for the natural residual* Φ *is given as*

$$\begin{aligned}\Phi_\mu(y,w) &:= y - P_\mu(y-w) \\ &= y - \mu g\left(\frac{y-w}{\mu}\right) \\ &= \begin{pmatrix} y^1 - \mu g\left(\dfrac{y^1-w^1}{\mu}\right) \\ \vdots \\ y^l - \mu g\left(\dfrac{y^l-w^l}{\mu}\right)\end{pmatrix}.\end{aligned} \tag{2.4}$$

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

Proposition 2.2. [6, 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* (2.4)*, respectively. Let* $\rho := \hat{g}(0)$*. Then, for any* $y,\ w \in \mathbb{R}^n$*,* $(\mu,\ \nu) \in \mathbb{R}\times\mathbb{R}$*, and* $\mu > \nu > 0$*, it follows that*

$$\begin{aligned}\rho(\mu-\nu)e &\succeq_\mathcal{K} \Phi_\nu(y,w) - \Phi_\mu(y,w) \succ_\mathcal{K} 0, \\ \rho\mu e &\succeq_\mathcal{K} \Phi(y,w) - \Phi_\mu(y,w) \succ_\mathcal{K} 0,\end{aligned}$$

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

Proposition 2.3. [9] *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* (2.4)*, respectively. Then, there exists* $\upsilon > 0$ *such that* $\|\Phi_\mu(y,w) - \Phi(y,w)\| \leq \upsilon\mu$ *for any* $(\mu, y, w) \in \mathbb{R}_{++} \times \mathbb{R}^m \times \mathbb{R}^m$*.*

Proposition 2.4. [6, Corollary 5.3, 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* (2.4)*, respectively. Let* $g : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ *be defined by* (2.3)*. Choose* $y, w \in \mathbb{R}^m$ *arbitrarily. Then, the following statements hold.*

(a) *For any* $z \in \mathbb{R}^m$*,* g *is continuous differentiable and* $\nabla g(z) = \mathrm{diag}(\nabla g^1(z^1), \ldots, \nabla g^l(z^l))) \in \mathbb{R}^m \times \mathbb{R}^m$ *is symmetric.*

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

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

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

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

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

Proposition 2.5. *Let $\Phi_\mu : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ be defined by (2.4). Then, for any $\mu > \mu' > 0$ and $(y,w) \in \mathbb{R}^m \times \mathbb{R}^m$, it holds that*

$$\|\Phi_{\mu'}(y,w)\|_1 - \|\Phi_\mu(y,w)\|_1 \leq m\rho(\mu - \mu'),$$

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

Proof. We first assume $\mathcal{K} = \mathcal{K}^m$. From Proposition 2.2, we have

$$\rho(\mu - \mu')e - (\Phi_{\mu'}(y,w) - \Phi_\mu(y,w)) \in \mathcal{K}^m, \tag{2.5}$$

$$\Phi_{\mu'}(y,w) - \Phi_\mu(y,w) \in \mathcal{K}^m, \tag{2.6}$$

where $e = (1, 0, \ldots, 0)^\top \in \mathbb{R}^m$. Moreover, for any $z = (z_1, z_2, \ldots, z_m)^\top \in \mathcal{K}^m$, we have

$$z_1 \geq |z_i| \quad (i = 1, \ldots, m), \tag{2.7}$$

since $z_1 \geq \sqrt{z_2^2 + \cdots + z_m^2}$. Therefore, for each $i = 1, 2, \ldots, m$, for each component of $\Phi_{\mu'}(y,w) - \Phi_\mu(y,w)$, we obtain the following inequalities.

$$\begin{aligned} \rho(\mu - \mu') &\geq \big(\Phi_{\mu'}(y,w) - \Phi_\mu(y,w)\big)_1 \\ &\geq \left|\big(\Phi_{\mu'}(y,w) - \Phi_\mu(y,w)\big)_i\right| \\ &\geq |\Phi_{\mu'}(y,w)_i| - |\Phi_\mu(y,w)_i| \end{aligned}$$

where the first inequality holds from (2.4) and (2.5). Moreover, the second equality holds from (2.6) and (2.7), and the last equality holds from the triangular inequality. Summing up (2.7) for all i, we obtain the desired conclusion. When $\mathcal{K} = \mathcal{K}^{m_1} \times \cdots \times \mathcal{K}^{m_l}$, we can prove it in a similar way. □

2.3 Cartesian P_0 and P matrices

In this subsection, we introduce the Cartesian P_0 and Cartesian P matrices. The concept of the Cartesian $P_0\,(P)$ matrix is a natural extension of the well-known $P_0\,(P)$ matrix that corresponds to the Cartesian structure of SOCs. Although the Cartesian $P_0\,(P)$ matrix can be defined not only for the SOC [3] [8], we restrict ourselves to the case of the SOCs.

Definition 7. *Suppose that the Cartesian structure of $\mathcal{K} \subset \mathbb{R}^m$ is given as $\mathcal{K} := \mathcal{K}^{m_1} \times \mathcal{K}^{m_2} \times \cdots \times \mathcal{K}^{m_l}$. Then, $M \in \mathbb{R}^{m \times m}$ is called*

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

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

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

Notice that the definition of the Cartesian P or P_0 property depends on the Cartesian structure of $\mathcal{K}$ in problem (1.1). In what follows, we assume that the Cartesian structure of $\mathcal{K}$ is always given as $\mathcal{K} = \mathcal{K}^{m_1} \times \mathcal{K}^{m_2} \times \cdots \times \mathcal{K}^{m_l}$. 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 P_0 matrix or Cartesian $P_0\,(P)$ matrix is a Cartesian P_0 matrix [13].

The following proposition implies that the Cartesian $P_0\,(P)$ matrix preserves its property after a nonsingular block-diagonal transformation.

Proposition 2.6. *Let M be a Cartesian P_0 matrix, and $B_1, \ldots, B_l \in \mathbb{R}^{m_i \times m_i}, i = 1, 2, \ldots, l$, be arbitrary nonsingular matrices. Let the matrix $M' \in \mathbb{R}^{m \times m}$ be defined by*

$$M' := B^\top M B, \quad B := \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_l \end{pmatrix}.$$

(a) *If $M \in \mathbb{R}^{m \times m}$ 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. Since (b) can be shown analogously, we only show (a). Let $z = (z^1, \ldots, z^l) \in \mathbb{R}^m = \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_l}$ be an arbitrary nonzero vector. Then it suffices to show that there exists an $i \in \{1, 2, \ldots, l\}$ such that $(z^i)^\top (M'z)^i \geq 0$. Note that

$$\begin{aligned}
(M'z)^i &= \left(\begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_l \end{pmatrix}^\top M \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_l \end{pmatrix} z \right)^i = \left(\begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_l \end{pmatrix}^\top M \begin{pmatrix} B_1 z^1 \\ \vdots \\ B_l z^l \end{pmatrix} \right)^i \\
&= \begin{pmatrix} B_1^\top \sum_{k=1}^{l} M_{1k} B_k z^k \\ \vdots \\ B_l^\top \sum_{k=1}^{l} M_{lk} B_k z^k \end{pmatrix}^i \\
&= \left(B_i^\top \sum_{k=1}^{l} M_{ik} B_k z^k \right).
\end{aligned}$$

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

$$(z^i)^\top (M'z)^i = (z^i)^\top B_i^\top \sum_{k=1}^{l} M_{ik} B_k z^k = \left(B_i z^i\right)^\top \sum_{k=1}^{l} M_{ik} B_k z^k = ((Bz)^i)^\top (MBz)^i.$$

Since M is a Cartesian P_0 matrix and $Bz \neq 0$ from the nonsingularity of B, we have $(z^i)^\top (M'z)^i = ((Bz)^i)^\top (MBz)^i \geq 0$ for some i. ☐

3 Algorithm

In this section, we propose an SQP type algorithm for problem (1.1). The SQP method solves a quadratic programming (QP) problem in each iteration to determine the search direction and the next point. This method is one of the most efficient methods for solving nonlinear programming problems. In order to apply the SQP method to problem (1.1), we need to transform it into the following problem where the SOC complementarity constraint is replaced by the equality with natural residual Φ:

$$\begin{aligned} \min_{x,y,w} \quad & f(x,y) \\ \text{s.t.} \quad & Ax \le b, \\ & w = Nx + My + q, \\ & \Phi(y,w) = 0. \end{aligned} \tag{3.1}$$

In the remainder of the paper, we only consider the problem (3.1) instead of (1.1). One may think that the SQP method can be applied to problem (3.1) in a direct manner. However, it is difficult since $\Phi(y,w)$ is nondifferentiable. We thus consider the following problem where the smoothing function Φ_{μ_k} is incorporated instead of Φ in each iteration k:

$$\begin{aligned} \min_{x,y,w} \quad & f(x,y) \\ \text{s.t.} \quad & Ax \le b, \\ & w = Nx + My + q, \\ & \Phi_{\mu_k}(y,w) = 0. \end{aligned} \tag{3.2}$$

Then we generate the search direction (dx^k, dy^k, dw^k) by solving the QP subproblem, which is composed by approximating the objective and constraint functions of problem (3.2) linearly or quadratically at (x^k, y^k, w^k). The QP subproblem is written concretely as follows:

$$\begin{aligned} \min_{dx,dy,dw} \quad & \nabla f(x^k,y^k)^\top \begin{pmatrix} dx \\ dy \end{pmatrix} + \frac{1}{2} \begin{pmatrix} dx \\ dy \\ dw \end{pmatrix}^\top B^k \begin{pmatrix} dx \\ dy \\ dw \end{pmatrix} \\ \text{s.t.} \quad & Adx \le b - Ax^k, \\ & \begin{pmatrix} N & M & -I_m \\ 0 & \nabla_y \Phi_{\mu_k}(y^k,w^k)^\top & \nabla_y \Phi_{\mu_k}(y^k,w^k)^\top \end{pmatrix} \begin{pmatrix} dx \\ dy \\ dw \end{pmatrix} = - \begin{pmatrix} 0 \\ \Phi_{\mu_k}(y^k,w^k) \end{pmatrix}, \end{aligned} \tag{3.3}$$

where B_k is a positive definite symmetric matrix. In the numerical experiment given in Section 5, we apply the Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula to generate B_k. Note that the Karush-

Kuhn-Tucker (KKT) conditions of the QP (3.3) can be written as

$$
\begin{aligned}
&\begin{pmatrix} \nabla_x f(x^k, y^k) \\ \nabla_y f(x^k, y^k) \\ 0 \end{pmatrix} + B_k \begin{pmatrix} dx \\ dy \\ dw \end{pmatrix} + \begin{pmatrix} N^\top \\ M^\top \\ -I_m \end{pmatrix} u + \begin{pmatrix} 0 \\ \nabla_y \Phi_{\mu_k}(y^k, w^k) \\ \nabla_w \Phi_{\mu_k}(y^k, w^k) \end{pmatrix} v + \begin{pmatrix} A^\top \\ 0 \\ 0 \end{pmatrix} \eta = 0 \\
&\begin{pmatrix} N & M & -I_m \\ 0 & \nabla_y \Phi_{\mu_k}(y^k, w^k) & \nabla_y \Phi_{\mu_k}(y^k, w^k) \end{pmatrix} \begin{pmatrix} dx \\ dy \\ dw \end{pmatrix} = -\begin{pmatrix} 0 \\ \Phi_{\mu_k}(y^k, w^k) \end{pmatrix} \\
&0 \le (b - Ax^k - Adx) \perp \eta \ge 0,
\end{aligned}
\tag{3.4}
$$

where $(\eta,\, u,\, v) \in \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^m$ denotes the Lagrange multipliers. In our algorithm, we also use the Lagrange multipliers for updating the penalty parameters.

For simplicity of notation, we denote

$$z := (x, y, w) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m, \qquad z^k := (x^k, y^k, w^k) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m,$$

hereafter. Also, we define the l_1 penalty function by

$$\theta_{\mu,\alpha}(z) := f(x, y) + \alpha \|\Phi_\mu(y, w)\|_1, \tag{3.5}$$

where $\|\cdot\|_1$ is the l_1 norm and α is the penalty parameter.

Algorithm 1.

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

Step 1: *Solve the QP subproblem* $QP(z^k : \mu_k)$ (3.3) (*equivalently the KKT system* (3.4)) *to obtain the optimum* $dz^k = (dx^k, dy^k, dw^k)$ *and the Lagrange multipliers* (η^k, u^k, v^k).

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

$$\alpha_k := \begin{cases} \alpha_{k-1} & \text{if } \alpha_{k-1} \ge \|v^k\|_\infty + \delta, \\ \max\{\|v^k\|_\infty + \delta,\ \alpha_{k-1} + 2\delta\} & \text{otherwise.} \end{cases} \tag{3.6}$$

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

$$\theta_{\mu_k,\alpha_k}(z^k + \rho^L dz^k) \le \theta_{\mu_k,\alpha_k}(z^k) + \sigma \rho^L \theta'_{\mu_k,\alpha_k}(z^k; dz^k). \tag{3.7}$$

Let $z^{k+1} := z^k + \tau_k dz^k$, $\tau_k := \rho^L$, *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* $k + 1$.

4 Convergence analysis

In the previous section, we proposed the SQP type algorithm for problem (3.1), which is equivalent to (1.1). In this section, we show that the sequence generated by the algorithm globally converges to a B-stationary point of problem (1.1) under the nondegeneracy assumption. The definition of the B-stationarity and the nondegeneracy are given as follows.

Definition 8 (B-stationarity). *Let $z^* := (x^*, y^*, w^*)$ be a feasible point of problem (1.1), and $\mathcal{T}(z^*)$ be the tangent cone at z^* for the feasible set of (1.1). We say that z^* is a B-stationary point of problem (1.1) if $\nabla f(x^*, y^*)^\top (dx, dy) \geq 0$ holds for any $(dx, dy, dw) \in \mathcal{T}(z^*)$.*

Definition 9 (Nondegeneracy). *Suppose that $(y, z) \in \mathbb{R}^m \times \mathbb{R}^m$ satisfies the SOC complementarity condition $\mathcal{K}^m \ni y \perp w \in \mathcal{K}^m$. Moreover, decompose y and w as $y = (y^1, y^2, \dots, y^l)$ and $w = (w^1, w^2, \dots, w^l) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \cdots \times \mathbb{R}^{m_l} = \mathbb{R}^m$ according to the Cartesian structure of $\mathcal{K}$. Then, (y, w) is said to be nondegenerate if, for every $i = 1, 2, \dots, l$, one of the following conditions holds;*

(i) $y^* \in \mathrm{int}\mathcal{K}^m,\ w^* = 0$.

(ii) $y^* = 0,\ w^* \in \mathrm{int}\mathcal{K}^m$.

(iii) $y^* \in \mathrm{bd}\mathcal{K}^m \setminus \{0\},\ w^* \in \mathrm{bd}\mathcal{K}^m \setminus \{0\},\ y^\top w = 0$.

4.1 Feasibility of QP subproblems

We first show the feasibility of QP subproblem (3.3). In general, the QP subproblem generated by the SQP method may not be feasible, even though the original nonlinear programming problem is feasible. However, in case of our algorithm, we can show that QP subproblem (3.3) is feasible if (x^k, y^k, w^k) is feasible to problem (3.2).

The following lemma will be used for showing the feasibility of QP subproblem.

Lemma 4.1. *Let $M \in \mathbb{R}^{m \times m}$ be a Cartesian P_0 matrix. Let $H_i \in \mathbb{R}^{m_i \times m_i}$ $(i = 1, 2, \dots, l)$ be positive definite matrices with $m = \sum_{i=1}^l m_i$, and $H \in \mathbb{R}^{m \times m}$ be a block diagonal matrix with block diagonal elements H_i $(i = 1, \dots, l)$. Then, $H + M$ is nonsingular.*

Proof. For contradiction, assume that $H + M$ is singular. Then, there exists $z = (z^1, z^2, \dots, z^l) \in \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_l} = \mathbb{R}^m$ such that $(H + M)z = 0$ and $z \neq 0$. Since $z \neq 0$, we can choose some i such that $z^i \neq 0$, which yields

$$(z^i)^\top (Mz)^i = -(z^i)^\top (Hz)^i = -(z^i)^\top H_i z^i < 0,$$

where the first equality follows from $(M + H)z = 0$, and the last inequality follows from $H_i \succ 0$ and $z^i \neq 0$. However, this contradicts the assumption that M is a Cartesian P_0 matrix. Hence, $H + M$ is nonsingular. □

By using this lemma, we show the feasibility of QP subproblem (3.3). In the proof of the following proposition, the matrix defined by

$$D_k := \begin{pmatrix} M & -I_m \\ \nabla_y \Phi_{\mu_k}(y^k, w^k) & \nabla_w \Phi_{\mu_k}(y^k, w^k) \end{pmatrix} \tag{4.1}$$

plays an important role.

Proposition 4.1. *Suppose that M is a Cartesian P_0 matrix, and $z^k = (x^k, y^k, w^k)$ satisfies $Ax^k \leq b$. Then, QP subproblem QP* (3.3) *is feasible and has a unique solution.*

Proof. Since the objective function of QP (3.3) is strongly convex, it suffices to show the feasibility. We first show that the matrix D_k defined by (4.1) is nonsingular. Let $\tilde{D}_k$ be the Schur complement of the matrix $\nabla_w \Phi_{\mu_k}(y^k, w^k)$ with respect to D_k, that is,

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

where $y^{i,k}$ and $w^{i,k}$ denote the i-th subvectors of y^k and w^k corresponding to the Cartesian structure of $\mathcal{K}$, respectively, and the invertibility of $\nabla_w \Phi_{\mu_k}(y^k, w^k)$ and each equality follow from Proposition 2.4. Since M is a Cartesian P_0 matrix and $\nabla g((y^{i,k} - w^{i,k})/\mu_k)^{-1} - I_{m_i} \in \mathbb{R}^{m_i \times m_i}$ is positive definite from Proposition 2.4 (c), it follows from Lemma 4.1 that $\tilde{D}_k$ is nonsingular, and hence D_k is nonsingular.

Now, the vectors

$$dx = 0, \qquad \begin{pmatrix} dy \\ dw \end{pmatrix} = -D_k^{-1} \begin{pmatrix} 0 \\ \Phi_{\mu_k}(y^k, w^k) \end{pmatrix}$$

are feasible to (3.3) since $Ax^k \leq b$ from the assumption. This completes the proof. □

4.2 Descent direction

We next show that the search direction dz^k generated in Step 1 of Algorithm 1 is a descent direction of the penalty function θ_{μ_k,α_k} defined by (3.5).

Proposition 4.2. *Let $\{z^k\} = \{(x^k, y^k, w^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ and $\{dz^k\} = \{(dx^k, dy^k, dw^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ be the sequences generated by Algorithm* 1*. Then, we have*

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

(b) $\theta'_{\mu_k,\alpha_k}(z^k; dz^k) \leq -(dz^k)^\top B_k dz^k$

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

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

$$J_+^k = \{j \mid \Phi_{\mu_k}(y^k, w^k)_j > 0\},$$
$$J_0^k = \{j \mid \Phi_{\mu_k}(y^k, w^k)_j = 0\},$$
$$J_-^k = \{j \mid \Phi_{\mu_k}(y^k, w^k)_j < 0\},$$

respectively, where $\Phi_{\mu_k}(y^k, w^k)_j \in \mathbb{R}$ denotes the j-th component of $\Phi_{\mu_k}(y^k, w^k) \in \mathbb{R}^m$. (Notice that it is not the j-th subvector.) Then, we have

$$\begin{aligned}\theta'_{\mu_k,\alpha_k}(z^k;\, dz^k) \;=\; & \nabla f(x^k, y^k)^\top \begin{pmatrix} dx^k \\ dy^k \end{pmatrix} + \alpha_k \sum_{j \in J_+^k} \left[\nabla \Phi_{\mu_k}(y^k, w^k)\right]_j^\top \begin{pmatrix} dy^k \\ dw^k \end{pmatrix} \\ & + \alpha_k \sum_{j \in J_0^k} \left| \left[\nabla \Phi_{\mu_k}(y^k, w^k)\right]_j^\top \begin{pmatrix} dy^k \\ dw^k \end{pmatrix} \right| - \alpha_k \sum_{j \in J_-^k} \left[\nabla \Phi_{\mu_k}(y^k, w^k)\right]_j^\top \begin{pmatrix} dy^k \\ dw^k \end{pmatrix}, \end{aligned} \tag{4.2}$$

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

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

from the constraint of QP subproblem (3.3), we have

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

We next show (b). Taking the inner product of $dz^k = (dx^k, dy^k, dw^k)$ and the first equality of the KKT conditions (3.4)for the subproblem (3.3), we obtain

$$\begin{aligned} & \nabla f(x^k, y^k)^\top \begin{pmatrix} dx^k \\ dy^k \end{pmatrix} + (dz^k)^\top B^k dz^k + (u^k)^\top (N dx^k + M dy^k - dw^k) \\ & + (v^k)^\top \nabla \Phi_{\mu_k}(y^k, w^k) \begin{pmatrix} dy^k \\ dw^k \end{pmatrix} + (\eta^k)^\top A dx^k = 0. \end{aligned} \tag{4.3}$$

Moreover, since from the constraints of the subproblem (3.3) and the KKT conditions (3.4), we have

$$N dx^k + M dy^k - dw^k = 0, \tag{4.4}$$

$$\nabla \Phi_{\mu_k}(y^k, w^k) \begin{pmatrix} dy^k \\ dw^k \end{pmatrix} = -\Phi_{\mu_k}(y^k, w^k), \tag{4.5}$$

and

$$0 = (\eta^k)^\top (b - Ax^k - A dx^k) = -(\eta^k)^\top A dx^k + (\eta^k)^\top (b - Ax^k) \geq -(\eta^k)^\top A dx^k, \tag{4.6}$$

where the inequality is due to $\eta \geq 0$ and $b - Ax^k \geq 0$ from (3.4). Substituting (4.4)–(4.6) into (4.3), we have

$$\nabla f(x^k, y^k)^\top \begin{pmatrix} dx^k \\ dy^k \end{pmatrix} + (dz^k)^\top B^k dz^k - (v^k)^\top \Phi_{\mu_k}(y^k, w^k) \leq 0. \tag{4.7}$$

Furthermore, substituting (4.7) into the equation (4.2), we obtain

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

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

This proposition guarantees that, for an arbitrary $\sigma \in (0,1)$, there exists a $\overline{\tau} > 0$ such that

$$\theta_{\mu_k,\alpha_k}(z_k + \tau dz^k) \leq \theta_{\mu_k,\alpha_k}(z_k) + \sigma\tau\theta'_{\mu_k,\alpha_k}(z_k;dz_k)$$

for all $\tau \in [0,\overline{\tau})$. This means that the line search in Step 2 is well-defined in the sense that there exists a finite L satisfying Armijo condition (3.7).

4.3 Some technical results for global convergence

In this section, we give some technical lemmas that will be used for showing the convergence property of the algorithm.

Lemma 4.2. (see [14]). *Assume that the sequences $\{v_k\}, \{\gamma_k\}, \{\beta_k\} \subset \mathbb{R}$ satisfy*

$$(1)\; v_{k+1} \leq (1+\gamma_k)v_k + \beta_k, \quad (2)\; \sum_{i=1}^{\infty} \beta_k < \infty, \quad (3)\; \sum_{i=1}^{\infty} \gamma_k < \infty, \quad (4)\; \{v_k\} \text{ is lower bounded.}$$

Then, $\{v^k\}$ is convergent.

Lemma 4.3. *Let $\{z^k\} := \{(x^k,y^k,w^k) \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m\}$ be an arbitrary sequence satisfying $Ax^k \leq b$, and $dz^k := (dx^k,dy^k,dw^k) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ be the unique optimum of QP subproblem (3.3). Let $\{(\mu_k,\tau_k)\} \subset \mathbb{R}_{++} \times \mathbb{R}_{++}$ be an arbitrary sequence converging to $(0,0)$. Let $\alpha > 0$ be an arbitrary fixed scalar. If $\{z^k\}$ and dz^k are bounded and $\lim_{k\to\infty} dz^k \neq 0$, then we have*

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

Proof. Let $\bar{z} := (\bar{x},\bar{y},\bar{w})$ and $\bar{dz} = (\bar{dx},\bar{dy},\bar{dw})$ be arbitrary accumulation points of $\{z^k\}$ and $\{dz^k\}$, respectively. Then taking subsequences if necessary, we have $z^k \to \bar{z}$ and $dz^k \to \bar{dz}$ as $k \to \infty$. Moreover, we can easily have $\theta'_{\mu_k,\alpha}(z^k;dz^k) = \nabla_x f(x^k,y^k)^\top dx^k + \nabla_y f(x^k,y^k)^\top dy^k - \alpha\|\Phi_{\mu_k}(y^k,w^k)\|_1$ in a way similar to the proof of Proposition 4.2(a). Thus, we only have to show

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

From the mean-value theorem, there exists $v^k \in (0,1)$ such that

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

Therefore, it suffices to show that

$$\limsup_{k\to\infty} \frac{|\Phi_{\mu_k}(y^k+\tau_k dy^k, w^k+\tau_k dw^k)_j| - |\Phi_{\mu_k}(y^k,w^k)_j|}{\tau_k} \le -|\Phi(\overline{y},\overline{w})_j|. \tag{4.8}$$

We first consider the case where $\Phi(\overline{y},\overline{w})_j = 0$. By Definition 5 and Proposition 2.4, we have $\nabla_y \Phi_{\mu_k}(y^k,w^k) = I_m - P_{\mu_k}(y^k-w^k)$ and $\nabla_w \Phi_{\mu_k}(y^k,w^k) = P_{\mu_k}(y^k-w^k)$, which together with the constraints of QP subproblem (3.3) yeilds

$$\begin{aligned} -\Phi_{\mu_k}(y^k,w^k) &= \nabla_y \Phi_{\mu_k}(y^k,w^k)dy^k + \nabla_w \Phi_{\mu_k}(y^k,w^k)dw^k \\ &= (I - \nabla P_{\mu_k}(y^k-w^k))dy^k + \nabla P_{\mu_k}(y^k-w^k)dw^k \\ &= dy^k - \nabla P_{\mu_k}(y^k-w^k)(dy^k-dw^k) \\ &= dy^k - \nabla P_{\mu_k}^\top(y^k-w^k)(dy^k-dw^k), \end{aligned} \tag{4.9}$$

where the last equality is due to the symmetry of ∇P_{μ_k}. Hence, we have

$$\begin{aligned} &\Phi_{\mu_k}(y^k+\tau_k dy^k, w^k+\tau_k dw^k)_j - \Phi_{\mu_k}(y^k,w^k)_j \\ &= \tau_k dy_j^k - \Big(P_{\mu_k}(y^k+\tau_k dy^k-(w^k+\tau_k dw^k))_j - P_{\mu_k}(y^k-w^k)_j\Big) \\ &= \tau_k \left(\left(-\Phi_{\mu_k}(y^k,w^k)_j + \nabla P_{\mu_k}(y^k-w^k)_j^\top(dy^k-dw^k) \right) \right. \\ &\qquad \left. - P_{\mu_k}(y^k-w^k+\tau_k(dy^k-dw^k))_j + P_{\mu_k}(y^k-w^k)_j \right) \\ &= -\tau_k \Phi_{\mu_k}(y^k,w^k)_j + \tau_k \delta_k, \end{aligned} \tag{4.10}$$

where the first equality is due to Definition 6, the second equality follows from (4.9) and the mean-value theorem with $\zeta_j^k \in (0,1)$, and δ_k means the term such that $\delta_k \to 0$ as $k \to 0$. Moreover, by the triangle inequality, we have

$$\|\Phi_{\mu_k}(y^k+\tau_k dy^k, w^k+\tau_k dw_k)_j\| - \|\Phi_{\mu_k}(y^k,w^k)_j\| \le \|\Phi_{\mu_k}(y^k+\tau_k dy^k, w^k+\tau_k dw_k)_j - \Phi_{\mu_k}(y^k,w^k)_j\|.$$

Dividing the both sides by τ_k and letting $k\to\infty$, we obtain (4.8) since $\Phi_{\mu_k}(y^k,w^k)_j$ tends to 0.

If $\Phi(\overline{y},\overline{w})_j > 0$, then we have, for all k sufficient large,

$$\begin{aligned} &|\Phi_{\mu_k}(y^k+\tau_k dy^k, w^k+\tau_k dw^k)_j| - |\Phi_{\mu_k}(y^k,w^k)_j| \\ &= \Phi_{\mu_k}(y^k+\tau_k dy^k, w^k+\tau_k dw^k)_j - \Phi_{\mu_k}(y^k,w^k)_j. \end{aligned}$$

Therefore we obtain the desired conclusion in a similar way by using (4.10).

If $\Phi(\overline{y},\overline{w})_j < 0$, then we have, for all k sufficiently large,

$$\begin{aligned} &|\Phi_{\mu_k}(y^k+\tau_k dy^k, w^k+\tau_k dw^k)_j| - |\Phi_{\mu_k}(y^k,w^k)_j| \\ &= -\Phi_{\mu_k}(y^k+\tau_k dy^k, w^k+\tau_k dw^k)_j + \Phi_{\mu_k}(y^k,w^k)_j. \end{aligned}$$

Therefore we obtain the desired conclusion in a similar way by using (4.10). □

Now, we make the following assumption on the sequence generated by Algorithm 1.

Assumption 2. *For the sequence generated by Algorithm* 1,

(a) $\{z^k\}$ *is bounded*;

(b) *There exists a positive scalar* $\gamma_1 > 0$ *such that* $\gamma_1 < \lambda_{\min}(B_k)$ *for all* k*, where* $\lambda_{\min}(\cdot)$ *denotes the smallest eigenvalue*;

(c) *There exists a positive scalar* $c > 0$ *such that* $\|D_k^{-1}\| \leq c$ *for any* k*, where* D_k *is the matrix defined by* (4.1)*.*

As shown in the proof of Proposition 4.1, D_k is nonsingular for any k. Therefore, Assumption 2 (c) means that any accumulation point of D_k is nonsingular. In the end of this subsection, we will provide a sufficient condition under which Assumption 2 (c) is satisfied.

The following lemma plays a crucial role in showing the convergence theorem in the next subsection.

Lemma 4.4. *Let* $\{z^k\}$ *be the sequence generated by Algorithm* 1*. Suppose that Assumption* 2 *holds. Then, we have the following statements.*

(i) $\{dz^k\}$ *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* $\{\theta_{\mu_k,\alpha_k}(z^k)\}$ *and* $\{\theta_{\mu_k,\alpha_k}(z^{k+1})\}$ *converge to the same limit.*

Proof. We first prove (a). Let

$$\begin{pmatrix} d\tilde{y}^k \\ d\tilde{w}^k \end{pmatrix} := \begin{pmatrix} M & -I_m \\ \nabla_y \Phi_{\mu_k}(y^k, w^k) & \nabla_w \Phi_{\mu_k}(y^k, w^k) \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \Phi_{\mu_k}(y^k, w^k) \end{pmatrix}$$

and $d\tilde{z}^k := (0,\ d\tilde{y}^k,\ d\tilde{w}^k)$. Then $d\tilde{z}^k$ is feasible to QP (3.3), and $d\tilde{z}^k$ is bounded from Assumptions 2 (a), (c) and Proposition 2.3. Since the objective function of QP subproblem (3.3) is rewritten as

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

with

$$g^k = \begin{pmatrix} \nabla f(x^k, y^k) \\ 0 \end{pmatrix},$$

the optimum dz^k of QP (3.3) satisfies

$$\begin{aligned} \frac{1}{2}(d\tilde{z}^k - B_k^{-1}g^k)^\top B_k(d\tilde{z}^k - B_k^{-1}g^k) &\geq \frac{1}{2}(dz^k - B_k^{-1}g^k)^\top B_k(dz^k - B_k^{-1}g^k) \\ &\geq \gamma_1 \|dz^k - B_k^{-1}g^k\|^2 \\ &\geq \gamma_1 \|dz^k\|^2 - \gamma_1 \|B_k^{-1}\|^2 \|g^k\|^2 \\ &\geq \gamma_1 \|dz^k\| - \gamma_1^{-1} \|g^k\|^2, \end{aligned}$$

where the first inequality is due to the optimality of dz^k, the second and the last inequalities follow from Assumption 2 (a), and the third inequality holds from the triangle inequality. Since $\{g^k\}$ is bounded from Assumption 2 (a), $\{dz^k\}$ is also bounded. From the first equality of the KKT conditions (3.4), we have

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

where D_k is defined by (4.1). Hence, by Assumption 2 (c), $\{(u^k, v^k)\}$ is also bounded.

We next prove (b). From the update rule (3.6), $\{\alpha_k\}$ is nondecreasing. Moreover, if

$$\|v^k\|_\infty > \alpha_{k-1} - \delta, \tag{4.11}$$

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 increase at least 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 (4.11). However this contradicts (a). Thus we have (b).

Finally we show (c). Since we have (b), there exist $\overline{\alpha}$ and k_0 such that $\overline{\alpha} = \alpha_k$ for all $k \geq k_0$. In what follows, we only consider the case $k \geq k_0$. From Proposition 2.5 together with (3.6), we have

$$\theta_{\mu_{k+1},\overline{\alpha}}(z^{k+1}) \leq \theta_{\mu_k,\overline{\alpha}}(z^{k+1}) + \overline{\alpha} m\rho(1-\beta)\mu_k \tag{4.12}$$

$$\leq \theta_{\mu_k,\overline{\alpha}}(z^k) + \overline{\alpha} m\rho(1-\beta)\mu_k, \tag{4.13}$$

where the last inequality follows from the Armijo condition (3.7) and Proposition 4.2 (b). Now, notice that $\sum_{k=1}^{\infty} \overline{\alpha} m\rho(1-\beta)\mu_k = \overline{\alpha} m\rho\mu_0 < \mu_0$. Therefore, letting $v^k := \theta_{\mu_k,\overline{\alpha}}(z^k)$, $\gamma_k \equiv 0$, and $\beta_k := \overline{\alpha} m\rho(1-\beta)\mu_k$ in Lemma 4.2, we have from (4.13) that $\{\theta_{\mu_k,\overline{\alpha}}\}$ is convergent. Moreover, from (4.12) and (4.13), $\theta_{\mu_k,\overline{\alpha}}(z^{k+1})$ and $\theta_{\mu_k,\overline{\alpha}}(z^k)$ must converge to the same limit, since $\mu_k - \mu_{k+1} = (1-\beta\mu_k)$ converges to 0. □

Before closing this subsection, we provide a sufficient condition under which Assumption 2 (c) holds. In the following proposition, we consider the orthogonal matrix $H \in \mathbb{R}^{m\times m}$ such that $H\nabla_y\Phi_y\Phi(\bar{y}, \bar{w})H^\top$ is diagonal. Here, we note that H already has a block-digonal structure with respect to $\mathcal{K} = \mathcal{K}^{m_1} \times \cdots \times \mathcal{K}^{m_l}$, since $\nabla_y\Phi$ is block-diagonal by Proposition 2.4 (a) and (b). We also use the following notation: For any given matrix $M \in \mathbb{R}^{m\times m}$ and index sets $\bar{J} \subseteq \{1, 2, \ldots, l\}$ and $\bar{K} \subseteq \{1, 2, \ldots, l\}$, we denote by $M_{\bar{J}\bar{K}}$ the submatrix consisting of M_{ij} with $(i, j) \in \bar{J} \times \bar{K}$.

Proposition 4.3. *Let M be a Cartesian P_0 matrix, and $(\bar{y}, \bar{w}) \in \mathbb{R}^m \times \mathbb{R}^m$ be arbitrary vectors at which Φ is differentiable. Let $H \in \mathbb{R}^m \times \mathbb{R}^m$ be the orthogonal matrix such that $D_{\bar{y}} := H\nabla\Phi(\bar{y}, \bar{w})H^\top$ is diagonal. If $(HMH^\top)_{\bar{J}\bar{J}} \in \mathbb{R}^{|\bar{J}|\times|\bar{J}|}$ is nonsingular with $\bar{J} := \{i \mid (D_{\bar{y}})_{ii} = 0\}$, then the matrix*

$$D_\infty = \begin{pmatrix} M & -I_m \\ \nabla_y\Phi(\overline{y}, \overline{w}) & \nabla_w\Phi(\overline{y}, \overline{w}) \end{pmatrix}$$

is nonsingular.

Proof. It suffices to show that

$$
\begin{aligned}
D' &= \begin{pmatrix} H & 0 \\ 0 & H \end{pmatrix} D_\infty \begin{pmatrix} H^\top & 0 \\ 0 & H^\top \end{pmatrix} \\
&= \begin{pmatrix} H & 0 \\ 0 & H \end{pmatrix} \begin{pmatrix} M & -I_m \\ \nabla_y \Phi(\overline{y}, \overline{w}) & \nabla_w \Phi(\overline{y}, \overline{w}) \end{pmatrix} \begin{pmatrix} H^\top & 0 \\ 0 & H^\top \end{pmatrix} \\
&= \begin{pmatrix} HMH^\top & -I_m \\ H\nabla_y \Phi(\overline{y}, \overline{w}) H^\top & H\nabla_w \Phi(\overline{y}, \overline{w}) H^\top \end{pmatrix} \\
&= \begin{pmatrix} HMH^\top & -I_m \\ D_{\overline{y}} & D_{\overline{w}} \end{pmatrix} \\
&= \begin{pmatrix} M' & -I_m \\ D_{\overline{y}} & D_{\overline{w}} \end{pmatrix}
\end{aligned}
$$

is nonsingular, where $M' := HMH^\top$, $D_{\overline{y}} := H\nabla_y \Phi(\overline{y}, \overline{w})H^\top$ and $D_{\overline{w}} := H\nabla_w \Phi(\overline{y}, \overline{w})H^\top$. Let $\bar{K} := \{1, 2, , \dots, l\} \setminus \bar{J}$. Then, $(D_{\bar{w}})_{\bar{J}\bar{J}} = I_{\bar{J}}$ and $(D_{\bar{y}})_{\bar{K}\bar{K}}$ is a positive diagonal matrix since we have $D_{\bar{y}} + D_{\bar{w}} = I_m$, $O \preceq D_{\bar{y}} \preceq I_m$, and $O \preceq D_{\bar{w}} \preceq I_m$ from Propositions 2.1, 2.3 and 2.4(c). Moreover, M' is a Cartesian P_0 matrix from Proposition 2.6 and the block-diagonal structure of H corresponding to $\mathcal{K} = \mathcal{K}^{m_1} \times \cdots \times \mathcal{K}^{m_l}$. Now, notice that M' is also a P_0 matrix, since any Cartesian P_0 matrix is a P_0 matrix. Since every principal minor of P_0 matrix is nonnegative and $M'_{\bar{J}\bar{J}}$ is nonsingular, we must have $\det M'_{\bar{J}\bar{J}} > 0$. Let $M'/M'_{\bar{J}\bar{J}}$ denote the Schur complement of $M'_{\bar{J}\bar{J}}$ in M, i.e.,

$$
M'/M'_{\bar{J}\bar{J}} = M'_{\bar{K}\bar{K}} - M'_{\bar{K}\bar{J}} \left(M'_{\bar{J}\bar{J}}\right)^{-1} M'_{\bar{J}\bar{K}}
$$

Then, we can easily show that $M'/M'_{\bar{J}\bar{J}}$ is a P_0 matrix.[1] Hence,

$$
(D_{\bar{y}})_{\bar{K}\bar{K}} + (M'/M'_{\bar{J}\bar{J}})(D_{\bar{w}})_{\bar{K}\bar{K}}
$$

is a P matrix.[2]

By a direct calculation, we have

$$
D_{\overline{y}} + M' D_{\overline{w}} = \begin{pmatrix} (D_{\overline{y}})_{\bar{K}\bar{K}} + M'_{\bar{K}\bar{K}}(D_{\overline{w}})_{\bar{K}\bar{K}} & M'_{\bar{K}\bar{J}}(D_{\overline{w}})_{\bar{J}\bar{J}} \\ M'_{\bar{J}\bar{J}}(D_{\overline{w}})_{\bar{K}\bar{K}} & M'_{\bar{J}\bar{J}}(D_{\overline{w}})_{\bar{J}\bar{J}} \end{pmatrix}.
$$

The Schur complement of $(M'_{\bar{J}\bar{J}})(D_{\overline{w}})_{\bar{J}\bar{J}}$ in $D_{\overline{y}} + M' D_{\overline{w}}$ is

$$
\begin{aligned}
&(D_{\overline{y}})_{\bar{K}\bar{K}} + M'_{\bar{K}\bar{K}}(D_{\overline{w}})_{\bar{K}\bar{K}} - M'_{\bar{K}\bar{J}}(D_{\overline{w}})_{\bar{J}\bar{J}}(M'_{\bar{J}\bar{J}}(D_{\overline{w}})_{\bar{J}\bar{J}})^{-1} M'_{\bar{J}\bar{K}}(D_{\overline{w}})_{\bar{K}\bar{K}} \\
&= (D_{\overline{y}})_{\bar{K}\bar{K}} + (M'_{\bar{K}\bar{K}} - M'_{\bar{K}\bar{J}}(M'_{\bar{J}\bar{J}})^{-1} M'_{\bar{J}\bar{K}})(D_{\overline{w}})_{\bar{K}\bar{K}} \\
&= (D_{\overline{y}})_{\bar{K}\bar{K}} + (M'/M'_{\bar{J}\bar{J}})(D_{\overline{w}})_{\bar{K}\bar{K}}.
\end{aligned}
$$

[1] This can be shown as follows. Choose $\bar{K}' \subseteq \bar{K}$ arbitrarily. Then, by an easy calculation, we have $(M'/M_{\bar{J}\bar{J}})_{\bar{K}'\bar{K}'} = M_{\bar{K}'\bar{K}'} - M_{\bar{K}'\bar{J}} M_{\bar{J}\bar{J}}^{-1} M_{\bar{J}\bar{K}'} = M'_{(\bar{J}\cup\bar{K}')(\bar{J}\cup\bar{K}')}/M'_{\bar{J}\bar{J}}$. Hence, $\det (M'/M'_{\bar{J}}\bar{J})_{\bar{K}\bar{K}} = \det M'_{(\bar{J}\cup\bar{K}')(\bar{J}\cup\bar{K}')}/\det M'_{\bar{J}\bar{J}} \geq 0$ from $\det M'_{(\bar{J}\cup\bar{K}')(\bar{J}\cup\bar{K}')} \geq 0$ and $\det M'_{\bar{J}\bar{J}} > 0$. Since $\bar{K}' \subseteq \bar{K}$ was arbitrary chosen, $M'/M'_{\bar{J}\bar{J}}$ is a P_0 matrix.

[2] Since $(M'/M'_{\bar{J}\bar{J}})$ is a P_0 matrix and $(D_{\bar{w}})_{\bar{K}\bar{K}}$ is a nonnegative diagonal matix, $(M'/M'_{\bar{J}\bar{J}})$ is a P_0 matrix. Moreover, $(D_{\bar{y}})_{\bar{K}\bar{K}}$ a positive diagonal matrix. Hence, we have P property.

where the second equality follows from $(D_{\bar{w}})_{\bar{J}\bar{J}} = I_{|J|}$. This matrix is a P matrix as shown above, and hence is nonsingular. Since $M'_{\bar{J}\bar{J}}(D_{\bar{w}})_{\bar{J}\bar{J}}$ is nonsingular, $D_{\bar{y}} + M'D_{\bar{w}}$ is nonsingular. Therefore, by an easy calculation, we can show that D' is nonsingular.[3] □

By this proposition, we readily obtain the following two corollaries.

Corollary 4.1. *Let* $\{z^k\} = \{(x^k, y^k, w^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ *be the sequence generated by Algorithm* 1*. Suppose that Assumption 1 and* 2 (a) *hold. Moreover, suppose that any accumulation point* $\bar{z} = (\bar{x}, \bar{y}, \bar{w})$ *satisfies the following statements:*

(i) $(\bar{y}, \bar{w})$ *satisfies the nondegeneracy condition,*

(ii) $(HMH^\top)_{\bar{J}\bar{J}} \in \mathbb{R}^{|\bar{J}|\times|\bar{J}|}$ *is nonsingularity with* $\bar{J} := \{i | (D_{\bar{y}})_{ii} = 0\}$ *where* $H \in \mathbb{R}^{m\times m}$ *is the orthogonal matrix such that* $D_{\bar{y}} := H\nabla\Phi(\bar{y}, \bar{w})H^\top$ *is diagonal.*

Then, Assumption 2 (c) *holds.*

Corollary 4.2. *Let* $\{z^k\} = \{(x^k, y^k, w^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ *be the sequence generated by Algorithm* 1*. Suppose that M is a Cartesian P matrix and Assumption* 2 (a) *holds. Moreover, suppose that any accumulation point* $\bar{z} = (\bar{x}, \bar{y}, \bar{w})$ *satisfies the following statement:*

(i) $(\bar{y}, \bar{w})$ *satisfies the nondegeneracy condition.*

Then, Assumption 2 (c) *holds.*

4.4 Global convergence

Finally, we show that the algorithm converges globally. It is known that the natural residual Φ is differentiable at (y, w) if it is nondegenerate. Moreover, if (x^*, y^*, w^*) satisfies the KKT conditions of problem (3.1), then it is a B-stationary point. Hence, we will show that the sequence generated by the algorithm converges to a KKT point (x^*, y^*, w^*) satisfying

$$\begin{pmatrix} \nabla_x f(x^*, y^*) \\ \nabla_y f(x^*, y^*) \\ 0 \end{pmatrix} + \begin{pmatrix} N^T \\ M^T \\ -I \end{pmatrix} u^* + \begin{pmatrix} 0 \\ \nabla_y \Phi(y^*, w^*) \\ \nabla_w \Phi(y^*, w^*) \end{pmatrix} v^* + \begin{pmatrix} A^T \\ 0 \\ 0 \end{pmatrix} \eta^* = 0,$$
$$\eta^* \geq 0 \quad (\eta^*)^\top (Ax^* - b) = 0,$$

where u^*, v^* and η^* are Lagrange multipliers.

Theorem 4.1. *Suppose that* $\{z^k\}$ *satisfies Assumptions* 1 *and* 2*. Let* $\bar{z} = (\bar{x}, \bar{y}, \bar{w})$ *be an arbitrary accumulation point of* $\{z^k\}$*. If* $(\overline{y}, \overline{w})$ *satisfies the nondegeneracy condition, then* $\overline{z}$ *is the B-stationary point of problem* (1.1)*.*

[3] Assuming $D\binom{\xi}{\eta} = 0$, we have $M'\xi - \eta = 0$ and $D_{\bar{y}}\xi + D_{\bar{w}}\eta = 0$, which are equivalent to $\eta = M'\xi$ and $(D_{\bar{y}} + D_{\bar{w}}M')\xi = 0$. Hence, if $D_{\bar{y}} + D_{\bar{w}}M'$ is nonsingular, then D' is also nonsingular.

Proof. We first show that

$$\lim_{k\to\infty} \|dz^k\| = 0. \tag{4.14}$$

From Proposition 4.2 (b) and Assumption 2 (b), we have

$$\theta'_{\mu_k,\alpha_k}(z^k; dz^k) \le -(dz^k)^\top B_k dz^k \le -\gamma_1 \|dz^k\|^2, \tag{4.15}$$

which together with Armijo condition (3.7) yields

$$\begin{aligned}\theta_{\mu_k,\alpha_k}(z^{k+1}) &\le \theta_{\mu_k,\alpha_k}(z^k) + \sigma\tau_k\theta'_{\mu_k,\alpha_k}(z^k; dz^k)\\ &\le \theta_{\mu_k,\alpha_k}(z^k) - \gamma_1\sigma\tau_k\|dz^k\|^2.\end{aligned}$$

Hence, from $\gamma_1\sigma > 0$ and Lemma 4.4 (iii), we obtain

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

Now, assume for contradiction that (4.14) does not hold. Then, there exists an index set $K \subset \{0, 1, \dots\}$ such that

$$\lim_{\substack{k\to\infty\\ k\in K}} \|dz^k\| > 0, \tag{4.16}$$

and hence

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

Let l_k be the minimum nonnegative interger L satisfying (3.7), i.e., $\rho^{l_k} = \tau_k$. Then, from the definition of l_k, we have

$$\theta_{\mu_k,\alpha_k}(z^k + \rho^{l_k-1}dz^k) > \theta_{\mu_k,\alpha_k}(z^k) + \sigma\rho^{l_k-1}\theta'_{\mu_k,\alpha_k}(z^k; dz^k),$$

that is,

$$\xi_k := \frac{\theta_{\mu_k,\alpha_k}(z^k + \rho^{l_k-1}dz^k) - \theta_{\mu_k,\alpha_k}(z^k)}{\rho^{l_k-1}} - \theta'_{\mu_k,\alpha_k}(z^k; dz^k) > -(1-\sigma)\theta'_{\mu_k,\alpha_k}(z^k; dz^k). \tag{4.17}$$

By Lemma 4.3 together with $\lim_{k\to\infty}\rho^{l_k-1}$=0, there exists $K' \subseteq K$ such that $\lim_{k\to\infty,\ k\in K'} \xi_k \le 0$. Moreover, we have from (4.15)

$$-(1-\delta)\theta'_{\mu_k,\alpha}(z^k; dz^k) \ge (1-\sigma)\gamma_1\|dz^k\|^2, \tag{4.18}$$

which implies $\lim_{k\to\infty,\ k\in K'} \|dz^k\| = 0$ from (4.17) and (4.18). However, this contradicts (4.16) since $K' \subseteq K$. Hence we have (4.14).

Next, we show that $\overline{z}$ is a B-stationry point of (1.1). Let $\{(\eta^k, u^k, v^k)\}$ be the sequence of multipliers. Then, $\{(u^k, v^k)\}$ is bounded from Lemma 4.4 (i). Hence, there exist vectors $\bar{u}$, $\bar{w}$ and an index set $K \subseteq \{0, 1, \dots\}$ such that $\lim_{k\to\infty}(z^k, u^k, v^k) = (\bar{z}, \bar{u}, \bar{v})$. Moreover, by an easy disscussion, there must exist $\bar{\eta}$ such that

$$\begin{aligned}&\begin{pmatrix}\nabla_x f(\overline{x},\overline{y})\\ \nabla_y f(\overline{x},\overline{y})\\ 0\end{pmatrix} + \begin{pmatrix}N^T\\ M^T\\ -I_m\end{pmatrix}\bar{u} + \begin{pmatrix}0\\ \nabla_y\Phi(\overline{y},\overline{w})\\ \nabla_w\Phi(\overline{y},\overline{w})\end{pmatrix}\bar{v} + \begin{pmatrix}A^T\\ 0\\ 0\end{pmatrix}\bar{\eta} = 0,\\ &\Phi(\overline{y},\overline{w}) = 0\\ &0 \le (b - A\bar{x}) \perp \bar{\eta} \ge 0.\end{aligned}$$

Therefore, $\bar{z}$ satisfies the KKT condition of problem (3.1), and hence $\bar{z}$ is a B-stationary point of problem (1.1). □

5 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 2010a 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

$$\delta := 1, \quad \alpha_{-1} := 10, \quad \beta := 0.8, \quad \mu_0 := 10m, \quad \sigma := 10^{-3}, \quad \rho := 0.9,$$

where m denotes a dimension of $\mathcal{K}$, i.e., Cartesian products of second-order cones. As an initial value B_0 for $\{B_k\}$, we choose the identity matrix, and update B_k by the BFGS formula:

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

where $s^k = z^{k+1} - z^k$ and $u^k = \nabla f(z^{k+1}) - \nabla f(z^k)$. The choice of z^0 will be mentioned later. 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:

$$\left\|\Phi(y^k, w^k)\right\|_\infty + \|dx^k\|_\infty \le 10^{-5}. \tag{5.1}$$

Note that $\Phi(y^k, w^k) = 0$ implies the feasiblity of (x^k, y^k, w^k) for problem (1.1) since the remaining constraints $Ax^k \le b$ and $w^k = Nx^k + My^k + q$ always hold, and $\|dx^k\|_\infty = 0$ yields that the KKT conditions for (3.2) are satisfied by (x^k, y^k, w^k). Thus, $\left\|\Phi(y^k, w^k)\right\|_\infty + \|dx^k\|_\infty = 0$ indicates that (x^k, y^k, w^k) is a B-stationary point under the nondegenency condition. Hence, (5.1) is appropriate for a stopping criterion of the algorithm. For the CM function, we choose $g(\alpha) := ((\alpha^2 + 4)^{1/2} + \alpha)/2$ [9].

In this experiment, we solve the following test problem of the form (1.1) with a convex quadratic objective function:

$$\begin{aligned} \min\ & x^\top x + y^\top y \\ \text{s.t.}\ & Ax \le b \\ & w = Nx + My + q \\ & \mathcal{K} \ni y \perp w \in \mathcal{K}, \end{aligned} \tag{5.2}$$

where $(x, y, z) \in \mathbb{R}^2 \times \mathbb{R}^m \times \mathbb{R}^m$, and each element of $A \in \mathbb{R}^{2\times 2}$, $b \in \mathbb{R}^2$ and $N \in \mathbb{R}^{m\times 2}$ is randomly chosen from $[-50, 50]$, $[0, 100]$ and $[-5, 5]$, respectively. In addition, $M \in \mathbb{R}^{m\times m}$ is a positive semi-definite symmetric matrix generated by $M = (M_1 + M_1^\top) + 2\,|\lambda_{\min}(M_1 + M_1^\top)|I$, where $\lambda_{\min}$ denotes the minimum eigenvalue, I is the identity matrix, and $M_1 \in \mathbb{R}^{m\times m}$ is a matrix whose entries are randomly chosen from $[-50, 50]$. The vector $q \in \mathbb{R}^m$ is set to be $q := \xi_w - M\xi_y$ where each component of $\xi_y \in \mathbb{R}^m$ and $\xi_w \in \mathbb{R}^m$ is randomly chosen from $[-50, 50]$. We set $z^0 = (x^0, y^0, w^0) := (0, \xi_y, \xi_w) \in \mathbb{R}^2 \times \mathbb{R}^m \times \mathbb{R}^m$, so that $Ax^0 \le b$ and $w^0 = Nx^0 + My^0 + q$ are satisfied. As to the choice of $\mathcal{K}$, we set $\mathcal{K} := (\mathcal{K}^\nu)^\kappa$ with

$2 \leq \nu \leq 10$ and $\kappa = 10, 15$, and 20. We then generate 50 test problems for each $\mathcal{K}$ and solve them by the proposed algorithm. The obtained results are shown in Tables 1–3, where each column represents the following:

- m: the total dimension of $(\mathcal{K}^\nu)^\kappa$, i.e., $m = \nu\kappa$,
- $\sharp$ite: the average number of iterations for 50 problems,
- cpu(s): the average cpu-time in second for 50 problems,
- non(%): percentage of obtained solutions which are nondegenerate for the complementarity constraints.

Moreover, the histograms of "non(%)" are given by Figures 1–3. Tables 1–3, as well as Figures 1–3, correspond to the cases of $\kappa = 10$, 15, and 20, respectively, where κ is the number of SOCs in $\mathcal{K} = (\mathcal{K}^\nu)^\kappa$. Recall that convergence to the 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 the following. First, the values of cpu(s) tends to be larger as m increases, while $\sharp$ite does not change so much. Due to this observation, we can conclude that the total computational cost mainly depends not on the number of iterations, but on the size of QP subproblems. As the second observation, we can see that non(%) tends to be larger as the dimension ν of each SOC increases. Indeed, when $\mathcal{K} = (\mathcal{K}^2)^{10}$ and $(\mathcal{K}^3)^{10}$, the values of "non(%)" are just 2 and 14, respectively, while it becomes 100 when $\mathcal{K} = (\mathcal{K}^8)^{10}, (\mathcal{K}^9)^{10}$, and $(\mathcal{K}^{10})^{10}$. Thus, the result indicates that the proposed algorithm works more effectively in the case where the size of each SOC in $\mathcal{K}$ is large. We can see a similar tendency when $\kappa = 15$ or 20, which can be observed in Tables 2 and 3.

Table 1: Results for $\mathcal{K} = (\mathcal{K}^\nu)^{10}$

m	$(\mathcal{K}^\nu)^\kappa$	$\sharp$ite	cpu(s)	non(%)
20	$(\mathcal{K}^2)^{10}$	49.58	2.57	2
30	$(\mathcal{K}^3)^{10}$	50.16	3.07	14
40	$(\mathcal{K}^4)^{10}$	51.16	3.43	66
50	$(\mathcal{K}^5)^{10}$	50.96	3.80	80
60	$(\mathcal{K}^6)^{10}$	51.14	5.24	86
70	$(\mathcal{K}^7)^{10}$	51.14	8.89	94
80	$(\mathcal{K}^8)^{10}$	51.24	7.85	100
90	$(\mathcal{K}^9)^{10}$	50.90	12.31	100
100	$(\mathcal{K}^{10})^{10}$	50.98	13.37	100

Table 2: Results for $\mathcal{K} = (\mathcal{K}^{\nu})^{15}$

m	$(\mathcal{K}^{\nu})^{\kappa}$	♯ite	cpu(s)	non(%)
30	$(\mathcal{K}^{2})^{15}$	51.58	3.61	0
45	$(\mathcal{K}^{3})^{15}$	52.80	4.57	12
60	$(\mathcal{K}^{4})^{15}$	53.32	6.02	40
75	$(\mathcal{K}^{5})^{15}$	54.12	8.60	56
90	$(\mathcal{K}^{6})^{15}$	53.60	11.04	78
105	$(\mathcal{K}^{7})^{15}$	53.58	16.86	90
120	$(\mathcal{K}^{8})^{15}$	52.94	19.84	98
135	$(\mathcal{K}^{9})^{15}$	53.02	24.84	100
150	$(\mathcal{K}^{10})^{15}$	53.18	31.33	100

Table 3: Results for $\mathcal{K} = (\mathcal{K}^{\nu})^{20}$

m	$(\mathcal{K}^{\nu})^{\kappa}$	♯ite	cpu(s)	non(%)
40	$(\mathcal{K}^{2})^{20}$	53.70	7.52	0
60	$(\mathcal{K}^{3})^{20}$	54.84	8.89	6
80	$(\mathcal{K}^{4})^{20}$	55.94	13.13	26
100	$(\mathcal{K}^{5})^{20}$	56.06	15.67	46
120	$(\mathcal{K}^{6})^{20}$	55.34	21.14	80
140	$(\mathcal{K}^{7})^{20}$	55.38	31.93	82
160	$(\mathcal{K}^{8})^{20}$	55.06	41.10	94
180	$(\mathcal{K}^{9})^{20}$	54.76	50.31	98
200	$(\mathcal{K}^{10})^{20}$	55.00	63.43	98

6 Conclusion

In this paper, we have considered the mathematical program with second-order cone (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 a vector equation by means of the natural residual and its smoothing function, and apply the SQP method with decreasing the smoothing parameter through iterations. The SQP type method proposed in this paper has the advantage that the exact solution of each subproblem can be calculated easily since it is a convex quadratic programming problem. We have further showed that the proposed algorithm possesses the global convergence property under the Cartesian P_0 and the nondegeneracy assumptions. We have confirmed the efficiency of the algorithm by means of numerical experiments.

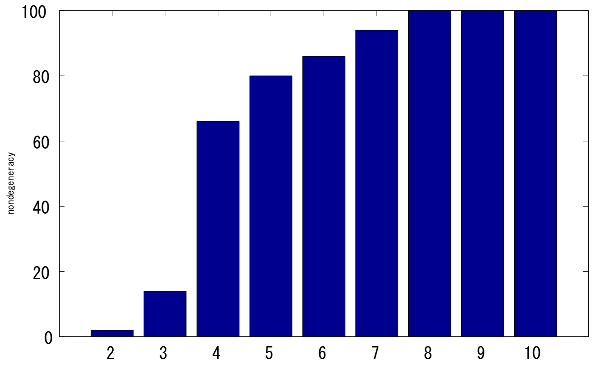

Figure 1: Nondegeneracy rate when $\kappa = 10$

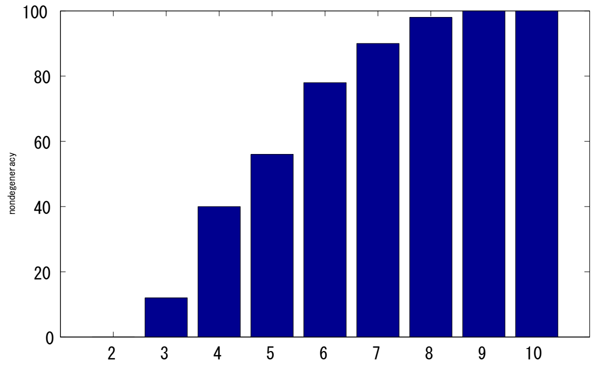

Figure 2: Nondegeneracy rate when $\kappa = 15$

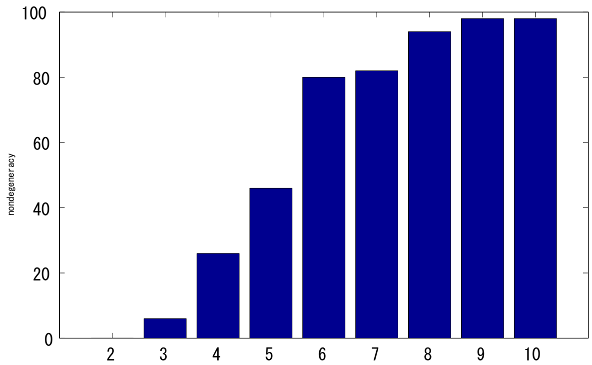

Figure 3: Nondegeneracy rate when $\kappa = 20$

7 Acknowledgments

First of all, I would like to express my sincere gratitude to my supervisor Assistant Professor Shunsuke Hayashi. He looked after me all night just before I turned in the master thesis. I would also like to express my gratitude to Professor Masao Fukushima. He gave me a lot of advice. Without his considerable advice, I could not achieve this paper. Moreover, I am also very grateful to Associate Professor Nobuo Yamashita. He gave me valuable comments from many various viewpoints in and after workshop. I also would like to thank to Okuno. He supported my study in several phase. Without his help, I could not complete this paper. Finally I am also very grateful to all members of Fukushima's Laboratory, my friends, and my family for help and encouragement.

References

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

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

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

[4] 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), pp. 39–56.

[5] 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), pp. 5–34.

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

[7] 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), pp. 724–739.

[8] 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), pp. 203–232.

[9] 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), pp. 593–615.

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

[11] N. Machida, M. Fukushima, and T. Ibaraki, *A multisplitting method for symmetric linear complementarity problems*, Journal of Computational and Applied Mathematics, 62 (1995), pp. 217–227.

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

[13] 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), pp. 1475–1491.

[14] B. T. Polyak, *Introduction to Optimization*, Optimization Software, Publications Division, New York, 1987.

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