A Combined Smoothing and Regularization Method for Monotone Second-Order Cone Complementarity Problems*

Shunsuke Hayashi, Nobuo Yamashita and Masao Fukushima†

January 20, 2003

Abstract. The Second-Order Cone Complementarity Problem (SOCCP) is a wide class of problems containing the Nonlinear Complementarity Problem (NCP) and the Second-Order Cone Programming Problem (SOCP). Recently, Fukushima, Luo and Tseng extended some merit functions and their smoothing functions for NCP to SOCCP. Moreover, they derived computable formulas for the Jacobians of the smoothing functions and gave conditions for the Jacobians to be invertible. In this paper, we propose a globally and quadratically convergent algorithm, which is based on smoothing and regularization methods, for solving monotone SOCCP. In particular, we study strong semismoothness and Jacobian consistency, which play an important role in establishing quadratic convergence of the algorithm. Furthermore, we examine effectiveness of the algorithm by means of numerical experiments.

Key words. second-order cone, complementarity problem, smoothing method, regularization method

AMS subject classifications. 90C33, 65K05

1 Introduction

In this paper, we consider the *second-order cone complementarity problem* (SOCCP) [11]:

$$\begin{array}{rl} \text{Find} & (x,y,\zeta) \in \Re^n \times \Re^n \times \Re^l \\ \text{such that} & x \in \mathcal{K},\ y \in \mathcal{K},\ x^T y = 0,\ F(x,y,\zeta) = 0, \end{array} \tag{1.1}$$

where $F : \Re^n \times \Re^n \times \Re^l \to \Re^n \times \Re^l$ is a continuously differentiable mapping, and $\mathcal{K} \subset \Re^n$ is the Cartesian product of second-order cones, that is,

$$\mathcal{K} = \mathcal{K}^{n_1} \times \mathcal{K}^{n_2} \times \cdots \times \mathcal{K}^{n_m}$$

with $n = n_1 + \cdots + n_m$ and $\mathcal{K}^{n_i} \subset \Re^{n_i}$ being the n_i-dimensional second-order cone defined by

$$\mathcal{K}^{n_i} := \left\{ (z_1, z_2^T)^T \in \Re \times \Re^{n_i - 1} \;\middle|\; \|z_2\| \le z_1 \right\}.$$

Throughout this paper, $\|\cdot\|$ denotes the 2-norm defined by $\|z\| := \sqrt{z^T z}$ for a vector z and $\|M\| := \max_{z \in \Re^n \setminus \{0\}} \|Mz\|/\|z\|$ for a matrix M. We will often write $z = (z_1, z_2)$ for $(z_1, z_2^T)^T$ for simplicity of notation.

The SOCCP contains a wide class of problems such as the Nonlinear Complementarity Problem (NCP) and the Second-Order Cone Programming Problem (SOCP) [13]. To see this, consider the

*This research was supported in part by a Grant-in-Aid for Scientific Research (B) from the Ministry of Education, Science, Sports and Culture of Japan.

†Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan (hayashi@amp.i.kyoto-u.ac.jp, nobuo@i.kyoto-u.ac.jp, fuku@i.kyoto-u.ac.jp)

SOCCP with $n_1 = n_2 = \cdots = n_m = 1$ and $F(x, y, \zeta) = f(x) - y$, where $f : \Re^n \to \Re^n$. Then the problem reduces to the NCP : Find $x \in \Re^n$ such that

$$x \geq 0,\ f(x) \geq 0,\ x^T f(x) = 0.$$

Next, consider the SOCP :

$$\begin{array}{ll} \text{Minimize} & \theta(z) \\ \text{subject to} & \gamma(z) \in \mathcal{K}, \end{array} \tag{1.2}$$

where $\theta : \Re^s \to \Re$ and $\gamma : \Re^s \to \Re^t$. Then the Karush-Kuhn-Tucker (KKT) conditions for SOCP (1.2) can be written as

$$\begin{array}{l} \nabla\theta(z) - \nabla\gamma(z)\lambda = 0, \\ \lambda \in \mathcal{K},\ \ \gamma(z) \in \mathcal{K},\ \ \lambda^T\gamma(z) = 0, \end{array} \tag{1.3}$$

where $\lambda \in \Re^t$ is the Lagrange multiplier vector. Then, by setting $n := t$, $l := s$, $x := \lambda$, $y := \mu$, $\zeta := z$ and

$$F(x, y, \zeta) := \begin{pmatrix} \mu - \gamma(z) \\ \nabla\theta(z) - \nabla\gamma(z)\lambda \end{pmatrix},$$

the KKT conditions (1.3) can be rewritten as SOCCP (1.1). If θ is a convex function and $\{w \in \Re^n \mid \gamma(w) \in \mathcal{K}\}$ is a convex set, then a point z satisfying the KKT conditions (1.3) is a solution of SOCP (1.2).

Recently, a number of methods for solving SOCCP have been proposed. For SOCP (1.2) with affine θ and γ, primal-dual interior-point methods [13, 15, 22] are shown to be effective. For SOCCP, Fukushima, Luo and Tseng [11], Chen, Sun and Sun [5], and Chen, Chen and Tseng [2] studied smoothing and nonsmooth approaches. Especially, Fukushima, Luo and Tseng [11] showed that the min function and the Fischer-Burmeister function for the NCP can be extended to the SOCCP by means of Jordan algebra. Furthermore, they constructed smoothing functions for those functions and analyzed the properties of their Jacobians.

In this paper, we introduce not only smoothing methods but also regularization methods for SOCCPs. Smoothing methods have effectively been adopted to deal with nondifferentiable reformulations of complementarity problems [1, 3, 4, 12, 19, 21, 23]. On the other hand, regularization methods have provided a fundamental tool to deal with ill-posed problems [8, 9]. By combining these methods, we develop a hybrid algorithm for solving monotone SOCCPs. Moreover, we show conditions for the algorithm to be globally and quadratically convergent. Particularly, we highlight the two properties called *strong semismoothness* and *Jacobian consistency*, which play a crucial role in establishing quadratic convergence of the algorithm.

This paper is organized as follows. In Section 2, we review the spectral factorization associated with a second-order cone, which plays a key role in analyzing the properties of merit functions for the SOCCP. Moreover, we construct a merit function by means of the natural residual for the SOCCP. In Section 3, we propose a prototype algorithm for solving the SOCCP and show that it has global convergence when applied to monotone SOCCPs. In Section 4, by incorporating Newton's method in the algorithm proposed in Section 3, we construct a concrete algorithm and establish quadratic convergence. In Section 5, we present some numerical results with the latter algorithm. In Section 6, we conclude the paper with some remarks.

2 Preliminaries

2.1 Spectral Factorization

We briefly review some properties of the spectral factorization with respect to a second-order cone, which will be used in the subsequent analysis. Spectral factorization is one of the basic concepts in

Jordan algebra. For more detail, see [10, 11].

For any vector $z = (z_1, z_2) \in \Re \times \Re^{n-1}$ $(n \geq 2)$, its spectral factorization with respect to the second-order cone $\mathcal{K}^n$ is defined as

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

where λ_1 and λ_2 are the *spectral values* given by

$$\lambda_i \;=\; z_1 + (-1)^i \|z_2\|, \quad i = 1, 2, \tag{2.1}$$

and u^1 and u^2 are the *spectral vectors* given by

$$u^i \;=\; \begin{cases} \frac{1}{2}\left(1, (-1)^i \frac{z_2}{\|z_2\|}\right) & (z_2 \neq 0), \\ \frac{1}{2}(1, (-1)^i w) & (z_2 = 0), \end{cases} \qquad i = 1, 2, \tag{2.2}$$

with $w \in \Re^{n-1}$ such that $\|w\| = 1$. The spectral values and vectors have the following properties. For any $z \in \Re^n$, the inequality $\lambda_1 \leq \lambda_2$ holds, and

$$\lambda_1 \geq 0 \iff z \in \mathcal{K}^n.$$

Moreover, for any $z \in \Re^n$, we have $\|u^i\| = 1/\sqrt{2}$ for $i = 1, 2$, and $(u^1)^T u^2 = 0$.

For any $z \in \Re^n$, let $P_{\mathcal{K}^n}(z)$ denote the projection of z onto the second-order cone $\mathcal{K}^n$, that is,

$$P_{\mathcal{K}^n}(z) = \operatorname*{argmin}_{z' \in \mathcal{K}^n} \|z' - z\|. \tag{2.3}$$

In particular, when $n = 1$,

$$P_{\mathcal{K}^1}(\alpha) \;=\; \max\{0, \alpha\}$$

for any $\alpha \in \Re$. For $n \geq 2$, the projection function $P_{\mathcal{K}^n}$ can also be calculated easily as shown in the following proposition.

Proposition 2.1 [11, Proposition 3.3] *For any $z \in \Re^n$ $(n \geq 2)$,*

$$P_{\mathcal{K}^n}(z) \;=\; \max\{0, \lambda_1\} u^1 + \max\{0, \lambda_2\} u^2, \tag{2.4}$$

where λ_1 and λ_2 are the spectral values of z defined by (2.1)*, and u^1 and u^2 are the spectral vectors of z defined by* (2.2).

2.2 Merit Function

In this paper, we consider the unconstrained optimization reformulation of SOCCP (1.1):

$$\text{Minimize} \;\; \Psi(x, y, \zeta),$$

where Ψ is a real-valued function on $\Re^n \times \Re^n \times \Re^l$. The objective function Ψ is called a *merit function* for SOCCP (1.1). In order to construct a merit function for SOCCP (1.1), it is convenient to introduce a function $\Phi : \Re^n \times \Re^n \to \Re^n$ satisfying

$$\Phi(x, y) = 0 \iff x \in \mathcal{K},\; y \in \mathcal{K},\; x^T y = 0. \tag{2.5}$$

By using such a function, we define $H : \Re^n \times \Re^n \times \Re^l \to \Re^{2n+l}$ by

$$H(x, y, \zeta) := \begin{pmatrix} \Phi(x, y) \\ F(x, y, \zeta) \end{pmatrix}.$$

It is obvious that SOCCP (1.1) is equivalent to the equation $H(x,y,\zeta)=0$. Moreover, we define the function $\Psi:\Re^n\times\Re^n\times\Re^l\to\Re$ by

$$\begin{aligned}\Psi(x,y,\zeta) &:= \frac{1}{2}\|H(x,y,\zeta)\|^2\\ &= \frac{1}{2}\|\Phi(x,y)\|^2+\frac{1}{2}\|F(x,y,\zeta)\|^2.\end{aligned}\tag{2.6}$$

Then, it is easy to see that $\Psi(x,y,\zeta)\ge 0$ for any $(x,y,\zeta)\in\Re^n\times\Re^n\times\Re^l$, and that $\Psi(x,y,\zeta)=0$ if and only if (x,y,ζ) is a solution of SOCCP (1.1). Therefore, the function Ψ defined by (2.6) can serve as a merit function for SOCCP (1.1).

The following proposition shows that the complementarity condition on $\mathcal{K}=\mathcal{K}^{n_1}\times\cdots\times\mathcal{K}^{n_m}$ can be decomposed into complementarity conditions on each $\mathcal{K}^{n_i}$.

Proposition 2.2 *Let $\mathcal{K}=\mathcal{K}^{n_1}\times\cdots\times\mathcal{K}^{n_m}$, $x=(x^1,\dots,x^m)\in\Re^{n_1}\times\cdots\times\Re^{n_m}$ and $y=(y^1,\dots,y^m)\in\Re^{n_1}\times\cdots\times\Re^{n_m}$. Then the following relation holds:*

$$x\in\mathcal{K},\ y\in\mathcal{K},\ x^Ty=0 \quad \textit{if and only if} \quad x^i\in\mathcal{K}^{n_i},\ y^i\in\mathcal{K}^{n_i},\ (x^i)^Ty^i=0 \quad (i=1,\dots,m).$$

Proof. Since "if" part is evident, we only show "only if" part. Noticing that $x^i,y^i\in\mathcal{K}^{n_i}$ implies $x_1^i\ge\|x_2^i\|$ and $y_1^i\ge\|y_2^i\|$, we have $(x^i)^Ty^i=x_1^iy_1^i+(x_2^i)^Ty_2^i\ge\|x_2^i\|\|y_2^i\|+(x_2^i)^Ty_2^i\ge 0$ for each i, where the last inequality follows from the Cauchy-Schwarz inequality. On the other hand, we have $\sum_{i=1}^m(x^i)^Ty^i=0$ from $x^Ty=0$. Hence, it holds $(x^i)^Ty^i=0$ for each i. ∎

This proposition naturally leads us to construct a function Φ satisfying (2.5) as

$$\Phi(x,y):=\begin{pmatrix}\varphi^1(x^1,y^1)\\ \vdots\\ \varphi^m(x^m,y^m)\end{pmatrix},$$

where $\varphi^i:\Re^{n_i}\times\Re^{n_i}\to\Re^{n_i}$ is a function satisfying

$$\varphi^i(x^i,y^i)=0 \iff x^i\in\mathcal{K}^{n_i},\ y^i\in\mathcal{K}^{n_i},(x^i)^Ty^i=0\tag{2.7}$$

for each $i=1,\dots,m$. Fukushima, Luo and Tseng [11] showed that (2.7) holds for the *natural residual* function $\varphi^i_{\rm NR}:\Re^{n_i}\times\Re^{n_i}\to\Re^{n_i}$ defined by

$$\varphi^i_{\rm NR}(x^i,y^i):=x^i-P_{\mathcal{K}^{n_i}}(x^i-y^i).\tag{2.8}$$

Using this function, we define the function $\Phi_{\rm NR}:\Re^n\times\Re^n\to\Re^n$ by

$$\Phi_{\rm NR}(x,y):=\begin{pmatrix}\varphi^1_{\rm NR}(x^1,y^1)\\ \vdots\\ \varphi^m_{\rm NR}(x^m,y^m)\end{pmatrix}\tag{2.9}$$

and $H_{\rm NR}:\Re^n\times\Re^n\times\Re^l\to\Re^{2n+l}$ by

$$H_{\rm NR}(x,y,\zeta):=\begin{pmatrix}\Phi_{\rm NR}(x,y)\\ F(x,y,\zeta)\end{pmatrix}.\tag{2.10}$$

Then, we can construct a merit function $\Psi_{\rm NR}:\Re^n\times\Re^n\times\Re^l\to\Re$ for SOCCP (1.1) as

$$\begin{aligned}\Psi_{\rm NR}(x,y,\zeta) &:= \frac{1}{2}\|H_{\rm NR}(x,y,\zeta)\|^2\\ &= \frac{1}{2}\|\Phi_{\rm NR}(x,y)\|^2+\frac{1}{2}\|F(x,y,\zeta)\|^2\\ &= \frac{1}{2}\sum_{i=1}^m\|\varphi_{\rm NR}(x^i,y^i)\|^2+\frac{1}{2}\|F(x,y,\zeta)\|^2.\end{aligned}\tag{2.11}$$

In what follows, we write $\varphi_{\rm NR}$ for $\varphi^i_{\rm NR}$ for simplicity of notation.

3 Globally convergent method

In the previous section, we have constructed the merit function $\Psi_{\rm NR}$ from the natural residual $\varphi_{\rm NR}$. We can solve SOCCP (1.1) by minimizing $\Psi_{\rm NR}$ by an appropriate descent algorithm. However, the function $\Psi_{\rm NR}$ is not differentiable, and hence, methods that use the gradient of the function, such as the steepest descent method and Newton's method, are not applicable. In order to get rid of this difficulty, we first introduce a smoothing method that solves a sequence of differentiable approximations to the original nondifferentiable problem. To ensure global convergence of a descent method, the level-boundedness of the objective function plays an important role. As shown later, if the function involved in the SOCCP is strongly monotone, then the merit function $\Psi_{\rm NR}$ is level-bounded. But the assumption of strong monotonicity is quite restrictive from a practical standpoint. To be amenable to a merely monotone problem, we propose to combine a regularization method with a smoothing method.

In the remainder of the paper, we restrict ourselves to the SOCCP in which (i) F is given by $F(x,y,\zeta) = f(x) - y$ with a continuously differentiable function $f : \Re^n \to \Re^n$ and (ii) $\mathcal{K} = \mathcal{K}^n$. Then we can rewrite SOCCP (1.1) as follows: Find $(x,y) \in \Re^n \times \Re^n$ such that

$$x \in \mathcal{K}^n,\ \ y \in \mathcal{K}^n,\ \ x^T y = 0,\ \ y = f(x). \tag{3.1}$$

Moreover, since the assumption (ii) and (2.9) imply $\Phi_{\rm NR}(x,y) = \varphi_{\rm NR}(x,y)$, the vector-valued function $H_{\rm NR}$ defined by (2.10) is rewritten as

$$H_{\rm NR}(x,y) = \begin{pmatrix} \varphi_{\rm NR}(x,y) \\ f(x) - y \end{pmatrix} \tag{3.2}$$

and the merit function $\Psi_{\rm NR}$ is rewritten as

$$\begin{aligned} \Psi_{\rm NR}(x,y) &= \frac{1}{2}\|H_{\rm NR}(x,y)\|^2 \\ &= \frac{1}{2}\|\varphi_{\rm NR}(x,y)\|^2 + \frac{1}{2}\|f(x) - y\|^2. \end{aligned} \tag{3.3}$$

Note that the assumption (i) implies F does not involve the additional variable ζ. This assumption may seem rather restrictive. However, the KKT conditions for SOCP (1.2) can be written in the form of the SOCCP with the assumption (i); for more detail, see Section 6. The assumption (ii) is only for simplicity of presentation. In view of Proposition 2.2, the results obtained in the following can be extended to the general $\mathcal{K}$ in a straightforward manner.

3.1 Smoothing functions

In this subsection, we introduce a class of smoothing functions of the merit function $\Psi_{\rm NR}$. For a nondifferentiable function $h : \Re^n \to \Re^m$, we consider a function $h_\mu : \Re^n \to \Re^m$ with a parameter $\mu > 0$ that has the following properties:

(a) h_μ is differentiable for any $\mu > 0$,

(b) $\lim_{\mu \downarrow 0} h_\mu(x) = h(x)$ for any $x \in \Re^n$.

Such a function h_μ is called a *smoothing function* of h. Instead of handling the nonsmooth equation $h(x) = 0$ directly, the smoothing method solves a family of smoothed subproblems $h_\mu(x) = 0$ for $\mu > 0$, and obtain a solution of the original problem by letting $\mu \downarrow 0$. Fukushima, Luo and Tseng [11] extended Chen and Mangasarian's class [1] of smoothing functions for NCP to SOCCP, which may be regarded as a smoothing function of the natural residual $\varphi_{\rm NR}$.

First we define a smoothing function of the projection function $P_{\mathcal{K}^n}$ defined by (2.3). To this end, we consider a continuously differentiable convex function $\hat{g} : \Re \to \Re$ such that

$$\lim_{\alpha \to -\infty} \hat{g}(\alpha) = 0,\ \ \lim_{\alpha \to \infty} (\hat{g}(\alpha) - \alpha) = 0,\ \ 0 < \hat{g}'(\alpha) < 1. \tag{3.4}$$

For example, $\hat{g}_1(\alpha) = (\sqrt{\alpha^2+4}+\alpha)/2$ and $\hat{g}_2(\alpha) = \ln(e^\alpha + 1)$ satisfy (3.4). By using $\hat{g}$, we define the function $P_\mu : \Re^n \to \Re^n$ by

$$P_\mu(z) := \mu\hat{g}(\lambda_1/\mu)u^1 + \mu\hat{g}(\lambda_2/\mu)u^2, \tag{3.5}$$

where λ_1 and λ_2 are the spectral values of z given by (2.1), and u^1 and u^2 are the spectral vectors of z given by (2.2). Fukushima, Luo and Tseng [11] showed that the function P_μ defined by (3.5) is a smoothing function of $P_{\mathcal{K}^n}$, by using the facts that $\lim_{\mu\downarrow 0}\mu\hat{g}(\lambda/\mu) = \max\{0,\lambda\}$ and that $\gamma_\mu(\lambda) := \mu\hat{g}(\lambda/\mu)$ is differentiable for any $\mu > 0$. Hence, from the definition (2.8) of $\varphi_{\rm NR}$, the function $\varphi_\mu : \Re^n \to \Re^n$ defined by

$$\varphi_\mu(x,y) := x - P_\mu(x-y)$$

becomes a smoothing function of $\varphi_{\rm NR}$. In particular, by [11, Proposition 5.1], there exists a positive constant ν such that

$$\|\varphi_\mu(x,y) - \varphi_{\rm NR}(x,y)\| \le \nu\mu \tag{3.6}$$

for any $\mu > 0$ and $(x,y) \in \Re^n \times \Re^n$. Likewise, the function $\Psi_\mu : \Re^n \times \Re^n \to \Re$ given by

$$\Psi_\mu(x,y) := \frac{1}{2}\|\varphi_\mu(x,y)\|^2 + \frac{1}{2}\|f(x)-y\|^2$$

serves as a smoothing function of the merit function $\Psi_{\rm NR}$. In the following, we denote $P_0(z) := P_{\mathcal{K}^n}(z)$, $\varphi_0(x,y) := \varphi_{\rm NR}(x,y)$ and $\Psi_0(x,y) := \Psi_{\rm NR}(x,y)$.

3.2 Level-boundedness

Since Ψ_μ is differentiable, we may apply an appropriate gradient-based descent method to obtain a minimum (x_μ, y_μ) of the function Ψ_μ for $\mu > 0$. In order that a sequence generated by such a descent method has an accumulation point, it is desirable that Ψ_μ is level-bounded, that is, the level sets $\mathcal{L}_\alpha := \{(x,y) \mid \Psi_\mu(x,y) \le \alpha\}$ are bounded for all $\alpha \in \Re$. Note that the level-boundedness of a function $\Psi : \Re^n \to \Re$ is equivalent to

$$\lim_{\|z\|\to\infty} \Psi(z) = +\infty. \tag{3.7}$$

We first recall the concepts of monotonicity and strong monotonicity of vector-valued functions.

Definition 3.1 *The function $f : \Re^n \to \Re^n$ is said to be*

(a) *monotone if, for any $(x,y) \in \Re^n \times \Re^n$,*

$$(x-y)^T(f(x)-f(y)) \ge 0;$$

(b) *strongly monotone with modulus $\varepsilon > 0$ if, for any $(x,y) \in \Re^n \times \Re^n$,*

$$(x-y)^T(f(x)-f(y)) \ge \varepsilon\|x-y\|^2.$$

It is obvious that a strongly monotone function is monotone. Moreover, if f is monotone, then the mapping $f_\varepsilon : \Re^n \to \Re^n$ defined by $f_\varepsilon(x) := f(x) + \varepsilon x$ is strongly monotone for any $\varepsilon > 0$.

Now, let the function $\tilde{\Psi}_{\rm NR} : \Re^n \to \Re$ be defined by

$$\tilde{\Psi}_{\rm NR}(x) := \|\varphi_{\rm NR}(x, f(x))\|. \tag{3.8}$$

Then the following two lemmas hold.

Lemma 3.1 *If f is strongly monotone, then $\tilde{\Psi}_{\rm NR}$ is level-bounded.*

This result was proved for the variational inequality problem by Peng and Fukushima [17]. Recently, Yamashita and Fukushima [24] generalized this result by introducing a concept of strong coerciveness, which is weaker than the strong monotonicity.

Lemma 3.2 *$\tilde{\Psi}_{\rm NR}$ is level-bounded if and only if $\Psi_{\rm NR}$ is level-bounded.*

Proof. We first show "if" part. Suppose that $\Psi_{\rm NR}$ is level-bounded. Then we have

$$\lim_{\|(x,y)\|\to\infty} \Psi_{\rm NR}(x,y) = +\infty. \tag{3.9}$$

Let $\{x^{(k)}\}$ be an arbitrary sequence such that $\|x^{(k)}\| \to \infty$ and $\{y^{(k)}\}$ be the corresponding sequence such that $y^{(k)} = f(x^{(k)})$ for all k. Then we have

$$\begin{aligned}
\frac{1}{2}\tilde{\Psi}_{\rm NR}(x^{(k)})^2 &= \frac{1}{2}\|\varphi_{\rm NR}(x^{(k)}, f(x^{(k)}))\|^2 \\
&= \frac{1}{2}\|\varphi_{\rm NR}(x^{(k)}, y^{(k)})\|^2 + \frac{1}{2}\|f(x^{(k)}) - y^{(k)}\|^2 \\
&= \Psi_{\rm NR}(x^{(k)}, y^{(k)}),
\end{aligned}$$

where the second equality follows from $y^{(k)} = f(x^{(k)})$. Noticing that $\|(x^{(k)}, y^{(k)})\| \to \infty$ as $k \to \infty$ and that (3.9) holds, we have $\lim_{k\to\infty} \tilde{\Psi}_{\rm NR}(x^{(k)}) = +\infty$, that is, $\tilde{\Psi}_{\rm NR}$ is level-bounded.

We next show the "only if" part. Suppose that $\tilde{\Psi}_{\rm NR}$ is level-bounded. Then $\tilde{\Psi}_{\rm NR}$ satisfies (3.7). Noticing $\sqrt{2\|\xi\|^2 + 2\|\eta\|^2} \geq \|\xi\| + \|\eta\|$ for any $\xi,\ \eta \in \Re^n$, we have

$$\begin{aligned}
2\sqrt{\Psi_{\rm NR}(x,y)} &\geq \|x - P_{\mathcal{K}^n}(x-y)\| + \|f(x) - y\| \\
&\geq \|x - P_{\mathcal{K}^n}(x - f(x))\| - \|P_{\mathcal{K}^n}(x - f(x)) - P_{\mathcal{K}^n}(x-y)\| + \|f(x) - y\| \\
&\geq \|x - P_{\mathcal{K}^n}(x - f(x))\| - \|f(x) - y\| + \|f(x) - y\| \\
&= \tilde{\Psi}_{\rm NR}(x),
\end{aligned}$$

where the second inequality follows from the triangle inequality, and the third inequality follows from the nonexpansiveness of the projection operator. Hence, $\Psi_{\rm NR}$ also satisfies (3.7), that is, $\Psi_{\rm NR}$ is level-bounded. ∎

We now give the main theorem of this subsection.

Theorem 3.1 *If f is strongly monotone, then Ψ_μ is level-bounded for any $\mu \geq 0$.*

Proof. By (3.6), we have

$$\begin{aligned}
\sqrt{2\Psi_{\rm NR}(x,y)} &= \left\| \begin{pmatrix} \varphi_{\rm NR}(x,y) \\ f(x) - y \end{pmatrix} \right\| \\
&\leq \left\| \begin{pmatrix} \varphi_\mu(x,y) \\ f(x) - y \end{pmatrix} \right\| + \left\| \begin{pmatrix} \varphi_{\rm NR}(x,y) - \varphi_\mu(x,y) \\ 0 \end{pmatrix} \right\| \\
&\leq \sqrt{2\Psi_\mu(x,y)} + \nu\mu.
\end{aligned}$$

From Lemmas 3.1 and 3.2, $\Psi_{\rm NR}$ is level-bounded if f is strongly monotone. This implies the level-boundedness of Ψ_μ. ∎

3.3 Regularization method

Theorem 3.1 implies that, if f is strongly monotone, any sequence generated by an appropriate descent method applied to Ψ_μ has an accumulation point. However, the strong monotonicity is quite a severe condition. As a remedy for this inconvenience, we employ a regularization method.

Let the function $f_\varepsilon : \Re^n \to \Re^n$ be defined by $f_\varepsilon(x) := f(x) + \varepsilon x$ with a positive parameter ε. The regularization method solves a sequence of SOCCPs involving f_ε with $\varepsilon > 0$, so that a solution of the original SOCCP is obtained by taking the limit $\varepsilon \downarrow 0$. Define the functions $H_{\mu,\varepsilon} : \Re^n \times \Re^n \to \Re^{2n}$ and $\Psi_{\mu,\varepsilon} : \Re^n \times \Re^n \to \Re$ by

$$H_{\mu,\varepsilon}(x,y) \;:=\; \begin{pmatrix} \varphi_\mu(x,y) \\ f_\varepsilon(x) - y \end{pmatrix}, \tag{3.10}$$

$$\begin{aligned} \Psi_{\mu,\varepsilon}(x,y) \;&:=\; \frac{1}{2}\|H_{\mu,\varepsilon}(x,y)\|^2 \\ &=\; \frac{1}{2}\|\varphi_\mu(x,y)\|^2 + \frac{1}{2}\|f_\varepsilon(x) - y\|^2. \end{aligned} \tag{3.11}$$

If f is monotone, then f_ε is strongly monotone for any $\varepsilon > 0$, and hence, by Theorem 3.1, the function $\Psi_{\mu,\varepsilon}$ is level-bounded for any $\mu \geq 0$ and $\varepsilon > 0$.

3.4 Prototype algorithm

As we have shown in the previous subsections, for any $\mu > 0$ and $\varepsilon > 0$, the function $\Psi_{\mu,\varepsilon}$ defined by (3.11) is differentiable and level-bounded, provided f is monotone. Therefore, by applying an appropriate descent method, we may obtain a minimum $(x_{\mu,\varepsilon}, y_{\mu,\varepsilon})$ of the function $\Psi_{\mu,\varepsilon}$. Moreover, letting (μ,ε) converge to $(0,0)$, we can expect that $(x_{\mu,\varepsilon}, y_{\mu,\varepsilon})$ converges to a solution of the original SOCCP. However, in practice, it is usually impossible to compute an exact minimum of $\Psi_{\mu,\varepsilon}$. So, we consider the following algorithm in which the function $\Psi_{\mu,\varepsilon}$ is minimized only approximately at each iteration.

Algorithm 3.1

(Step 0) *Choose $(x^{(0)}, y^{(0)}) \in \Re^n \times \Re^n$, $\mu_0 \in (0,\infty)$, $\varepsilon_0 \in (0,\infty)$ and $\alpha_0 \in (0,\infty)$. Set $k := 0$.*
(Step 1) *Terminate the iteration if $\Psi_{\mathrm{NR}}(x^{(k)}, y^{(k)}) = 0$.*
(Step 2) *Find a pair $(x^{(k+1)}, y^{(k+1)}) \in \Re^n \times \Re^n$ such that*

$$\Psi_{\mu_{k+1},\,\varepsilon_{k+1}}(x^{(k+1)}, y^{(k+1)}) \;\leq\; \alpha_{k+1}. \tag{3.12}$$

(Step 3) *Update parameters $\mu_{k+1} \in (0,\mu_k)$, $\varepsilon_{k+1} \in (0,\varepsilon_k)$ and $\alpha_{k+1} \in (0,\alpha_k)$. Set $k := k+1$. Go back to Step 1.*

In this algorithm, $\{\mu_k\}$, $\{\varepsilon_k\}$ and $\{\alpha_k\}$ are chosen so as to tend to 0 eventually. To obtain $(x^{(k+1)}, y^{(k+1)})$ in Step 2, we may use any suitable unconstrained minimization technique. These issues will be discussed in detail in Section 4.

In order for Algorithm 3.1 to be well-defined, there must exist a pair $(x^{(k+1)}, y^{(k+1)})$ satisfying (3.12) for any $\alpha_{k+1} > 0$. For this purpose, we first show that any stationary point of $\Psi_{\mu,\varepsilon}$ is a global minimum of $\Psi_{\mu,\varepsilon}$ by using the nonsingularity of $\nabla H_{\mu,\varepsilon}(x,y)$.

Remark 3.1 *From the definition of $H_{\mu,\varepsilon}$, φ_μ and f_ε, $\nabla H_{\mu,\varepsilon}(x,y)$ can be written as*

$$\begin{aligned} \nabla H_{\mu,\varepsilon}(x,y) \;&=\; \begin{pmatrix} \nabla_x \varphi_\mu(x,y) & \nabla f(x) + \varepsilon I \\ \nabla_y \varphi_\mu(x,y) & -I \end{pmatrix} \\ &=\; \begin{pmatrix} I - \nabla P_\mu(x-y) & \nabla f(x) + \varepsilon I \\ \nabla P_\mu(x-y) & -I \end{pmatrix}. \end{aligned} \tag{3.13}$$

Let $g : \Re^n \to \Re^n$ *be defined by* $g(z) := \hat{g}(\lambda_1)u^1 + \hat{g}(\lambda_2)u^2$*, where* λ_1 *and* λ_2 *are the spectral values of* z*, and* u^1 *and* u^2 *are the spectral vectors of* z*. Then we have* $\nabla P_\mu(z) = \nabla g(z/\mu)$ *since* $P_\mu(z) = \mu g(z/\mu)$ *from* (3.5)*. Therefore, by* [11, Proposition 5.2]*,* $\nabla P_\mu(z)$ *is written as*

$$\nabla P_\mu(z) \;=\; \begin{cases} \hat{g}'(z_1/\mu)I & \text{if } z_2 = 0, \\ \begin{pmatrix} b_\mu & \dfrac{c_\mu z_2^T}{\|z_2\|} \\ \dfrac{c_\mu z_2}{\|z_2\|} & a_\mu I + (b_\mu - a_\mu)\dfrac{z_2 z_2^T}{\|z_2\|^2} \end{pmatrix} & \text{if } z_2 \neq 0, \end{cases} \tag{3.14}$$

where

$$a_\mu = \frac{\hat{g}(\lambda_2/\mu) - \hat{g}(\lambda_1/\mu)}{\lambda_2/\mu - \lambda_1/\mu},\;\; b_\mu = \frac{1}{2}(\hat{g}'(\lambda_2/\mu) + \hat{g}'(\lambda_1/\mu)),\;\; c_\mu = \frac{1}{2}(\hat{g}'(\lambda_2/\mu) - \hat{g}'(\lambda_1/\mu)). \tag{3.15}$$

Proposition 3.1 *If* $f : \Re^n \to \Re^n$ *is monotone, then the Jacobian* $\nabla H_{\mu,\varepsilon}(x, y)$ *is nonsingular for any* $\mu > 0$*,* $\varepsilon \geq 0$ *and* $(x, y) \in \Re^n \times \Re^n$*.*

Proof. Firstly, we show that $O \prec \nabla P_\mu(z) \prec I$ for any $z \in \Re^n$. When $z_2 = 0$, it is clear that $O \prec \nabla P_\mu(z) \prec I$ since $\nabla P_\mu(z) = \hat{g}'(z_1/\mu)I$ from (3.14) and $0 < \hat{g}'(z_1/\mu) < 1$ from (3.4). So we only consider the case where $z_2 \neq 0$. Noticing that $\nabla P_\mu(z)$ is given by (3.14), in order to show $\nabla P_\mu(z) \succ O$, it is sufficient to show that b_μ is positive and the Schur complement of $\nabla P_\mu(z)$ with respect to b_μ is positive definite. Since $b_\mu = (1/2)(\hat{g}'(\lambda_1/\mu) + \hat{g}'(\lambda_2/\mu))$, we have $b_\mu > 0$ from (3.4). On the other hand, the Schur complement of $\nabla P_\mu(z)$ with respect to b_μ is given by

$$\left\{ a_\mu I + (b_\mu - a_\mu)\frac{z_2 z_2^T}{\|z_2\|^2} \right\} - \frac{c_\mu^2}{b_\mu}\frac{z_2 z_2^T}{\|z_2\|^2} \;=\; a_\mu \left(I - \frac{z_2 z_2^T}{\|z_2\|^2} \right) + \frac{b_\mu^2 - c_\mu^2}{b_\mu}\frac{z_2 z_2^T}{\|z_2\|^2}.$$

If $a_\mu = (\hat{g}(\lambda_2/\mu) - \hat{g}(\lambda_1/\mu))/(\lambda_2/\mu - \lambda_1/\mu) \leq 0$, then the continuous differentiability of $\hat{g}$ and the mean value theorem guarantee the existence of $\tau \in [\lambda_1/\mu, \lambda_2/\mu]$ such that $\hat{g}'(\tau) \leq 0$. Since this fact contradicts (3.4), we have $a_\mu > 0$. Moreover, we have $(b_\mu^2 - c_\mu^2)/b_\mu = 2/(1/\hat{g}(\lambda_1) + 1/\hat{g}(\lambda_2)) > 0$ from (3.4). Furthermore, both $z_2 z_2^T/\|z_2\|^2$ and $I - z_2 z_2^T/\|z_2\|^2$ are positive semidefinite and their sum is the identity matrix. Hence, any positive linear combination of $z_2 z_2^T/\|z_2\|^2$ and $I - z_2 z_2^T/\|z_2\|^2$ is positive definite. Therefore, the Schur complement of $\nabla P_\mu(z)$ with respect to b_μ is positive definite, and hence we obtain $\nabla P_\mu(z) \succ O$. In a similar way, we can show

$$I - \nabla P_\mu(z) \;=\; \begin{pmatrix} 1 - b_\mu & -\dfrac{c_\mu z_2^T}{\|z_2\|} \\ -\dfrac{c_\mu z_2}{\|z_2\|} & (1 - a_\mu)I - (b_\mu - a_\mu)\dfrac{z_2 z_2^T}{\|z_2\|^2} \end{pmatrix} \succ O,$$

from $1 - b_\mu > 0$ and the positive definiteness of the Schur complement of $I - \nabla P_\mu(z)$ with respect to $1 - b_\mu$.

Secondly, we show the nonsingularity of $\nabla H_{\mu,\varepsilon}(x, y)^T$ instead of $\nabla H_{\mu,\varepsilon}(x, y)$. Let us denote $\mathcal{P} := \nabla P_\mu(z)$ and $\mathcal{F} := \nabla f(x)$ for convenience. Then $\nabla H_{\mu,\varepsilon}(x, y)^T$ can be written as

$$\nabla H_{\mu,\varepsilon}(x, y)^T \;=\; \begin{pmatrix} I - \mathcal{P} & \mathcal{P} \\ \mathcal{F} + \varepsilon I & -I \end{pmatrix}.$$

Let $\xi, \eta \in \Re^n$ satisfy

$$\nabla H_{\mu,\varepsilon}(x, y)^T \begin{pmatrix} \xi \\ \eta \end{pmatrix} \;=\; 0,$$

that is,

$$(I-\mathcal{P})\xi+\mathcal{P}\eta \;=\; 0, \tag{3.16}$$
$$(\mathcal{F}+\varepsilon I)\xi-\eta \;=\; 0. \tag{3.17}$$

Multiplying the left-hand side of (3.16) by $\mathcal{P}^{-1}$ and combining with (3.17), we have

$$(\mathcal{P}^{-1}-I+\mathcal{F}+\varepsilon I)\xi=0.$$

Since $O\prec\mathcal{P}\prec I$ implies $\mathcal{P}^{-1}\succ I$ and monotonicity of f implies $\mathcal{F}\succeq O$, $\mathcal{P}^{-1}-I+\mathcal{F}+\varepsilon I$ is positive definite. So we have $\xi=0$, and then $\eta=0$ from (3.17). Hence, $\nabla H_{\mu,\varepsilon}(x,y)^T$ is nonsingular, that is, $\nabla H_{\mu,\varepsilon}(x,y)$ is nonsingular. ∎

Proposition 3.2 *If $f:\Re^n\to\Re^n$ is monotone, then, for any $\mu>0$ and $\varepsilon\geq 0$, every stationary point $(\overline{x},\overline{y})$ of the function $\Psi_{\mu,\varepsilon}$ satisfies $\Psi_{\mu,\varepsilon}(\overline{x},\overline{y})=0$.*

Proof. Note that $\nabla\Psi_{\mu,\varepsilon}(\overline{x},\overline{y})=\nabla H_{\mu,\varepsilon}(\overline{x},\overline{y})\,H_{\mu,\varepsilon}(\overline{x},\overline{y})=0$. By Proposition 3.1, $\nabla H_{\mu,\varepsilon}(\overline{x},\overline{y})$ is nonsingular. Hence, we have $H_{\mu,\varepsilon}(\overline{x},\overline{y})=0$, that is, $\Psi_{\mu,\varepsilon}(\overline{x},\overline{y})=(1/2)\|H_{\mu,\varepsilon}(\overline{x},\overline{y})\|^2=0$. ∎

We proceed to showing the global convergence property of Algorithm 3.1 by extending the result of [9] for NCP to SOCCP. To this end, we give two lemmas. The first lemma implies that $\Psi_{\mu,\varepsilon}$ is uniformly continuous on a compact set not only in x and y but also in μ and ε.

Lemma 3.3 *Let $C\subset\Re^n\times\Re^n$ be a compact set. Then, for any $\delta>0$, there exist $\varepsilon'>0$ and $\mu'>0$ such that*

$$|\Psi_{\mu,\varepsilon}(x,y)-\Psi_{\rm NR}(x,y)|\;\leq\;\delta$$

for any $(x,y)\in C$, $\varepsilon\in[0,\varepsilon']$ and $\mu\in[0,\mu']$.

Proof. Define the function $\Omega:\Re^n\times\Re^n\times[0,+\infty)\times[0,+\infty)\to\Re$ by $\Omega(x,y,\mu,\varepsilon):=\Psi_{\mu,\varepsilon}(x,y)$. Then, Ω is continuous and satisfies $\Omega(x,y,0,0)=\Psi_{\rm NR}(x,y)$. Since any continuous function is uniformly continuous on a compact set, Ω is uniformly continuous on $C\times[0,\varepsilon']\times[0,\mu']$. ∎

The next lemma is known as the Mountain Pass Theorem, which is useful for our analysis. For more detail, see Theorem 9.2.7 in [16].

Lemma 3.4 (Mountain Pass Theorem) *Let $\theta:\Re^n\to\Re$ be a continuously differentiable and level-bounded function. Let $C\subset\Re^n$ be a nonempty and compact set and let m be the minimum value of θ on the boundary of C, that is,*

$$m:=\min_{x\in\partial C}\theta(x).$$

Assume that there exist points $p\in C$ and $q\notin C$ such that $\theta(p)<m$ and $\theta(q)<m$. Then, there exists a point $r\in\Re^n$ such that $\nabla\theta(r)=0$ and $\theta(r)\geq m$.

Finally, by using the above lemmas, we establish the global convergence of Algorithm 3.1.

Theorem 3.2 *Let $f:\Re^n\to\Re^n$ be a monotone function. Assume that the solution set $\mathcal{S}$ of SOCCP* (3.1) *is nonempty and bounded. Let $\{(x^{(k)},y^{(k)})\}$ be a sequence generated by Algorithm 3.1. Then, $\{(x^{(k)},y^{(k)})\}$ is bounded, and every accumulation point is a solution of* (3.1).

Proof. From a simple continuity argument, we can easily show that every accumulation point of $\{(x^{(k)}, y^{(k)})\}$ is a solution of SOCCP (3.1). So we only show the boundedness of $\{(x^{(k)}, y^{(k)})\}$. For a contradiction purpose, we assume that $\{(x^{(k)}, y^{(k)})\}$ is not bounded. Then, there exists a subsequence $\{(x^{(k)}, y^{(k)})\}_{k\in K}$ such that $\lim_{k\to\infty,\, k\in K} \|(x^{(k)}, y^{(k)})\| = \infty$. Moreover, noticing that there exists a compact set $C \subset \Re^n \times \Re^n$ such that $\mathcal{S} \subset \operatorname{int} C$ because of the boundedness of $\mathcal{S}$, we have the following two facts:

(a) $(x^{(k)}, y^{(k)}) \notin C$ for all $k \in K$ sufficiently large;

(b) $\overline{m} := \min_{(x,y)\in\partial C} \Psi_{\rm NR}(x, y) > 0$.

Now, applying Lemma 3.3 with $\delta := \overline{m}/4 > 0$, we have

$$\Psi_{\mu_k,\varepsilon_k}(x, y) - \Psi_{\rm NR}(x, y) \le \frac{1}{4}\overline{m} \tag{3.18}$$

and

$$\Psi_{\mu_k,\varepsilon_k}(x, y) - \Psi_{\rm NR}(x, y) \ge -\frac{1}{4}\overline{m} \tag{3.19}$$

for any $(x, y) \in C$ and $k \in K$ sufficiently large. Let $(\overline{x}, \overline{y}) \in \mathcal{S} \subset C$ be a solution of SOCCP. Then, from (3.18), we have

$$\Psi_{\mu_k,\varepsilon_k}(\overline{x}, \overline{y}) - \Psi_{\rm NR}(\overline{x}, \overline{y}) = \Psi_{\mu_k,\varepsilon_k}(\overline{x}, \overline{y}) \le \frac{1}{4}\overline{m} \tag{3.20}$$

for all $k \in K$ sufficiently large. On the other hand, letting $(\tilde{x}^{(k)}, \tilde{y}^{(k)})$ be a minimizer of $\Psi_{\mu_k,\varepsilon_k}(x, y)$ on the set ∂C, we have, for all $k \in K$ sufficiently large,

$$\begin{aligned} \min_{(x,y)\in\partial C} \Psi_{\mu_k,\varepsilon_k}(x, y) &= \Psi_{\mu_k,\varepsilon_k}(\tilde{x}^{(k)}, \tilde{y}^{(k)}) \\ &\ge -\frac{1}{4}\overline{m} + \Psi_{\rm NR}(\tilde{x}^{(k)}, \tilde{y}^{(k)}) \\ &\ge -\frac{1}{4}\overline{m} + \overline{m} \\ &= \frac{3}{4}\overline{m}, \end{aligned} \tag{3.21}$$

where the first inequality follows from (3.19) and the second inequality follows from (b) and $(\tilde{x}^{(k)}, \tilde{y}^{(k)}) \in \partial C$. Furthermore, since $\Psi_{\mu_k,\varepsilon_k}(x^{(k)}, y^{(k)}) \le \alpha_k$ from Step 2 of Algorithm 3.1, we have

$$\Psi_{\mu_k,\varepsilon_k}(x^{(k)}, y^{(k)}) \le \frac{1}{4}\overline{m} \tag{3.22}$$

for all $k \in K$ sufficiently large. Now, let $\overline{k} \in K$ be a sufficiently large integer satisfying (a), (3.20), (3.21) and (3.22). Then, applying Lemma 3.4 to $\Psi_{\mu_{\bar{k}},\varepsilon_{\bar{k}}}$ with $p := (\overline{x}, \overline{y})$, $q := (x^{(\bar{k})}, y^{(\bar{k})})$ and $m := \min_{(x,y)\in\partial C} \Psi_{\mu_{\bar{k}},\varepsilon_{\bar{k}}}(x, y) \ge (3/4)\overline{m}$, we obtain the existence of $(\hat{x}^{(k)}, \hat{y}^{(k)}) \in \Re^n \times \Re^n$ such that

$$\nabla\Psi_{\mu_{\bar{k}},\varepsilon_{\bar{k}}}(\hat{x}^{(k)}, \hat{y}^{(k)}) = 0 \quad \text{and} \quad \Psi_{\mu_{\bar{k}},\varepsilon_{\bar{k}}}(\hat{x}^{(k)}, \hat{y}^{(k)}) \ge \frac{3}{4}\overline{m} > 0.$$

However, this contradicts Proposition 3.2. Hence, $\{(x^{(k)}, y^{(k)})\}$ is bounded. ∎

4 Quadratically convergent algorithm

In the previous section, we have presented a hybrid algorithm with smoothing and regularization methods. Moreover, we have shown that the algorithm is globally convergent under the monotonicity assumption. In this section, we consider how it can be equipped with a rapid convergence property. To this end, the properties called *strong semismoothness* and *Jacobian consistency* play an important role. We first show that the functions H_{NR} and $H_{\mu,\varepsilon}$ have those properties. Then, we demonstrate that the hybrid algorithm that incorporates Newton's method to solve subproblems has a quadratic rate of convergence.

4.1 Differential properties of the projection function

To start with, we study differential properties of the function $P_{\mathcal{K}^n}$ defined by (2.3). Specifically, we give an explicit representation of Clarke subdifferential of $P_{\mathcal{K}^n}$. For the sake of convenience, we divide $\Re^n$ into six subsets as follows:

$$\begin{array}{lclclcl} S_1 & := & \{z \,|\, 0 < \lambda_1 \le \lambda_2\} & = & \{z \,|\, z_1 > \|z_2\|\} & = & \operatorname{int}\mathcal{K}^n \\ S_2 & := & \{z \,|\, \lambda_1 < 0 < \lambda_2\} & = & \{z \,|\, -\|z_2\| < z_1 < \|z_2\|\} & = & \Re^n \backslash (\mathcal{K}^n \cup -\mathcal{K}^n) \\ S_3 & := & \{z \,|\, \lambda_1 \le \lambda_2 < 0\} & = & \{z \,|\, z_1 < -\|z_2\|\} & = & -\operatorname{int}\mathcal{K}^n \\ S_4 & := & \{z \,|\, \lambda_1 = 0 < \lambda_2\} & = & \{z \,|\, 0 < z_1 = \|z_2\|\} & = & \operatorname{bd}\mathcal{K}^n \backslash \{0\} \\ S_5 & := & \{z \,|\, \lambda_1 < 0 = \lambda_2\} & = & \{z \,|\, -\|z_2\| = z_1 < 0\} & = & -\operatorname{bd}\mathcal{K}^n \backslash \{0\} \\ S_6 & := & \{z \,|\, \lambda_1 = 0 = \lambda_2\} & = & \{z \,|\, z_1 = \|z_2\| = 0\} & = & \{0\}, \end{array} \tag{4.1}$$

where λ_1 and λ_2 are the spectral values of z given by (2.1), and bd and int denote the boundary and the interior, respectively. Note that $\cup_{i=1}^6 S_i = \Re^n$, $S_i \cap S_j = \emptyset$ $(i \ne j)$, and that S_1, S_2 and S_3 are open sets. Define, moreover, the functions $\Pi_1 : \Re^n \to \Re^n$, $\Pi_2 : D := \{z \in \Re^n | z_2 \ne 0\} \to \Re^n$ and $\Pi_3 : \Re^n \to \Re^n$ as

$$\Pi_1(z) := \lambda_1 u^1 + \lambda_2 u^2 = z \tag{4.2}$$

$$\Pi_2(z) := \lambda_2 u^2 = \frac{1}{2}(z_1 + \|z_2\|)\left(1, \frac{z_2}{\|z_2\|}\right) \tag{4.3}$$

$$\Pi_3(z) := 0. \tag{4.4}$$

Then (2.4) is rewritten as

$$P_{\mathcal{K}^n}(z) = \begin{cases} \Pi_1(z) & (z \in S_1) \\ \Pi_2(z) & (z \in S_2) \\ \Pi_3(z) & (z \in S_3) \\ \Pi_1(z) = \Pi_2(z) & (z \in S_4) \\ \Pi_2(z) = \Pi_3(z) & (z \in S_5) \\ \Pi_1(z) = \Pi_3(z) & (z \in S_6)\,. \end{cases} \tag{4.5}$$

The following proposition refers to the differential properties of Π_1, Π_2 and Π_3.

Proposition 4.1 *Let $z = (z_1, z_2) \in \Re \times \Re^{n-1}$ and $D = \{z \in \Re^n | z_2 \ne 0\}$. Let Π_1, Π_2 and Π_3 be the functions defined by (4.2), (4.3) and (4.4), respectively. Then, Π_1 and Π_3 are differentiable on $\Re^n$, and Π_2 is differentiable on D. Moreover, their Jacobians are given as*

$$\nabla \Pi_1(z) = I \tag{4.6}$$

$$\nabla \Pi_2(z) = \frac{1}{2}\left(1 + \frac{z_1}{\|z_2\|}\right) I + \frac{1}{2\|z_2\|}\begin{pmatrix} -z_1 & z_2^T \\ z_2 & -\dfrac{z_1 z_2 z_2^T}{\|z_2\|^2} \end{pmatrix} \tag{4.7}$$

$$\nabla \Pi_3(z) = O. \tag{4.8}$$

Proof. It is evident from $\Pi_1(z) = z$ and $\Pi_3(z) = 0$ that $\nabla\Pi_1(z) = I$ and $\nabla\Pi_3(z) = O$. By using the fact that

$$\nabla_w\left(\frac{w}{\|w\|}\right) = \frac{1}{\|w\|}\left(I - \frac{ww^T}{\|w\|^2}\right)$$

for any $w \in \Re^{n-1}\backslash\{0\}$, we have

$$\nabla_z\left(1,\ \frac{z_2^T}{\|z_2\|}\right)^T = \frac{1}{\|z_2\|}\begin{pmatrix} 0 & 0 \\ 0 & I - \dfrac{z_2 z_2^T}{\|z_2\|^2}\end{pmatrix}$$

for $z_2 \neq 0$. Moreover, since

$$\nabla_z(z_1 + \|z_2\|) = \left(1,\ \frac{z_2^T}{\|z_2\|}\right)^T,$$

we have

$$\begin{aligned}
\nabla\Pi_2(z) &= \frac{1}{2}\nabla_z(z_1+\|z_2\|)\left(1,\ \frac{z_2^T}{\|z_2\|}\right) + \frac{1}{2}(z_1+\|z_2\|)\nabla_z\left(1,\ \frac{z_2^T}{\|z_2\|}\right)^T \\
&= \frac{1}{2}\left(1,\ \frac{z_2^T}{\|z_2\|}\right)^T\left(1,\ \frac{z_2^T}{\|z_2\|}\right) + \frac{z_1+\|z_2\|}{2\|z_2\|}\begin{pmatrix} 0 & 0 \\ 0 & I - \dfrac{z_2 z_2^T}{\|z_2\|^2}\end{pmatrix} \\
&= \frac{1}{2}\begin{pmatrix} 1 & \dfrac{z_2^T}{\|z_2\|} \\ \dfrac{z_2}{\|z_2\|} & \left(1+\dfrac{z_1}{\|z_2\|}\right)I - \dfrac{z_1}{\|z_2\|}\dfrac{z_2 z_2^T}{\|z_2\|^2}\end{pmatrix} \\
&= \frac{1}{2}\left(1+\frac{z_1}{\|z_2\|}\right)I + \frac{1}{2\|z_2\|}\begin{pmatrix} -z_1 & z_2^T \\ z_2 & -\dfrac{z_1 z_2 z_2^T}{\|z_2\|^2}\end{pmatrix}.
\end{aligned}$$

■

We have shown that the function $P_{\mathcal{K}^n}$ is composed of the differentiable functions Π_1, Π_2 and Π_3. These functions will be used to represent Clarke subdifferential of $P_{\mathcal{K}^n}$ explicitly.

Let $H : \Re^n \to \Re^m$ be a locally Lipschitzian function. Then H is differentiable almost everywhere by Rademacher's Theorem [6]. Let $D_H \subseteq \Re^n$ be the set of points where H is differentiable. Then the B(ouligand) subdifferential $\partial_B H(x)$ of H at x is defined by

$$\partial_B H(x) := \{\lim_{\substack{\hat{x}\to x \\ \hat{x}\in D_H}} \nabla H(\hat{x})\},$$

and Clarke subdifferential $\partial H(x)$ of H at x is defined by

$$\partial H(x) := \operatorname{co}\partial_B H(x), \tag{4.9}$$

where $\operatorname{co} S$ denotes the convex hull of a set S [6, 18]. We note that $\partial_B H(x) = \partial H(x) = \{\nabla H(x)\}$ if H is continuously differentiable at x. The next proposition gives an explicit representation of Clarke subdifferential of $P_{\mathcal{K}^n}$.

Proposition 4.2 *Clarke subdifferential of $P_{\mathcal{K}^n}$ is represented as*

$$\partial P_{\mathcal{K}^n}(z) \;=\; \begin{cases} \{I\} & (z \in S_1) \\ \left\{\dfrac{1}{2}(1+r_1)\,I + \dfrac{1}{2}\begin{pmatrix} -r_1 & r_2^T \\ r_2 & -r_1 r_2 r_2^T \end{pmatrix}\right\} & (z \in S_2) \\ \{O\} & (z \in S_3) \\ \mathrm{co}\left\{I,\;\; I + \dfrac{1}{2}\begin{pmatrix} -1 & r_2^T \\ r_2 & -r_2 r_2^T \end{pmatrix}\right\} & (z \in S_4) \\ \mathrm{co}\left\{O,\;\; \dfrac{1}{2}\begin{pmatrix} 1 & r_2^T \\ r_2 & r_2 r_2^T \end{pmatrix}\right\} & (z \in S_5) \\ \mathrm{co}\,(\{O\} \cup \{I\} \cup \mathcal{S}) & (z \in S_6)\,, \end{cases}$$

where $r_1 := z_1/\|z_2\|$, $r_2 := z_2/\|z_2\|$,

$$\mathcal{S} \;:=\; \left\{ \frac{1}{2}(1+\beta)I + \frac{1}{2}\begin{pmatrix} -\beta & w^T \\ w & -\beta w w^T \end{pmatrix} \middle|\, -1 \le \beta \le 1, \|w\| = 1 \right\},$$

and S_i $(i = 1, \ldots, 6)$ *are defined by* (4.1).

Proof. It follows from (4.2)$-$(4.5) and Proposition 4.1 that $P_{\mathcal{K}^n}$ is continuously differentiable at any point z in the open sets S_i $(i = 1, 2, 3)$. So it suffices to consider the case where $z \in S_i$ $(i = 4, 5, 6)$.

When $z \in S_4$, from (4.1), we have $z_1 = \|z_2\|$, namely $r_1 = 1$. Note that S_4 is adjacent to S_1 and S_2 but not to S_3. Since

$$\lim_{\substack{\hat{z} \to z \\ \hat{z} \in S_1}} \nabla P_{\mathcal{K}^n}(\hat{z}) \;=\; \lim_{\substack{\hat{z} \to z \\ \hat{z} \in S_1}} \nabla \Pi_1(\hat{z}) \;=\; I \tag{4.10}$$

and

$$\lim_{\substack{\hat{z} \to z \\ \hat{z} \in S_2}} \nabla P_{\mathcal{K}^n}(\hat{z}) \;=\; \lim_{\substack{\hat{z} \to z \\ \hat{z} \in S_2}} \nabla \Pi_2(\hat{z}) \;=\; I + \frac{1}{2}\begin{pmatrix} -1 & r_2^T \\ r_2 & -r_2 r_2^T \end{pmatrix},$$

Clarke subdifferential of $P_{\mathcal{K}^n}$ at $z \in S_4$ is given by

$$\partial P_{\mathcal{K}^n}(z) \;=\; \mathrm{co}\left\{I,\;\; I + \frac{1}{2}\begin{pmatrix} -1 & r_2^T \\ r_2 & -r_2 r_2^T \end{pmatrix}\right\}.$$

When $z \in S_5$, from (4.1), we have $z_1 = -\|z_2\|$, namely $r_1 = -1$. Note that S_5 is adjacent to S_2 and S_3 but not to S_1. Since

$$\lim_{\substack{\hat{z} \to z \\ \hat{z} \in S_3}} \nabla P_{\mathcal{K}^n}(\hat{z}) \;=\; \lim_{\substack{\hat{z} \to z \\ \hat{z} \in S_3}} \nabla \Pi_3(\hat{z}) \;=\; O \tag{4.11}$$

and

$$\lim_{\substack{\hat{z} \to z \\ \hat{z} \in S_2}} \nabla P_{\mathcal{K}^n}(\hat{z}) \;=\; \lim_{\hat{z} \to z} \Pi_2(\hat{z}) \;=\; \frac{1}{2}\begin{pmatrix} 1 & r_2^T \\ r_2 & r_2 r_2^T \end{pmatrix},$$

Clarke subdifferential of $P_{\mathcal{K}^n}$ at $z \in S_5$ is given by

$$\partial P_{\mathcal{K}^n}(z) \;=\; \mathrm{co}\left\{O,\;\; \frac{1}{2}\begin{pmatrix} 1 & r_2^T \\ r_2 & r_2 r_2^T \end{pmatrix}\right\}.$$

Let $z \in S_6$, i.e., $z = 0$. Note that $z = 0$ is adjacent to S_1, S_2 and S_3. Similar to (4.10) and (4.11), we have

$$\lim_{\substack{\hat{z}\to 0\\ \hat{z}\in S_1}} \nabla P_{\mathcal{K}^n}(\hat{z}) \;=\; I \quad \text{and} \quad \lim_{\substack{\hat{z}\to 0\\ \hat{z}\in S_3}} \nabla P_{\mathcal{K}^n}(\hat{z}) \;=\; O.$$

We will show $\mathcal{T} = \mathcal{S}$, where

$$\begin{aligned}
\mathcal{T} \;&:=\; \left\{ \lim_{\substack{\hat{z}\to 0\\ \hat{z}\in S_2}} \nabla \Pi_2(\hat{z}) \right\}, \\
\mathcal{S} \;&:=\; \left\{ \frac{1}{2}(1+\beta)I + \frac{1}{2}\begin{pmatrix} -\beta & w^T \\ w & -\beta w w^T \end{pmatrix} \middle| -1 \le \beta \le 1, \|w\| = 1 \right\}.
\end{aligned}$$

First we show $\mathcal{T} \subseteq \mathcal{S}$. Let $\{z^{(k)}\} \subseteq \Re^n$ be an arbitrary sequence such that $z^{(k)} \to 0$ and $z^{(k)} \in S_2$ for all k, and denote $(r_1^{(k)}, r_2^{(k)}) := (z_1^{(k)}, z_2^{(k)})/\|z_2^{(k)}\|$. Then we have

$$\nabla \Pi_2(z^{(k)}) \;=\; \frac{1}{2}\left(1 + r_1^{(k)}\right) I + \frac{1}{2}\begin{pmatrix} -r_1^{(k)} & (r_2^{(k)})^T \\ r_2^{(k)} & -r_1^{(k)} r_2^{(k)} (r_2^{(k)})^T \end{pmatrix}.$$

Since $z^{(k)} \in S_2$ implies $-1 < r_1^{(k)} < 1$, every accumulation point of $\{r_1^{(k)}\}$ belongs to the interval $[-1, 1]$. Moreover, every accumulation point of $\{r_2^{(k)}\}$ is a unit vector since $\|r_2^{(k)}\| = 1$ for all k. Hence, we have $\mathcal{T} \subseteq \mathcal{S}$. Next we show $\mathcal{T} \supseteq \mathcal{S}$. Let J be an arbitrary element of $\mathcal{S}$. Then, there exist a scalar $\beta \in [-1, 1]$ and a unit vector w such that

$$J \;=\; \frac{1}{2}(1+\beta)I + \frac{1}{2}\begin{pmatrix} -\beta & w^T \\ w & -\beta w w^T \end{pmatrix}.$$

Let $\{\xi^{(k)}\} = \{(\xi_1^{(k)}, \xi_2^{(k)})\} \subseteq \Re \times \Re^{n-1}$ be defined by

$$\begin{aligned}
\xi_1^{(k)} \;&:=\; \frac{\beta}{k+1}\left(1 - \frac{1}{k+1}\right) \\
\xi_2^{(k)} \;&:=\; \frac{w}{k+1},
\end{aligned}$$

and $\{\eta^{(k)}\} = \{(\eta_1^{(k)}, \eta_2^{(k)})\} \subseteq \Re \times \Re^{n-1}$ be defined by $(\eta_1^{(k)}, \eta_2^{(k)}) := (\xi_1^{(k)}, \xi_2^{(k)})/\|\xi_2^{(k)}\|$. Then, $\xi^{(k)} \in S_2$ for all k since $\xi_1^{(k)}/\|\xi_2^{(k)}\| = \beta(1 - 1/(k+1)) \in (-1, 1)$. Moreover, we have $\lim_{k\to\infty} \xi^{(k)} = 0$, $\lim_{k\to\infty} \eta_1^{(k)} = \beta$ and $\lim_{k\to\infty} \eta_2^{(k)} = w$. Hence, we have

$$\begin{aligned}
\lim_{k\to\infty} \nabla \Pi_2(\xi^{(k)}) \;&=\; \lim_{k\to\infty} \left\{ \frac{1}{2}\left(1 + \eta_1^{(k)}\right) I + \frac{1}{2}\begin{pmatrix} -\eta_1^{(k)} & (\eta_2^{(k)})^T \\ \eta_2^{(k)} & -\eta_1^{(k)} \eta_2^{(k)} (\eta_2^{(k)})^T \end{pmatrix} \right\} \\
&=\; \frac{1}{2}(1+\beta)I + \frac{1}{2}\begin{pmatrix} -\beta & w^T \\ w & -\beta w w^T \end{pmatrix} \\
&=\; J.
\end{aligned}$$

This implies $J \in \mathcal{T}$, that is, $\mathcal{T} \supseteq \mathcal{S}$. Thus, Clarke subdifferential of $P_{\mathcal{K}^n}$ at $z = 0$ is given by

$$\partial P_{\mathcal{K}^n}(z) \;=\; \mathrm{co}\left(\{O\} \cup \{I\} \cup \mathcal{S}\right).$$

This completes the proof. ■

4.2 Semismoothness and strong semismoothness

Semismoothness is a generalized concept of the smoothness, which was originally introduced by Mifflin [14] for functionals and extended to vector-valued functions by Qi and Sun [20]. Strong semismoothness is also a generalized concept of the smoothness, which is stronger than the semismoothness. These concepts play an important role in establishing fast local convergence of the algorithm presented later. In this section, we first show that the function $P_{\mathcal{K}^n}$ is strongly semismooth at an arbitrary point. Then, by using this result, we show the strong semismoothness of the function H_{NR} defined by (3.2).

Definition 4.1 *A directionally differentiable and locally Lipschitzian function $H : \Re^n \to \Re^m$ is said to be*

(a) *semismooth at z if*

$$V^T d - H'(z\,; d) = o(\|d\|) \tag{4.12}$$

for any $d \in \Re^n \backslash \{0\}$ sufficiently small and $V \in \partial H(z+d)$;

(b) *strongly semismooth at z if*

$$V^T d - H'(z\,; d) = O(\|d\|^2) \tag{4.13}$$

for any $d \in \Re^n \backslash \{0\}$ sufficiently small and $V \in \partial H(z+d)$,

where

$$H'(z; d) := \lim_{h \downarrow 0} \frac{H(z+hd) - H(z)}{h}$$

is the directional derivative of H at z along the direction d.

To show the strong semismoothness of $P_{\mathcal{K}^n}$, it will be helpful to study the strong semismoothness of the three functions Π_1, Π_2 and Π_3 that constitute $P_{\mathcal{K}^n}$, cf. (4.5).

Lemma 4.1 *Let Π_1, Π_2 and Π_3 be the functions defined by* (4.2), (4.3) *and* (4.4), *respectively. Then, Π_1 and Π_3 are strongly semismooth at any $z \in \Re^n$, and Π_2 is strongly semismooth at any $z \in \Re^n$ such that $z_2 \neq 0$.*

Proof. Obviously, Π_1 and Π_3 are strongly semismooth since their Jacobians are constants. So we only consider the function Π_2. Let $z = (z_1, z_2) \in \Re \times \Re^{n-1}$ be an arbitrary vector such that $z_2 \neq 0$, and $d = (d_1, d_2) \in \Re \times \Re^{n-1}$ be an arbitrarily small and nonzero vector such that $z_2 + d_2 \neq 0$. Moreover, let V be an arbitrary element of $\partial \Pi_2(z+d)$. Note that Π_2 is twice differentiable at z_2 and $z_2 + d_2$ since $z_2 \neq 0$ and $z_2 + d_2 \neq 0$. Then we have

$$\begin{aligned} \|V^T d - \Pi_2'(z\,; d)\| &= \|\nabla \Pi_2(z+d)^T d - \nabla \Pi_2(z)^T d\| \\ &\leq \|\nabla \Pi_2(z+d) - \nabla \Pi_2(z)\| \|d\| \\ &= O(\|d\|^2), \end{aligned}$$

which proves the strong semismoothness of Π_2. ∎

By using this lemma, we show the strong semismoothness of $P_{\mathcal{K}^n}$.

Theorem 4.1 *The function $P_{\mathcal{K}^n}$ defined by* (2.3) *is strongly semismooth at any point $z \in \Re^n$.*

Proof. Let S_i $(i = 1, \ldots, 6)$ be defined by (4.1). Let Π_1, Π_2 and Π_3 be defined by (4.2), (4.3) and (4.4), respectively. When $z \in S_1$, for any $d \in \Re^n$ sufficiently small, $z + d$ is an element of S_1, because S_1 is an open set. Hence, $P_{\mathcal{K}^n}$ is strongly semismooth at any $z \in S_1$ by (4.5) and Lemma 4.1. In a similar way, we can show the strong semismoothness of $P_{\mathcal{K}^n}$ at any $z \in S_2 \cup S_3$.

When $z \in S_4$, for any $d \in \Re^n$ sufficiently small, we have $z+d \in S_1 \cup S_2 \cup S_4$. Let V be an arbitrary element of $\partial P_{\mathcal{K}^n}(z + d)$. If $z + d \in S_1$, it then follows from Lemma 4.1 that $V^T d - P'_{\mathcal{K}^n}(z\,; d) = (\nabla\Pi_1(z + d) - \nabla\Pi_1(z))^T d = O(\|d\|^2)$. Similarly, if $z + d \in S_2$, it follows that $V^T d - P'_{\mathcal{K}^n}(z\,; d) = (\nabla\Pi_2(z + d) - \nabla\Pi_2(z))^T d = O(\|d\|^2)$. If $z + d \in S_4$, then we can find a scalar $\alpha \in [0, 1]$ such that $V = (1 - \alpha)\nabla\Pi_1(z + d) + \alpha\nabla\Pi_2(z + d)$. Since $P'_{\mathcal{K}^n}(z\,; d) = \nabla\Pi_1(z)^T d = \nabla\Pi_2(z)^T d$, we have $V^T d - P'_{\mathcal{K}^n}(z\,; d) = (1 - \alpha)\{\nabla\Pi_1(z + d)^T d - \nabla\Pi_1(z)^T d\} + \alpha\{\nabla\Pi_2(z + d)^T d - \nabla\Pi_2(z)^T d\} = O(\|d\|^2)$. Thus, $P_{\mathcal{K}^n}$ is strongly semismooth at $z \in S_4$. In a similar way, we can show the strong semismoothness of $P_{\mathcal{K}^n}$ at $z \in S_5$.

Finally, we consider the case where $z \in S_6$, that is, $z = 0$. From (4.2), (4.3), (4.4) and (4.5), $P_{\mathcal{K}^n}$ is directionally differentiable at 0, and its directional derivative along the direction $d = (d_1, d_2) \in \Re \times \Re^{n-1} \backslash \{(0, 0)\}$ is given by

$$\begin{aligned} P'_{\mathcal{K}^n}(0\,; d) &= \lim_{h \downarrow 0} P_{\mathcal{K}^n}(hd)/h \\ &= P_{\mathcal{K}^n}(d). \end{aligned} \tag{4.14}$$

When $d \in S_1$, it follows from (4.5), (4.6) and (4.14) that $V^T d - P'_{\mathcal{K}^n}(0\,; d) = \nabla\Pi_1(d)^T d - \Pi_1(d) = Id - d = 0$. When $d \in S_2$, it follows from (4.5), (4.7) and (4.14) that

$$\begin{aligned} & V^T d - P'_{\mathcal{K}^n}(0\,; d) \\ &= \nabla\Pi_2(d)^T d - \Pi_2(d) \\ &= \frac{1}{2}\begin{pmatrix} 1 & \dfrac{d_2^T}{\|d_2\|} \\ \dfrac{d_2}{\|d_2\|} & \left(1 + \dfrac{d_1}{\|d_2\|}\right) I - \dfrac{d_1}{\|d_2\|}\dfrac{d_2 d_2^T}{\|d_2\|^2} \end{pmatrix}\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} - \frac{1}{2}(d_1 + \|d_2\|)\begin{pmatrix} 1 \\ \dfrac{d_2}{\|d_2\|} \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} d_1 + \|d_2\| \\ \dfrac{d_1 d_2}{\|d_2\|} + d_2 \end{pmatrix} - \frac{1}{2}(d_1 + \|d_2\|)\begin{pmatrix} 1 \\ \dfrac{d_2}{\|d_2\|} \end{pmatrix} \\ &= 0. \end{aligned}$$

When $d \in S_3$, it follows from (4.5), (4.8) and (4.14) that $V^T d - P'_{\mathcal{K}^n}(0\,; d) = \nabla\Pi_3(d)^T d - \Pi_3(d) = Od - 0 = 0$. When $d \in S_4$, any $V \in \partial P_{\mathcal{K}^n}(d)$ can be written as $V = (1 - \alpha)\nabla\Pi_1(d) + \alpha\nabla\Pi_2(d)$ with $\alpha \in [0, 1]$. Furthermore, we have $P'_{\mathcal{K}^n}(0\,; d) = P_{\mathcal{K}^n}(d) = \Pi_1(d) = \Pi_2(d)$ from (4.14) and (4.5). Hence, we have $V^T d - P_{\mathcal{K}^n}(0\,; d) = (1 - \alpha)\{\nabla\Pi_1(d)^T d - \Pi_1(d)\} + \alpha\{\nabla\Pi_2(d)^T d - \Pi_2(d)\} = 0$. When $d \in S_5$, we can also show $V^T d - P_{\mathcal{K}^n}(0\,; d) = 0$ in a similar way. Consequently, we have $V^T d - P'_{\mathcal{K}^n}(0\,; d) = 0$ for any $d \in \Re^n \backslash \{0\}$. ■

Recently, Chen, Sun and Sun [5] also showed the strong semismoothness of $P_{\mathcal{K}^n}$, and Chen, Chen and Tseng [2] gave more general results. We note, however, that their approaches are quite different from ours. Now, we are ready to prove the strong semismoothness of H_{NR}.

Theorem 4.2 *The function H_{NR} defined by (3.2) is semismooth at every point $(x, y) \in \Re^n \times \Re^n$. Moreover, if $\nabla f : \Re^n \to \Re^{n\times n}$ is locally Lipschitzian, then H_{NR} is strongly semismooth at every point $(x, y) \in \Re^n \times \Re^n$.*

Proof. Let $(\xi, \eta) \in \Re^n \times \Re^n$ be a pair of vectors sufficiently small, and U be an arbitrary element of $\partial H_{\mathrm{NR}}(x+\xi, y+\eta)$. Note that, from Theorem 4.1, $P_{\mathcal{K}^n}$ is strongly semismooth at $x - y$. Then, since

$$\partial H_{\mathrm{NR}}(x + \xi, y + \eta) = \left\{ \begin{pmatrix} I - V & \nabla f(x + \xi) \\ V & -I \end{pmatrix} \,\middle|\, V \in \partial P_{\mathcal{K}^n}((x - y) + (\xi - \eta)) \right\}$$

and

$$H'_{\text{NR}}((x,y)\,;(\xi,\eta)) \;=\; \begin{pmatrix} \xi - P'_{\mathcal{K}^n}(x-y\,;\xi-\eta) \\ \nabla f(x)^T\xi - \eta \end{pmatrix},$$

we have

$$\begin{aligned} U^T\begin{pmatrix}\xi\\ \eta\end{pmatrix} - H'_{\text{NR}}((x,y)\,;(\xi,\eta)) &= \begin{pmatrix} I-V & \nabla f(x) \\ V & -I \end{pmatrix}^T \begin{pmatrix}\xi\\ \eta\end{pmatrix} - \begin{pmatrix} \xi - P'_{\mathcal{K}^n}(x-y\,;\xi-\eta) \\ \nabla f(x)^T\xi - \eta \end{pmatrix} \\ &= \begin{pmatrix} P'_{\mathcal{K}^n}(x-y\,;\xi-\eta) - V^T(\xi-\eta) \\ (\nabla f(x+\xi) - \nabla f(x))^T\xi \end{pmatrix} \\ &= \begin{pmatrix} O(\|\xi-\eta\|^2) \\ o(\|\xi\|) \end{pmatrix} \qquad (4.15) \\ &= o(\|(\xi,\eta)\|), \end{aligned}$$

where the last equality follows from $\|\xi-\eta\|^2 \le 2\|\xi\|^2 + 2\|\eta\|^2 = 2\|(\xi,\eta)\|^2$ and $\|\xi\| \le \|(\xi,\eta)\|$. Hence, H_{NR} is semismooth at (x,y). Furthermore, if ∇f is locally Lipschitzian, then the strong semismoothness of H_{NR} at (x,y) readily follows since $o(\|\xi\|)$ in (4.15) can be replaced by $O(\|\xi\|^2)$. ∎

4.3 Jacobian consistency of the smoothing function

Jacobian consistency, which was first introduced by Chen, Qi and Sun [4], is a concept relating the generalized Jacobian of a nonsmooth function with Jacobian of a smoothing function. Like the semismoothness and the strong semismoothness, the Jacobian consistency often plays an important role in establishing rapid convergence of smoothing methods. In this subsection, we show that the function $H_{\mu,\varepsilon}$ defined by (3.10) enjoys the Jacobian consistency, and give some results that will be useful in constructing a rapidly convergent algorithm.

Definition 4.2 *Let $F : \Re^n \to \Re^n$ be a continuous function. Let $F_{\mu,\varepsilon} : \Re^n \to \Re^n$ be a function such that $F_{\mu,\varepsilon}$ is continuously differentiable for any $\mu > 0$ and $\varepsilon \ge 0$, and that $\lim_{(\mu,\varepsilon)\downarrow(0,0)} F_{\mu,\varepsilon}(z) = F(z)$ for any $z \in \Re^n$. Then we say $F_{\mu,\varepsilon}$ satisfies the Jacobian consistency if*

$$\lim_{(\mu,\varepsilon)\downarrow(0,0)} \operatorname{dist}(\nabla F_{\mu,\varepsilon}(z),\, \partial F(z)) = 0 \qquad (4.16)$$

holds for any $z \in \Re^n$. Here $\operatorname{dist}(X,S)$ denotes $\min\{\|X-Y\| \mid Y \in S\}$.

We note that Chen, Qi and Sun [4] define the Jacobian consistency in terms of

$$\partial_C F(z) \;:=\; \partial F_1(z) \times \cdots \times \partial F_n(z) \supseteq \partial F(z)$$

instead of $\partial F(z)$. Moreover, their definition contains only a smoothing parameter μ.

In the remainder of the paper, we use the following notation for convenience. Let $\hat{g} : \Re \to \Re$ be a continuously differentiable convex function satisfying (3.4). Then we denote for any $\mu > 0$ that

$$\gamma_\mu(\alpha) := \mu \hat{g}(\alpha/\mu). \qquad (4.17)$$

Note that, as $\mu \downarrow 0$, there exist limits of $\gamma_\mu(\alpha)$ and $\gamma'_\mu(\alpha)$ for any fixed α, which we denote

$$\gamma_0(\alpha) \;:=\; \lim_{\mu\downarrow 0}\gamma_\mu(\alpha) \;=\; \max\{0,\alpha\}, \qquad (4.18)$$

$$\gamma_0^+(\alpha) \;:=\; \lim_{\mu\downarrow 0}\gamma'_\mu(\alpha) \;=\; \begin{cases} 0 & (\alpha < 0) \\ \hat{g}'(0) & (\alpha = 0) \\ 1 & (\alpha > 0). \end{cases} \qquad (4.19)$$

In order to show the Jacobian consistency of $H_{\mu,\varepsilon}$, we first investigate the limiting behavior of $\nabla P_\mu(z)$ as $\mu \downarrow 0$, where P_μ denotes the smoothing function of $P_{\mathcal{K}^n}$ defined by (3.5).

Proposition 4.3 *Fix $z = (z_1, z_2) \in \Re \times \Re^{n-1}$ arbitrarily. Then, as $\mu \downarrow 0$, $\nabla P_\mu(z)$ converges to the matrix $J_P^o(z)$ defined by*

$$J_P^o(z) := \begin{cases} I & (z \in S_1) \\ \frac{1}{2}(1+r_1)I + \frac{1}{2}\begin{pmatrix} -r_1 & r_2^T \\ r_2 & -r_1 r_2 r_2^T \end{pmatrix} & (z \in S_2) \\ O & (z \in S_3) \\ I + \frac{1}{2}(1-\hat{g}'(0))\begin{pmatrix} -1 & r_2^T \\ r_2 & -r_2 r_2^T \end{pmatrix} & (z \in S_4) \\ \frac{1}{2}\hat{g}'(0)\begin{pmatrix} 1 & r_2^T \\ r_2 & r_2 r_2^T \end{pmatrix} & (z \in S_5) \\ \hat{g}'(0)I & (z \in S_6)\,, \end{cases} \tag{4.20}$$

where $r_1 = z_1/\|z_2\| \in \Re$, $r_2 = z_2/\|z_2\| \in \Re^{n-1}$, S_i $(i = 1, \ldots, 6)$ are the sets defined by (4.1), *and $\hat{g} : \Re \to \Re$ is a function satisfying* (3.4).

Proof. Let λ_1 and λ_2 be the spectral values of z given by (2.1). Note that, if $z_2 \neq 0$, $\nabla P_\mu(z)$ is given by (3.14) with a_μ, b_μ and c_μ defined by (3.15). First we suppose $z_2 \neq 0$ and consider the limiting behavior of a_μ, b_μ and c_μ as $\mu \downarrow 0$. Since we have

$$a_\mu = \frac{\hat{g}(\lambda_2/\mu) - \hat{g}(\lambda_1/\mu)}{\lambda_2/\mu - \lambda_1/\mu} = \frac{\gamma_\mu(\lambda_2) - \gamma_\mu(\lambda_1)}{\lambda_2 - \lambda_1},$$

(4.18) implies that the limit $a_0 := \lim_{\mu\downarrow 0} a_\mu$ is given by

$$a_0 = \frac{\gamma_0(\lambda_2) - \gamma_0(\lambda_1)}{\lambda_2 - \lambda_1} = \begin{cases} 1 & (0 \le \lambda_1 < \lambda_2) \\ (1+r_1)/2 & (\lambda_1 < 0 < \lambda_2) \\ 0 & (\lambda_1 < \lambda_2 \le 0)\,, \end{cases} \tag{4.21}$$

where the second case follows from $\lambda_2/(\lambda_2 - \lambda_1) = (z_1 + \|z_2\|)/(2\|z_2\|) = (1+r_1)/2$. By (3.15) and (4.19), the limit $b_0 := \lim_{\mu\downarrow 0} b_\mu$ is given by

$$b_0 = \frac{1}{2}(\gamma_0^+(\lambda_2) + \gamma_0^+(\lambda_1)) = \begin{cases} 1 & (0 < \lambda_1 < \lambda_2) \\ (1+\hat{g}'(0))/2 & (\lambda_1 = 0 < \lambda_2) \\ 1/2 & (\lambda_1 < 0 < \lambda_2) \\ \hat{g}'(0)/2 & (\lambda_1 < 0 = \lambda_2) \\ 0 & (\lambda_1 < \lambda_2 < 0)\,, \end{cases} \tag{4.22}$$

and $c_0 := \lim_{\mu\downarrow 0} c_\mu$ is given by

$$c_0 = \frac{1}{2}(\gamma_0^+(\lambda_2) - \gamma_0^+(\lambda_1)) = \begin{cases} 0 & (0 < \lambda_1 < \lambda_2) \\ (1-\hat{g}'(0))/2 & (\lambda_1 = 0 < \lambda_2) \\ 1/2 & (\lambda_1 < 0 < \lambda_2) \\ \hat{g}'(0)/2 & (\lambda_1 < 0 = \lambda_2) \\ 0 & (\lambda_1 < \lambda_2 < 0)\,. \end{cases} \tag{4.23}$$

Next we describe the limit $\lim_{\mu\downarrow 0} P_\mu(z)$. We first consider the case where $z \in S_1$, that is, $0 < \lambda_1 \le \lambda_2$. When $z_2 = 0$, we have $0 < \lambda_1 = \lambda_2 = z_1$. Hence, by (3.14) and (4.19), we have

$$\begin{aligned} \lim_{\mu\downarrow 0} \nabla P_\mu(z) &= \lim_{\mu\downarrow 0} \hat{g}'(z_1/\mu)I \\ &= \gamma_0^+(z_1)I \\ &= I. \end{aligned}$$

When $z_2 \neq 0$, that is, $0 < \lambda_1 < \lambda_2$, we have $a_0 = b_0 = 1$ and $c_0 = 0$ by (4.21)–(4.23). This, together with (3.14), yields

$$\begin{aligned} \lim_{\mu \downarrow 0} \nabla P_\mu(z) &= \begin{pmatrix} b_0 & c_0 r_2^T \\ c_0 r_2 & a_0 I + (b_0 - a_0) r_2 r_2^T \end{pmatrix} \\ &= I. \end{aligned}$$

We thus have $J_P^o(z) = I$ for any $z \in S_1$. When $z \in S_3$, that is, $\lambda_1 \le \lambda_2 < 0$, we can show that $J_P^o(z) = O$ in a similar way. When $z \in S_2$, that is, $\lambda_1 < 0 < \lambda_2$ and $z_2 \neq 0$, we have $a_0 = (1 + r_1)/2$ and $b_0 = c_0 = 1/2$ from (4.21)–(4.23). It then follows from (3.14) that

$$J_P^o(z) = \frac{1}{2}(1 + r_1) I + \frac{1}{2} \begin{pmatrix} -r_1 & r_2^T \\ r_2 & -r_1 r_2 r_2^T \end{pmatrix}.$$

We can argue the case of $z \in S_4 \cup S_5$ in a similar way. When $z \in S_6$, we have $\lambda_1 = \lambda_2 = z_1 = 0$ and $z_2 = 0$. Therefore, by (3.14) and (4.19), we have $J_P^o(z) = g'(0) I$. This completes the proof. ■

By using Propositions 4.2 and 4.3, we show the Jacobian consistency of $H_{\mu,\varepsilon}$.

Theorem 4.3 *$H_{\mu,\varepsilon}$ satisfies the Jacobian consistency.*

Proof. Propositions 4.2, 4.3 and $\hat{g}'(0) \in (0,1)$ from (3.4) yield $J_P^o(z) \in \partial P_{\mathcal{K}^n}(z)$ for any $z \in \Re^n$. Moreover, we have

$$J_H^o(x,y) := \lim_{(\mu,\varepsilon)\downarrow(0,0)} \nabla H_{\mu,\varepsilon}(x,y) = \begin{pmatrix} I - J_P^o(x-y) & \nabla f(x) \\ J_P^o(x-y) & -I \end{pmatrix} \tag{4.24}$$

from (3.13), and

$$\partial H_{\rm NR}(x,y) = \left\{ \begin{pmatrix} I - D & \nabla f(x) \\ D & -I \end{pmatrix} \middle| D \in \partial P_{\mathcal{K}^n}(x-y) \right\}.$$

We thus have $J_H^o(x,y) \in \partial H_{\rm NR}(x,y)$ for any $(x,y) \in \Re^n \times \Re^n$, which implies that $H_{\mu,\varepsilon}$ satisfies the Jacobian consistency. ■

The Jacobian consistency of $H_{\mu,\varepsilon}$ guarantees that, for any $(x,y) \in \Re^n \times \Re^n$ and $\delta' > 0$, there exists $\mu > 0$ and $\varepsilon > 0$ such that $\operatorname{dist}(\nabla H_{\mu,\varepsilon}(x,y), \partial H_{\rm NR}(x,y)) < \delta'$. To find such μ and ε for any δ' is an essential issue for constructing a rapidly convergent algorithm. By (3.13), (4.24) and the Jacobian consistency of $H_{\mu,\varepsilon}$, we can easily see that there exists $M > 0$ such that

$$\operatorname{dist}(\nabla H_{\mu,\varepsilon}(x,y), \partial H_{\rm NR}(x,y)) \le M(\|\nabla P_\mu(x-y) - J_P^o(x-y)\| + \varepsilon). \tag{4.25}$$

Next we consider to estimate the value of $\|\nabla P_\mu(z) - J_P^o(z)\|$ by providing its upper bound.

In what follows, we suppose that the function $\hat{g}$ satisfies not only (3.4) but also

$$\hat{g}(\alpha) - \alpha = \hat{g}(-\alpha) \tag{4.26}$$

for any $\alpha \in \Re$. For example, $\hat{g}_1(\alpha) = (\sqrt{\alpha^2 + 4} + \alpha)/2$ and $\hat{g}_2(\alpha) = \ln(e^\alpha + 1)$ satisfy (3.4) and (4.26). For such functions, we have the following lemma.

Lemma 4.2 *Let $\hat{g}$ be a continuously differentiable convex function satisfying* (3.4) *and* (4.26). *Let γ_μ, γ_0 and γ_0^+ be defined by* (4.17)–(4.19). *Then it holds that*

$$\gamma_\mu(\alpha) - \gamma_0(\alpha) = \gamma_\mu(-\alpha) - \gamma_0(-\alpha) \tag{4.27}$$

for any $\alpha \in \Re$, and

$$|\gamma_\mu'(0) - \gamma_0^+(0)| = 0 < |\gamma_\mu'(\alpha_2) - \gamma_0^+(\alpha_2)| \le |\gamma_\mu'(\alpha_1) - \gamma_0^+(\alpha_1)| \tag{4.28}$$

for any $(\alpha_1, \alpha_2) \in \Re \times \Re$ such that $0 < |\alpha_1| \le |\alpha_2|$.

Proof. Since (4.27) can be easily seen from (4.17), (4.18) and (4.26), we only show (4.28). Let $(\alpha_1, \alpha_2) \in \Re \times \Re$ be arbitrary scalars such that $0 < |\alpha_1| \le |\alpha_2|$. From (4.26), we have $\hat{g}'(\alpha) - 1 = -\hat{g}'(-\alpha)$ for any $\alpha \in \Re$. This, together with $\gamma'_\mu(\alpha) = \hat{g}'(\alpha/\mu)$ and (4.19), yields

$$\gamma'_\mu(\alpha) - \gamma_0^+(\alpha) = -\{\gamma'_\mu(-\alpha) - \gamma_0^+(-\alpha)\}.$$

We therefore have $|\gamma'_\mu(0) - \gamma_0^+(0)| = 0$ and

$$\left|\gamma'_\mu(\alpha_i) - \gamma_0^+(\alpha_i)\right| \;=\; \left|\gamma'_\mu(|\alpha_i|) - \gamma_0^+(|\alpha_i|)\right| \tag{4.29}$$

for each $i = 1, 2$. Now, we note that, for any $\beta > 0$,

$$|\gamma'_\mu(\beta) - \gamma_0^+(\beta)| \;=\; 1 - \gamma'_\mu(\beta) > 0,$$

where the inequality follows from $\gamma'_\mu(\beta) = \hat{g}'(\beta/\mu)$ and (3.4). Moreover, $1 - \gamma'_\mu(\beta)$ is monotonically nonincreasing since γ_μ is convex. Hence, we have

$$0 \;<\; \left|\gamma'_\mu(|\alpha_2|) - \gamma_0^+(|\alpha_2|)\right| \;\le\; \left|\gamma'_\mu(|\alpha_1|) - \gamma_0^+(|\alpha_1|)\right|.$$

Combining this inequality with (4.29), we have (4.28). ∎

Now, we define the function $\tilde{\lambda} : \Re^n \to [0, +\infty)$ by

$$\tilde{\lambda}(z) \;:=\; \begin{cases} \min\limits_{i \in \mathcal{I}(z)} |\lambda_i(z)| & (\mathcal{I}(z) \neq \emptyset) \\ 0 & (\mathcal{I}(z) = \emptyset), \end{cases} \tag{4.30}$$

where $\lambda_i(z)$ $(i = 1, 2)$ are the spectral values of z, and $\mathcal{I}(z) \subseteq \{1, 2\}$ is the index set defined by $\mathcal{I}(z) := \{i \,|\, \lambda_i(z) \neq 0\}$. Then we can easily see from (4.28) that

$$|\gamma'_\mu(\lambda_i(z)) - \gamma_0^+(\lambda_i(z))| \;\le\; |\gamma'_\mu(\tilde{\lambda}(z)) - \gamma_0^+(\tilde{\lambda}(z))| \qquad (i = 1, 2). \tag{4.31}$$

From this fact, we have the following proposition.

Proposition 4.4 *Let $\hat{g}$ be a function satisfying* (3.4) *and* (4.26)*, and $\tilde{\lambda}$ be the function defined by* (4.30)*. Then, there exists $M' > 0$ such that*

$$\|\nabla P_\mu(z) - J_P^o(z)\| \;\le\; M'|\gamma'_\mu(\tilde{\lambda}(z)) - \gamma_0^+(\tilde{\lambda}(z))|$$

for any $\mu > 0$ and $z \in \Re^n$.

Proof. Let λ_1 and λ_2 be the spectral values of z. When $z_2 = 0$, that is, $z_1 = \lambda_1 = \lambda_2$, we have

$$\begin{aligned} \|\nabla P_\mu(z) - J_P^o(z)\| &= \|g'(z_1/\mu)I - \gamma_0^+(z_1)I\| \\ &= |\gamma'_\mu(z_1) - \gamma_0^+(z_1)| \\ &= |\gamma'_\mu(\tilde{\lambda}(z)) - \gamma_0^+(\tilde{\lambda}(z))|, \end{aligned}$$

where the first equality follows from (3.14), (4.20) and (4.19), and the second equality follows from $\tilde{\lambda}(z) = |z_1|$.

When $z_2 \neq 0$, that is, $\lambda_1 < \lambda_2$, by (3.14) and (4.20)–(4.23), we have for some $M'' > 0$

$$\begin{aligned} &\|\nabla P_\mu(z) - J_P^o(z)\| \\ &= \left\|(a_\mu - a_0)I + \begin{pmatrix} (b_\mu - b_0) - (a_\mu - a_0) & (c_\mu - c_0)r_2^T \\ (c_\mu - c_0)r_2 & ((b_\mu - b_0) - (a_\mu - a_0))r_2 r_2^T \end{pmatrix}\right\| \\ &\le M''(|a_\mu - a_0| + |b_\mu - b_0| + |c_\mu - c_0|), \end{aligned} \tag{4.32}$$

where $r_2 = z_2/\|z_2\|$. First we consider the case where $0 \le \lambda_1 < \lambda_2$. Note that there exists $\overline{\lambda} \in [\lambda_1, \lambda_2]$ such that

$$\gamma'_\mu(\overline{\lambda}) \;=\; \frac{\gamma_\mu(\lambda_2) - \gamma_\mu(\lambda_1)}{\lambda_2 - \lambda_1} \;=\; a_\mu.$$

Moreover, we have from (4.19), (4.21) and $\overline{\lambda} \in [\lambda_1, \lambda_2]$ that

$$a_0 \;=\; 1 \;=\; \gamma_0^+(\overline{\lambda}).$$

Hence, we obtain

$$|a_\mu - a_0| \;=\; |\gamma'_\mu(\overline{\lambda}) - \gamma_0^+(\overline{\lambda})|.$$

Furthermore, by (3.15) and (4.22), we have

$$\begin{aligned} |b_\mu - b_0| &= \frac{1}{2}\left|\Big(\gamma'_\mu(\lambda_2) + \gamma'_\mu(\lambda_1)\Big) - \Big(\gamma_0^+(\lambda_2) + \gamma_0^+(\lambda_1)\Big)\right| \\ &\le \frac{1}{2}|\gamma'_\mu(\lambda_2) - \gamma_0^+(\lambda_2)| + \frac{1}{2}|\gamma'_\mu(\lambda_1) - \gamma_0^+(\lambda_1)|, \end{aligned}$$

and, by (3.15) and (4.23), we have

$$\begin{aligned} |c_\mu - c_0| &= \frac{1}{2}\left|\Big(\gamma'_\mu(\lambda_2) - \gamma'_\mu(\lambda_1)\Big) - \Big(\gamma_0^+(\lambda_2) - \gamma_0^+(\lambda_1)\Big)\right| \\ &\le \frac{1}{2}|\gamma'_\mu(\lambda_2) - \gamma_0^+(\lambda_2)| + \frac{1}{2}|\gamma'_\mu(\lambda_1) - \gamma_0^+(\lambda_1)|. \end{aligned}$$

Combining these inequalities, we have

$$\begin{aligned} &|a_\mu - a_0| + |b_\mu - b_0| + |c_\mu - c_0| \\ &\le\; |\gamma'_\mu(\lambda_2) - \gamma_0^+(\lambda_2)| + |\gamma'_\mu(\lambda_1) - \gamma_0^+(\lambda_1)| + |\gamma'_\mu(\overline{\lambda}) - \gamma_0^+(\overline{\lambda})| \\ &\le\; 3|\gamma'_\mu(\tilde{\lambda}(z)) - \gamma_0^+(\tilde{\lambda}(z))|, \end{aligned}$$

where the last inequality follows from $0 \le \lambda_1 < \lambda_2$, $\overline{\lambda} \in [\lambda_1, \lambda_2]$, (4.28) and (4.31). Hence, by (4.32), we have for some $M' > 0$ that $\|\nabla P_\mu(z) - J_P^o(z)\| \;\le\; M'|\gamma'_\mu(\tilde{\lambda}(z)) - \gamma_0^+(\tilde{\lambda}(z))|$. When $\lambda_1 < \lambda_2 \le 0$, we can obtain the desired result in a similar way.

Finally, we consider the case where $\lambda_1 < 0 < \lambda_2$. First we assume $0 < |\lambda_1| \le \lambda_2$. When $|\lambda_1| < \lambda_2$, there exists $\overline{\lambda} \in [|\lambda_1|, \lambda_2] = [-\lambda_1, \lambda_2]$ such that

$$\begin{aligned} |a_\mu - a_0| &= \left|\frac{\gamma_\mu(\lambda_2) - \gamma_\mu(\lambda_1)}{\lambda_2 - \lambda_1} - \frac{\gamma_0(\lambda_2) - \gamma_0(\lambda_1)}{\lambda_2 - \lambda_1}\right| \\ &= \left|\frac{\gamma_\mu(\lambda_2) - \gamma_\mu(-\lambda_1)}{\lambda_2 - \lambda_1} - \frac{\gamma_0(\lambda_2) - \gamma_0(-\lambda_1)}{\lambda_2 - \lambda_1}\right| \\ &= \frac{\lambda_2 + \lambda_1}{\lambda_2 - \lambda_1}\left|\frac{\gamma_\mu(\lambda_2) - \gamma_\mu(-\lambda_1)}{\lambda_2 - (-\lambda_1)} - 1\right| \\ &= \frac{\lambda_2 + \lambda_1}{\lambda_2 - \lambda_1}|\gamma'_\mu(\overline{\lambda}) - \gamma_0^+(\overline{\lambda})| \\ &\le |\gamma'_\mu(\overline{\lambda}) - \gamma_0^+(\overline{\lambda})|, \end{aligned} \tag{4.33}$$

where the second equality follows from (4.27), the third equality follows from (4.18), the fourth equality follows from (4.19), and the inequality follows from $(\lambda_2+\lambda_1)/(\lambda_2-\lambda_1) = (|\lambda_2|-|\lambda_1|)/(|\lambda_2|+|\lambda_1|) \le 1$. When $|\lambda_1| = \lambda_2$, we have $|a_\mu - a_0| = 0$ from (4.33) and (4.27). In a way similar to the previous case, we have $|b_\mu - b_0| \le (1/2)|\gamma'_\mu(\lambda_2) - \gamma_0^+(\lambda_2)| + (1/2)|\gamma'_\mu(\lambda_1) - \gamma_0^+(\lambda_1)|$ and $|c_\mu - c_0| \le (1/2)|\gamma'_\mu(\lambda_2) -$

$\gamma_0^+(\lambda_2)|\ +(1/2)|\gamma_\mu'(\lambda_1) - \gamma_0^+(\lambda_1)|$. Hence, by $0 < |\lambda_1| \le \lambda_2$, $\overline{\lambda} \in [|\lambda_1|, \lambda_2]$, (4.28) and (4.31), we have $\|\nabla P_\mu(z) - J_P^o(z)\| \le M'|\gamma_\mu'(\tilde{\lambda}(z)) - \gamma_0^+(\tilde{\lambda}(z))|$ for some $M' > 0$. In the case where $\lambda_1 < 0 < \lambda_2$ and $0 < \lambda_2 \le |\lambda_1|$, we can also show the desired result in a similar manner. ■

Proposition 4.4 indicates that, if we can minify the value of $|\gamma_\mu'(\alpha) - \gamma_0^+(\alpha)|$ arbitrarily by choosing a sufficiently small $\mu > 0$, then we can also minify the value of $\|\nabla P_\mu(z) - J_P^o(z)\|$ arbitrarily. Since $\gamma_0^+(\alpha) = \lim_{\mu\downarrow 0} \gamma_\mu'(\alpha)$, for any $\delta > 0$ and $\alpha \in \Re$, there exists $\overline{\mu}(\alpha, \delta)$ such that

$$|\gamma_\mu'(\alpha) - \gamma_0^+(\alpha)| < \delta$$

holds for any $\mu \in (0, \overline{\mu}(\alpha, \delta))$. The following proposition gives an explicit expression of such $\overline{\mu}(\alpha, \delta)$ when $\hat{g}(\alpha)$ is given by $\hat{g}(\alpha) = (\alpha + \sqrt{\alpha^2 + 4})/2$.

Proposition 4.5 *Let the function $\hat{g}$ be defined by $\hat{g}(\alpha) = (\alpha + \sqrt{\alpha^2 + 4})/2$, and γ_μ and γ_0^+ be defined by* (4.17) and (4.19), *respectively. Moreover, let $\overline{\mu}(\alpha, \delta)$ be defined by*

$$\overline{\mu}(\alpha, \delta) := \begin{cases} +\infty & (\delta \ge 1/2 \text{ or } \alpha = 0) \\ \dfrac{1}{2}|\alpha|\sqrt{\delta} & (\delta < 1/2 \text{ and } \alpha \neq 0)\,. \end{cases}$$

Then, for any $\alpha \in \Re$, $\delta > 0$ and $\mu \in (0, \overline{\mu}(\alpha, \delta))$, we have

$$|\gamma_\mu'(\alpha) - \gamma_0^+(\alpha)| < \delta. \tag{4.34}$$

Proof. Let μ be an arbitrary scalar in the interval $(0, \overline{\mu}(\alpha, \delta))$. By easy calculation, we have

$$\begin{aligned} \gamma_\mu(\alpha) &= \frac{1}{2}\left(\alpha + \sqrt{\alpha^2 + 4\mu^2}\right), \\ \gamma_\mu'(\alpha) &= \frac{1}{2}\left(1 + \frac{\alpha}{\sqrt{\alpha^2 + 4\mu^2}}\right), \\ \gamma_0^+(\alpha) &= \begin{cases} 1 & (\alpha > 0) \\ 1/2 & (\alpha = 0) \\ 0 & (\alpha < 0)\,. \end{cases} \end{aligned}$$

Moreover, these equalities yield

$$|\gamma_\mu'(\alpha) - \gamma_0^+(\alpha)| = \begin{cases} 0 & (\alpha = 0) \\ \dfrac{1}{2}\left(1 - \dfrac{|\alpha|}{\sqrt{\alpha^2 + 4\mu^2}}\right) & (\alpha \neq 0)\,. \end{cases}$$

When $\alpha = 0$, (4.34) trivially holds since $|\gamma_\mu'(\alpha) - \gamma_0^+(\alpha)| = 0 < \delta$. When $\delta \ge 1/2$, (4.34) always holds since $|\gamma_\mu'(\alpha) - \gamma_0^+(\alpha)| < 1/2$ for any $\delta > 0$ and $\mu > 0$. When $\alpha \neq 0$ and $0 < \delta < 1/2$, we have

$$\begin{aligned} |\gamma_\mu'(\alpha) - \gamma_0^+(\alpha)| - \delta &= \frac{1}{2}\left(1 - \frac{|\alpha|}{\sqrt{\alpha^2 + 4\mu^2}}\right) - \delta \\ &< \frac{1}{2}\left(1 - \frac{|\alpha|}{\sqrt{\alpha^2 + \alpha^2\delta}}\right) - \delta \\ &= \frac{1}{2}\left(1 - \frac{1}{\sqrt{1 + \delta}}\right) - \delta \\ &< 0, \end{aligned}$$

where the first inequality follows from $\mu < \overline{\mu}(\alpha, \delta) = (1/2)|\alpha|\sqrt{\delta}$, and the last inequality follows since the function $f(\delta) := (1/2)(1 - (1 + \delta)^{-1/2}) - \delta$ satisfies $f(0) = 0$ and $f'(\delta) = (1/4)(1 + \delta)^{-3/2} - 1 < 0$ for any $\delta > 0$. This completes the proof. ■

4.4 Quadratically convergent algorithm with Newton's method

In this subsection, based on Algorithm 3.1, we propose a more specific algorithm in which Newton's method is applied for solving subproblem (3.12). Moreover, we show that the algorithm is quadratically convergent under appropriate assumptions including the strong semismoothness and the Jacobian consistency. We first state the algorithm. For convenience, we denote

$$w := \begin{pmatrix} x \\ y \end{pmatrix}, \quad w^{(k)} := \begin{pmatrix} x^{(k)} \\ y^{(k)} \end{pmatrix}.$$

Algorithm 4.1 *Choose* η, $\rho \in (0,1)$, $\overline{\eta} \in (0,\eta]$, $\sigma \in (0,1/2)$, $\kappa > 0$, $\overline{\kappa} > 0$ *and* $\hat{\kappa} > 0$.

(Step 0) *Choose* $w^{(0)} \in \Re^{2n}$ *and* $\beta_0 \in (0,\infty)$. *Let* $\mu_0 := \|H_{\rm NR}(w^{(0)})\|$ *and* $\varepsilon_0 := \|H_{\rm NR}(w^{(0)})\|$. *Set* $k := 0$.

(Step 1) *Terminate if* $\|H_{\rm NR}(w^{(k)})\| = 0$.

(Step 2)

(Step 2.0) *Set* $v^{(0)} := w^{(k)}$ *and* $j := 0$.

(Step 2.1) *Find a vector* $\hat{d}^{(j)}$ *such that*

$$H_{\mu_k,\varepsilon_k}(v^{(j)}) + \nabla H_{\mu_k,\varepsilon_k}(v^{(j)})^T \hat{d}^{(j)} = 0. \tag{4.35}$$

(Step 2.2) *If* $\|H_{\mu_k,\varepsilon_k}(v^{(j)} + \hat{d}^{(j)})\| \le \beta_k$, *then let* $w^{(k+1)} := v^{(j)} + \hat{d}^{(j)}$ *and go to* Step 3. *Otherwise, go to* Step 2.3.

(Step 2.3) *Find the smallest nonnegative integer* m *such that*

$$\Psi_{\mu_k,\,\varepsilon_k}(v^{(j)} + \rho^m \hat{d}^{(j)}) \le (1 - 2\sigma\rho^m)\Psi_{\mu_k,\,\varepsilon_k}(v^{(j)}). \tag{4.36}$$

Let $m_j := m$, $\tau_j := \rho^{m_j}$ *and* $v^{(j+1)} := v^{(j)} + \tau_j \hat{d}^{(j)}$.

(Step 2.4) *If*

$$\|H_{\mu_k,\varepsilon_k}(v^{(j+1)})\| \le \beta_k, \tag{4.37}$$

then let $w^{(k+1)} := v^{(j+1)}$ *and go to* Step 3. *Otherwise, set* $j := j+1$ *and go back to* Step 2.1.

(Step 3) *Update the parameters as follows*:

$$\begin{aligned} \mu_{k+1} &:= \min\left\{\kappa\|H_{\rm NR}(w^{(k+1)})\|^2,\ \mu_0\overline{\eta}^{k+1},\ \overline{\mu}\Big(\tilde{\lambda}(x^{(k+1)} - y^{(k+1)}), \hat{\kappa}\|H_{\rm NR}(w^{(k+1)})\|\Big)\right\} \\ \varepsilon_{k+1} &:= \min\left\{\overline{\kappa}\|H_{\rm NR}(w^{(k+1)})\|^2, \varepsilon_0\overline{\eta}^{k+1}\right\} \\ \beta_{k+1} &:= \beta_0\eta^{k+1}. \end{aligned}$$

Set $k := k+1$. *Go back to* Step 1.

In Step 3, $\tilde{\lambda}$ is the function given by (4.30), and $\overline{\mu}(\alpha,\delta)$ is determined so that $|\gamma'_\mu(\alpha) - \gamma_0^+(\alpha)| < \delta$ for any $\mu \in (0, \overline{\mu}(\alpha,\delta))$. An explicit formula of $\overline{\mu}(\alpha,\delta)$ is given as in Proposition 4.5 when $\hat{g}(\alpha) = (\alpha + \sqrt{\alpha^2+4})/2$.

In the inner iterations Steps 2.0 − 2.4, a damped Newton method seeks a point $w^{(k+1)}$ such that $\|H_{\mu_k,\varepsilon_k}(w^{(k+1)})\| \le \beta_k$. Note that, by letting $\alpha_k := \beta_k^2/2$, the termination criterion (4.37) for the inner iterations becomes equivalent to (3.12) in Algorithm 3.1. Step 3 specifies the update rule of the parameters. We note that $\{\beta_k\}$, $\{\mu_k\}$ and $\{\varepsilon_k\}$ converge to 0 since $0 < \overline{\eta} \le \eta < 1$.

The next proposition shows that Algorithm 4.1 is well-defined in the sense that the number of inner iterations Steps 2.0 − 2.4 is finite for each k.

Proposition 4.6 *For each k,* Step 2 *of Algorithm* 4.1 *finds a $w^{(k+1)} := v^{(j+1)}$ for some j.*

Proof. Fix k arbitrarily. Notice that we always have $\|H_{\rm NR}(w^{(k)})\| > 0$ in Step 2. First we show the existence of a nonnegative integer m satisfying (4.36). Without loss of generality, we assume $\Psi_{\mu,\varepsilon}(v^{(j)}) > 0$. Noticing that

$$\begin{aligned}
\nabla\Psi_{\mu_k,\varepsilon_k}(v^{(j)})^T\hat{d}^{(j)} &= H_{\mu_k,\varepsilon_k}(v^{(j)})^T\nabla H_{\mu_k,\varepsilon_k}(v^{(j)})^T\hat{d}^{(j)} \\
&= -H_{\mu_k,\varepsilon_k}(v^{(j)})^T H_{\mu_k,\varepsilon_k}(v^{(j)}) \\
&= -2\Psi_{\mu_k,\varepsilon_k}(v^{(j)}), \qquad (4.38)
\end{aligned}$$

we have

$$\begin{aligned}
&\Psi_{\mu_k,\varepsilon_k}(v^{(j)}+\tau\hat{d}^{(j)}) - (1-2\sigma\tau)\Psi_{\mu_k,\varepsilon_k}(v^{(j)}) \\
&\quad= \left\{\Psi_{\mu_k,\varepsilon_k}(v^{(j)}) + \tau\nabla\Psi_{\mu_k,\varepsilon_k}(v^{(j)})^T\hat{d}^{(j)} + o(\tau)\right\} - (1-2\sigma\tau)\Psi_{\mu_k,\varepsilon_k}(v^{(j)}) \\
&\quad= -2\tau\Psi_{\mu_k,\varepsilon_k}(v^{(j)}) + 2\sigma\tau\Psi_{\mu_k,\varepsilon_k}(v^{(j)}) + o(\tau) \\
&\quad= \tau\left\{-2(1-\sigma)\Psi_{\mu_k,\varepsilon_k}(v^{(j)}) + o(\tau)/\tau\right\}. \qquad (4.39)
\end{aligned}$$

Since $1-\sigma > 0$ and $\Psi_{\mu_k,\varepsilon_k}(v^{(j)}) > 0$, the right-hand side of (4.39) is negative for all $\tau > 0$ sufficiently small. Hence, (4.36) holds for some integer $m \geq 0$.

Next, for the contradiction purpose, we assume that the inner iterations of Step 2 never terminate, that is,

$$\|H_{\mu_k,\varepsilon_k}(v^{(j)})\| > \beta_k > 0 \qquad (4.40)$$

for all j at Step 2.4. We consider two cases (i) $\liminf_{j\to\infty}\tau_j > 0$ and (ii) $\liminf_{j\to\infty}\tau_j = 0$. In case (i), there exists $\overline{\tau} \in (0,(2\sigma)^{-1})$ such that $\tau_j \geq \overline{\tau}$ for all j. Since

$$\begin{aligned}
\Psi_{\mu_k,\varepsilon_k}(v^{(j)}) &= \Psi_{\mu_k,\varepsilon_k}(v^{(j-1)} + \tau_{j-1}\hat{d}^{(j-1)}) \\
&\leq (1-2\sigma\tau_{j-1})\Psi_{\mu_k,\varepsilon_k}(v^{(j-1)}),
\end{aligned}$$

we have

$$\begin{aligned}
\left\|H_{\mu_k,\varepsilon_k}(v^{(j)})\right\| &\leq \sqrt{1-2\sigma\tau_{j-1}}\,\left\|H_{\mu_k,\varepsilon_k}(v^{(j-1)})\right\| \\
&= \left\|H_{\mu_k,\varepsilon_k}(v^{(0)})\right\| \prod_{i=0}^{j-1}\sqrt{1-2\sigma\tau_i} \\
&\leq \left\|H_{\mu_k,\varepsilon_k}(v^{(0)})\right\| (1-2\sigma\overline{\tau})^{j/2}.
\end{aligned}$$

The right-hand side of this inequality tends to 0 as $j\to\infty$, which contradicts (4.40). In case (ii), there exists a subsequence such that $\{\tau_j\}_{j\in J} \to 0$. Note that Step 2.3 implies

$$\Psi_{\mu_k,\varepsilon_k}(v^{(j)} + \rho^{m_j-1}\hat{d}^{(j)}) > (1-2\sigma\rho^{m_j-1})\Psi_{\mu_k,\varepsilon_k}(v^{(j)})$$

for all j. Since $\rho^{m_j-1} = \tau_j\rho^{-1}$, we then have

$$\begin{aligned}
-2\sigma\Psi_{\mu_k,\varepsilon_k}(v^{(j)}) &< \frac{\Psi_{\mu_k,\varepsilon_k}(v^{(j)}+\tau_j\rho^{-1}\hat{d}^{(j)}) - \Psi_{\mu_k,\varepsilon_k}(v^{(j)})}{\tau_j\rho^{-1}} \\
&= \nabla\Psi_{\mu_k,\varepsilon_k}(v^{(j)})^T\hat{d}^{(j)} + o(\tau_j)/\tau_j \\
&= -2\Psi_{\mu_k,\varepsilon_k}(v^{(j)}) + o(\tau_j)/\tau_j,
\end{aligned}$$

that is,

$$2(1-\sigma)\Psi_{\mu_k,\,\varepsilon_k}(v^{(j)}) \;<\; o(\tau_j)/\tau_j. \tag{4.41}$$

However, (4.41) cannot hold when $j \in J$ is sufficiently large, since $1-\sigma > 0$ and $\Psi_{\mu_k,\,\varepsilon_k}(v^{(j)}) > \beta_k^2/2 > 0$. Thus, we have a contradiction. This completes the proof. ■

Finally, we show that Algorithm 4.1 is quadratically convergent under appropriate assumptions. Note that there exists $\nu > 0$ such that, for any $w \in \Re^{2n}$, the following inequality holds:

$$\begin{aligned}
\|H_{\mu,\varepsilon}(w) - H_{\rm NR}(w)\| &\le \left\| \begin{pmatrix} \varphi_\mu(w) \\ f_\varepsilon(x) - y \end{pmatrix} - \begin{pmatrix} \varphi_{\rm NR}(w) \\ f_\varepsilon(x) - y \end{pmatrix} \right\| + \left\| \begin{pmatrix} \varphi_{\rm NR}(w) \\ f_\varepsilon(x) - y \end{pmatrix} - \begin{pmatrix} \varphi_{\rm NR}(w) \\ f(x) - y \end{pmatrix} \right\| \\
&= \|\varphi_\mu(w) - \varphi_{\rm NR}(w)\| + \|\{f_\varepsilon(x) - y\} - \{f(x) - y\}\| \\
&\le \nu\mu + \varepsilon\|x\| \\
&\le \nu\mu + \varepsilon\|w\|,
\end{aligned} \tag{4.42}$$

where the first inequality follows from the triangle inequality, the second inequality follows from (3.6), and the third inequality follows from $\|x\| \le \|(x,y)\|$. By using the inequality (4.42) and the two properties shown in Subsections 4.2 and 4.3, we establish the main theorem of this subsection.

Theorem 4.4 *Let $f : \Re^n \to \Re^n$ be a monotone function such that ∇f is locally Lipschitzian. Let $\{w^{(k)}\}$ be a sequence generated by Algorithm* 4.1*. Moreover, suppose that the following assumptions hold true:*

(i) *The solution set of SOCCP* (3.1) *is nonempty and bounded.*

(ii) *Every accumulation point of $\{\nabla H_{\mu_k,\varepsilon_k}(w^{(k)})\}$ is nonsingular.*

Then we have the following two statements:

(a) *For all k sufficiently large, the inequality in* Step 2.2 *of Algorithm* 4.1 *holds for $j = 0$;*

(b) *The sequence $\{w^{(k)}\}$ converges to a solution w^* of SOCCP* (3.1) *quadratically.*

Proof. By assumption (i) and Theorem 3.2, $\{w^{(k)}\}$ is bounded and an arbitrary accumulation point w^* is a solution of SOCCP (3.1), that is, $H_{\rm NR}(w^*) = 0$. Let B be a positive number such that $\|w^{(k)}\| \le B$ for all k. Let $L > 0$ be a Lipschitzian constant of $H_{\rm NR}$ on a bounded set $\Omega \supset \{w^{(k)}\}$. Let $d^{(k)} := \hat{d}^{(0)} = -\nabla H_{\mu_k,\varepsilon_k}(w^{(k)})^{-T} H_{\mu_k,\varepsilon_k}(w^{(k)})$, where J^{-T} denotes $J^{-T} := (J^T)^{-1} = (J^{-1})^T$. From Proposition 3.1 and assumption (ii), there exists $C > 0$ such that

$$\|\nabla H_{\mu_k,\varepsilon_k}(w^{(k)})^{-T}\| \;=\; \|\nabla H_{\mu_k,\varepsilon_k}(w^{(k)})^{-1}\| \;\le\; C \tag{4.43}$$

for all k.

First we show $\|w^{(k)} + d^{(k)} - w^*\| = O(\|w^{(k)} - w^*\|^2)$. Let V_k be an element of $\partial H_{\rm NR}(w^{(k)})$ such that $\|\nabla H_{\mu_k,\varepsilon_k}(w^{(k)}) - V_k\| = {\rm dist}(\nabla H_{\mu_k,\varepsilon_k}(w^{(k)}), \partial H_{\rm NR}(w^{(k)}))$. It then follows from (4.43) and $H_{\rm NR}(w^*) = 0$ that

$$\begin{aligned}
&\|w^{(k)} + d^{(k)} - w^*\| \\
&\quad = \left\| w^{(k)} - \nabla H_{\mu_k,\varepsilon_k}(w^{(k)})^{-T} H_{\mu_k,\varepsilon_k}(w^{(k)}) - w^* \right\| \\
&\quad \le \left\| \nabla H_{\mu_k,\varepsilon_k}(w^{(k)})^{-T} \right\| \left\| \nabla H_{\mu_k,\varepsilon_k}(w^{(k)})^T (w^{(k)} - w^*) - H_{\mu_k,\varepsilon_k}(w^{(k)}) \right\| \\
&\quad \le C\Big\{ \left\| (\nabla H_{\mu_k,\varepsilon_k}(w^{(k)}) - V_k)^T (w^{(k)} - w^*) \right\| + \left\| V_k^T (w^{(k)} - w^*) - H'_{\rm NR}(w^*; w^{(k)} - w^*) \right\| \\
&\qquad + \left\| H'_{\rm NR}(w^*; w^{(k)} - w^*) - H_{\rm NR}(w^{(k)}) + H_{\rm NR}(w^*) \right\| + \left\| H_{\rm NR}(w^{(k)}) - H_{\mu_k,\varepsilon_k}(w^{(k)}) \right\| \Big\}.
\end{aligned} \tag{4.44}$$

Moreover, each term of (4.44) is $O(\|w^{(k)} - w^*\|^2)$ by the following reasons. Since $H_{\mu,\varepsilon}$ satisfies the Jacobian consistency, we have for some $M > 0$ and $M' > 0$ that

$$\begin{aligned}
\|\nabla H_{\mu_k,\varepsilon_k}(w^{(k)}) - V_k\| &\leq M\Big(M'\left|\gamma'_{\mu_k}(\tilde{\lambda}(x^{(k)} - y^{(k)})) - \gamma_0^+(\tilde{\lambda}(x^{(k)} - y^{(k)}))\right| + \varepsilon_k\Big) \\
&\leq M(M'\hat{\kappa}\|H_{\mathrm{NR}}(w^{(k)})\| + \overline{\kappa}\|H_{\mathrm{NR}}(w^{(k)})\|^2) \\
&\leq M(M'\hat{\kappa} + \overline{\kappa}\|H_{\mathrm{NR}}(w^{(k)})\|)\|H_{\mathrm{NR}}(w^{(k)}) - H_{\mathrm{NR}}(w^*)\| \\
&\leq M(M'\hat{\kappa} + \overline{\kappa}\|H_{\mathrm{NR}}(w^{(k)})\|)L\|w^{(k)} - w^*\|, \qquad (4.45)
\end{aligned}$$

where the first inequality follows from (4.25) and Proposition 4.4, the second inequality follows from Step 3 of the algorithm, the third inequality follows from $H_{\mathrm{NR}}(w^*) = 0$, and the last inequality is due to the local Lipschitz continuity of H_{NR}. By (4.45) and the boundedness of $\{\|H_{\mathrm{NR}}(w^{(k)})\|\}$, we have $\|(\nabla H_{\mu_k,\varepsilon_k}(w^{(k)}) - V_k)^T(w^{(k)} - w^*)\| = O(\|w^{(k)} - w^*\|^2)$. From Theorem 4.2, H_{NR} is directionally differentiable and strongly semismooth, which implies

$$\|H'_{\mathrm{NR}}(w^*; w^{(k)} - w^*) - H_{\mathrm{NR}}(w^{(k)}) + H_{\mathrm{NR}}(w^*)\| = O(\|w^{(k)} - w^*\|^2)$$

and

$$\|V_k^T(w^{(k)} - w^*) - H'_{\mathrm{NR}}(w^*; w^{(k)} - w^*)\| = O(\|w^{(k)} - w^*\|^2).$$

By (4.42), we have

$$\begin{aligned}
\left\|H_{\mathrm{NR}}(w^{(k)}) - H_{\mu_k,\varepsilon_k}(w^{(k)})\right\| &\leq \nu\mu_k + B\varepsilon_k \\
&\leq (\nu\kappa + B\overline{\kappa})\|H_{\mathrm{NR}}(w^{(k)}) - H_{\mathrm{NR}}(w^*)\|^2 \\
&\leq L^2(\nu\kappa + B\overline{\kappa})\|w^{(k)} - w^*\|^2,
\end{aligned}$$

where the second inequality follows from $H_{\mathrm{NR}}(w^*) = 0$ and Step 3 of Algorithm 4.1. The above inequalities together with (4.44) yield

$$\|w^{(k)} + d^{(k)} - w^*\| = O(\|w^{(k)} - w^*\|^2). \qquad (4.46)$$

Next we show that $\{d^{(k)}\}$ converges to 0. Note that there exists $\Theta > 0$ such that the following inequalities hold for all k:

$$\begin{aligned}
\|d^{(k)}\| &= \|-\nabla H_{\mu_k,\varepsilon_k}(w^{(k)})^{-T} H_{\mu_k,\varepsilon_k}(w^{(k)})\| \\
&\leq \|\nabla H_{\mu_k,\varepsilon_k}(w^{(k)})^{-T}\|\|H_{\mu_k,\varepsilon_k}(w^{(k)})\| \\
&\leq C(\|H_{\mu_k,\varepsilon_k}(w^{(k)}) - H_{\mathrm{NR}}(w^{(k)})\| + \|H_{\mathrm{NR}}(w^{(k)})\|) \\
&\leq C(\nu\mu_k + B\varepsilon_k + \|H_{\mathrm{NR}}(w^{(k)})\|) \\
&\leq C\{1 + (\nu\kappa + B\overline{\kappa})\|H_{\mathrm{NR}}(w^{(k)})\|\}\|H_{\mathrm{NR}}(w^{(k)})\| \\
&\leq \Theta\|H_{\mathrm{NR}}(w^{(k)})\|, \qquad (4.47)
\end{aligned}$$

where the second inequality follows from (4.43) and the triangle inequality, the third inequality follows from (4.42) and $\|w^{(k)}\| \leq B$, the fourth inequality follows from Step 3 of Algorithm 4.1, and the last inequality follows from the boundedness of $\{\|H_{\mathrm{NR}}(w^{(k)})\|\}$. Moreover, the following inequalities hold:

$$\begin{aligned}
\|H_{\mathrm{NR}}(w^{(k)})\| &\leq \|H_{\mu_{k-1},\varepsilon_{k-1}}(w^{(k)})\| + \|H_{\mu_{k-1},\varepsilon_{k-1}}(w^{(k)}) - H_{\mathrm{NR}}(w^{(k)})\| \\
&\leq \beta_{k-1} + \nu\mu_{k-1} + B\varepsilon_{k-1} \\
&\leq \beta_0\eta^{k-1} + \nu\mu_0\overline{\eta}^{k-1} + B\varepsilon_0\overline{\eta}^{k-1} \\
&\leq \beta_0\eta^{k-1} + \nu\mu_0\eta^{k-1} + B\varepsilon_0\eta^{k-1} \\
&= \eta^{-1}(\beta_0 + \nu\mu_0 + B\varepsilon_0)\eta^k,
\end{aligned}$$

where the second inequality follows from Step 2.4 of Algorithm 4.1 and (4.42), the third inequality follows from Step 3 in Algorithm 4.1, and the fourth inequality follows from $0 < \overline{\eta} \le \eta < 1$. Hence, by setting $\Gamma := \eta^{-1}(\beta_0 + \nu\mu_0 + B\varepsilon_0)$, we have

$$\|H_{\rm NR}(w^{(k)})\| \ \le\ \Gamma\eta^k \tag{4.48}$$

for all k. Combining (4.47) and (4.48), we have

$$\|d^{(k)}\| \ \le\ \Theta\Gamma\eta^k \tag{4.49}$$

for all k. Thus, $\{d^{(k)}\}$ converges to 0.

Next we show $\|H_{\rm NR}(w^{(k)} + d^{(k)})\| = O(\|H_{\rm NR}(w^{(k)})\|^2)$ in a similar way to [18, Theorem 3.1]. By (4.46), there exists $\gamma > 0$ such that

$$\|w^{(k)} + d^{(k)} - w^*\| \le \gamma\|w^{(k)} - w^*\|^2. \tag{4.50}$$

Now, let k be sufficiently large so that $0 < \gamma\|w^{(k)} - w^*\| \le 1/2$. By the directional differentiability of $H_{\rm NR}$, we have

$$H_{\rm NR}(w^{(k)} + d^{(k)}) - H_{\rm NR}(w^*) = H'_{\rm NR}(w^*; w^{(k)} + d^{(k)} - w^*) + o(\|w^{(k)} + d^{(k)} - w^*\|).$$

Therefore, there exists a positive sequence $\{\delta_k\}$ tending to 0 such that

$$\begin{aligned}\|H_{\rm NR}(w^{(k)} + d^{(k)})\| &\le \|H'_{\rm NR}(w^*; w^{(k)} + d^{(k)} - w^*)\| + \delta_k\|w^{(k)} + d^{(k)} - w^*\| \\ &\le (\overline{L} + \delta_k)\|w^{(k)} + d^{(k)} - w^*\| \\ &\le (\overline{L} + \delta_k)\gamma\|w^{(k)} - w^*\|^2,\end{aligned} \tag{4.51}$$

where $\overline{L}$ is Lipschitzian constant for $H_{\rm NR}$ on a bounded set $\hat{\Omega} \supset \{w^{(k)} + d^{(k)}\}$, and the third inequality follows from (4.50). From (4.47) and (4.50), we have

$$\begin{aligned}\|w^{(k)} - w^*\| &\le \|d^{(k)}\| + \|w^{(k)} + d^{(k)} - w^*\| \\ &\le \Theta\|H_{\rm NR}(w^{(k)})\| + \gamma\|w^{(k)} - w^*\|^2 \\ &\le \Theta\|H_{\rm NR}(w^{(k)})\| + \frac{1}{2}\|w^{(k)} - w^*\|,\end{aligned}$$

that is,

$$\|w^{(k)} - w^*\| \ \le\ 2\Theta\|H_{\rm NR}(w^{(k)})\|. \tag{4.52}$$

Thus, (4.51) and (4.52) yield

$$\begin{aligned}\|H_{\rm NR}(w^{(k)} + d^{(k)})\| &\le 4\Theta^2\gamma(\overline{L} + \delta_k)\|H_{\rm NR}(w^{(k)})\|^2 \\ &= O(\|H_{\rm NR}(w^{(k)})\|^2).\end{aligned} \tag{4.53}$$

Finally, we show that, for all k sufficiently large, Step 2.2 of Algorithm 4.1 holds for $j = 0$, that is, $\|H_{\mu_k,\varepsilon_k}(w^{(k)} + d^{(k)})\| \le \beta_k$. From (4.53), there exists $\theta > 0$ such that $\|H_{\rm NR}(w^{(k)} + d^{(k)})\| \le \theta\|H_{\rm NR}(w^{(k)})\|^2$. Then we have

$$\begin{aligned}\|H_{\mu_k,\varepsilon_k}(w^{(k)} + d^{(k)})\| &\le \|H_{\mu_k,\varepsilon_k}(w^{(k)} + d^{(k)}) - H_{\rm NR}(w^{(k)} + d^{(k)})\| + \|H_{\rm NR}(w^{(k)} + d^{(k)})\| \\ &\le \nu\mu_k + (B + \Theta\Gamma\eta^k)\varepsilon_k + \theta\|H_{\rm NR}(w^{(k)})\|^2 \\ &\le (\nu\kappa + (B + \Theta\Gamma\eta^k)\overline{\kappa} + \theta)\|H_{\rm NR}(w^{(k)})\|^2 \\ &\le (\nu\kappa + (B + \Theta\Gamma\eta^k)\overline{\kappa} + \theta)\Gamma\eta^k\|H_{\rm NR}(w^{(k)})\| \\ &= (\nu\kappa + (B + \Theta\Gamma\eta^k)\overline{\kappa} + \theta)\Gamma\beta_0^{-1}\beta_k\|H_{\rm NR}(w^{(k)})\|,\end{aligned} \tag{4.54}$$

where the second inequality follows from (4.42), (4.49) and (4.53), the third inequality follows from Step 3 of Algorithm 4.1, the fourth inequality follows from (4.48), and the equality follows from Step 3 of Algorithm 4.1. Furthermore, since $(\nu\kappa + (B + \Theta\Gamma\eta^k)\overline{\kappa} + \theta)\Gamma\beta_0^{-1}\|H_{\rm NR}(w^{(k)})\| \le 1$ for all k sufficiently large, (4.54) yields $\|H_{\mu_k,\varepsilon_k}(w^{(k)} + d^{(k)})\| \le \beta_k$, which shows (a). Moreover, this implies $w^{(k+1)} = w^{(k)} + d^{(k)}$, and hence, we have from (4.46) that $\|w^{(k+1)} - w^*\| = O(\|w^{(k)} - w^*\|^2)$. This shows (b). ∎

5 Numerical experiments

In order to evaluate the efficiency of Algorithm 4.1, we have conducted some numerical experiments. In our experiments, we chose $\hat{g}(\alpha) = (\sqrt{\alpha^2+4}+\alpha)/2$ as the function satisfying (3.4) and (4.26), and employed $\|H_{\rm NR}(w^{(k)})\| < 10^{-8}$ as the termination criterion. Moreover, we adopted

$$\overline{\mu}(\alpha,\delta) := \begin{cases} 10^{10} & (\delta \ge 1/2 \ \text{ or } \ \alpha = 0) \\ \dfrac{1}{2}|\alpha|\sqrt{\delta} & (\delta < 1/2 \ \text{ and } \ \alpha \ne 0)\,. \end{cases}$$

in Step 3 of the algorithm. Note that, when $\tilde{\lambda}(x^{(k)} - y^{(k)})$ defined by (4.30) is very small but positive while $\|H_{\rm NR}(x^{(k)}, y^{(k)})\|$ is not small enough, μ_k becomes almost as small as $\tilde{\lambda}(x^{(k)} - y^{(k)})$, although $(x^{(k)}, y^{(k)})$ is not sufficiently close to a solution. Since this may diminish the effects of smoothing, we regard $\lambda_i(x^{(k)} - y^{(k)})$ as 0 when $\lambda_i(x^{(k)} - y^{(k)})/\|H_{\rm NR}(x^{(k)}, y^{(k)})\| \le \overline{\delta}$ holds for a sufficiently small $\overline{\delta} > 0$. In our numerical experiments, we therefore modified the definition (4.30) of $\tilde{\lambda}(z)$ such that the index set $\mathcal{I}(z) := \{i \mid |\lambda_i(z)| > 10^{-4}\|H_{\rm NR}(x,y)\|\}$ is used instead of $\mathcal{I}(z) = \{i \mid \lambda_i(z) \ne 0\}$. The program was coded in MATLAB 6.5 and run on a machine with AMD Athlon(tm) XP 2000+ CPU and 1GB RAM.

5.1 Linear case

We solved the following linear SOCCP :

$$\begin{array}{rl} \text{Find} & (x,y) \in \Re^n \times \Re^n \\ \text{such that} & x \in \mathcal{K}^n,\ y \in \mathcal{K}^n,\ x^T y = 0,\ y = Mx + q, \end{array} \tag{5.1}$$

where $M \in \Re^{n\times n}$ is a rank-deficient positive semidefinite matrix. We chose $\eta = 0.01$, $\overline{\eta} = 0.001$, $\rho = 0.5$, $\sigma = 0.4$, $\kappa = \overline{\kappa} = 0.01$ and $\hat{\kappa} = 1$ in Algorithm 4.1.

In order to obtain a positive semidefinite matrix M with $\operatorname{rank} M = r < n$, we let $M := nBB^T/\|BB^T\|$, where $B \in \Re^{n\times r}$ is a matrix of which every components are randomly chosen from the interval $[-1,1]$. Furthermore, we let $q := 10^\alpha n^{1/2} p - Me$, where $e = (1,0,\ldots,0)^T \in \operatorname{int}\mathcal{K}^n$, p is a vector such that $p \in \operatorname{int}\mathcal{K}^n$ and $\|p\| = 1$, and α is randomly chosen from the interval $[-1,1]$. Then, SOCCP (5.1) has a solution since the assumption of Theorem 6.1 (to be shown in the next section) holds by letting $\overline{x} = e \in \operatorname{int}\mathcal{K}^n$ and $\overline{y} = Me + q = 10^\alpha n^{1/2} p \in \operatorname{int}\mathcal{K}^n$. In the experiments, we determined p as follows : Let θ be randomly chosen from $(0,\pi/2)$, let w be an $(n-1)$ - dimensional vector whose components are randomly chosen from $[-1,1]$, and put $p := 2^{-1/2}\cos\theta(1, w/\|w\|) + 2^{-1/2}\sin\theta(1, -w/\|w\|)$. It is not difficult to see that a vector p thus determined satisfies $\|p\| = 1$ and $p \in \operatorname{int}\mathcal{K}^n$.

In our first experiments, we generated 100 problem instances for each of $n = 100, 200, \ldots, 1000$, and solved each problem instance by using 100 randomly chosen initial points. We let the rank r of M be an integer randomly chosen from $[0.9n, n-1]$, and selected an initial point $(x^{(0)}, y^{(0)})$ as $(x^{(0)}, y^{(0)}) := 10^\beta (a,b)/\|(a,b)\|$, where β is randomly chosen from $[-3,3]$ and each component of $(a,b) \in \Re^n \times \Re^n$ is randomly chosen from $[-1,1]$. Table 1 shows the results of our experiments, in which n denotes the size of problems, ♯Ite denotes the number of outer iterations, ♯Newton denotes the total number of inner Newton iterations, and cpu(s) denotes the CPU time. The values of $\sharp$Ite, $\sharp$Newton and cpu(s) are the average of 100 runs for each n. Table 1 reveals that the problem size only slightly affects the number of iterations.

Table 1: Results for linear SOCCPs of various problem sizes

n	$\sharp$Ite	$\sharp$Newton	cpu(s)
100	5.28	7.12	0.151
200	5.58	7.93	1.026
300	5.86	8.48	2.848
400	5.65	8.80	5.960
500	5.77	8.89	10.603
600	5.80	9.12	17.346
700	5.96	9.37	26.252
800	5.99	9.36	37.185
900	5.93	9.71	52.223
1000	5.98	9.52	67.316

In our second experiments, we fix the size of problems at 100 and varied the rank of $M \in \Re^{100\times 100}$ as $r = 10, 20, \ldots, 90$ and 99. We solved 1000 different problem instances for each r, where initial points are selected in a way similar to the previous experiments. Table 2 shows the results of the experiments. The numbers of iterations, $\sharp$Ite and $\sharp$Newton, are the average of 1000 runs for each r. Table 2 indicates that neither $\sharp$Ite nor $\sharp$Newton are affected by the rank of matrix M.

Table 2: Results for linear SOCCPs with various degrees of rank deficiency

r	$\sharp$Ite	$\sharp$Newton
99	5.396	7.346
90	5.361	7.260
80	5.284	7.207
70	5.245	7.115
60	5.242	7.189
50	5.172	7.004
40	5.233	7.128
30	5.116	6.988
20	5.072	6.809
10	4.894	6.371

5.2 Nonlinear case

In our third experiments, we solved the following nonlinear SOCCP:

$$\begin{array}{rl} \text{Find} & (x, y) \in \Re^5 \times \Re^5 \\ \text{such that} & x \in \mathcal{K},\ y \in \mathcal{K},\ x^T y = 0,\ y = f(x), \end{array} \tag{5.2}$$

where $\mathcal{K} = \mathcal{K}^3 \times \mathcal{K}^2$ and $f : \Re^5 \to \Re^5$ is given by

$$f(x) := \begin{pmatrix} 24(2x_1 - x_2)^3 + \exp(x_1 - x_3) - 4x_4 + x_5 \\ -12(2x_1 - x_2)^3 + 3(3x_2 + 5x_3)/\sqrt{1 + (3x_2 + 5x_3)^2} - 6x_4 - 7x_5 \\ -\exp(x_1 - x_3) + 5(3x_2 + 5x_3)/\sqrt{1 + (3x_2 + 5x_3)^2} - 3x_4 + 5x_5 \\ 4x_1 + 6x_2 + 3x_3 - 1 \\ -x_1 + 7x_2 - 5x_3 + 2 \end{pmatrix}.$$

We note that, in view of the KKT conditions (6.3) for the SOCP (1.2) (see Section 6), SOCCP (5.2) is equivalent to the following SOCP:

$$\begin{array}{ll} \text{Minimize} & \exp(z_1 - z_3) + 3(2z_1 - z_2)^4 + \sqrt{1 + (3z_2 + 5z_3)^2} \\ \text{subject to} & \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} \in \mathcal{K}^3, \quad \begin{pmatrix} 4 & 6 & 3 \\ -1 & 7 & -5 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} + \begin{pmatrix} -1 \\ 2 \end{pmatrix} \in \mathcal{K}^2. \end{array}$$

Since the objective function of this SOCP is convex, we can easily see that the function f is monotone. In the experiments, we set the parameters as $\rho = 0.5$, $\sigma = 0.4$, $\kappa = \overline{\kappa} = 0.01$ and $\hat{\kappa} = 1$.

We used nine different pairs of $(\eta, \overline{\eta})$ and ran Algorithm 4.1 with 100 initial points $(x^{(0)}, y^{(0)})$ for each $(\eta, \overline{\eta})$ determined as $(x^{(0)}, y^{(0)}) := \Gamma(a, b)/\|(a, b)\|$, where Γ is randomly chosen from $[0, 10]$ and each component of $(a, b) \in \Re^n \times \Re^n$ is randomly chosen from $[-1, 1]$. Table 3 shows the results of the experiments, in which $\sharp$Ite and $\sharp$Newton are the average of 100 runs for each $(\eta, \overline{\eta})$. As Table 3 shows, the number of outer iterations decreases as η becomes smaller. On the other hand, the number of inner iterations decreases as $\overline{\eta}$ becomes smaller and η becomes larger. Since the computation time is largely dependent on the number of inner iterations, it would be recommended to choose η as large as possible and $\overline{\eta}$ as small as possible. However, it should be kept in mind that, when $\overline{\eta}$ is too small, μ and ε tend to 0 very rapidly, which may deteriorate the advantage of smoothing and regularization methods.

Table 3: Results for nonlinear SOCCPs with various choices of $(\eta, \overline{\eta})$

η	$\overline{\eta}$	$\sharp$Ite	$\sharp$Newton
0.5	0.5	16.79	26.34
0.5	0.1	12.20	13.70
0.5	0.01	10.50	10.75
0.5	0.001	9.58	9.85
0.1	0.1	8.35	19.11
0.1	0.01	8.13	13.51
0.1	0.001	7.66	11.25
0.01	0.01	5.99	14.32
0.01	0.001	5.73	12.35

6 Concluding Remarks

In this section, we make some comments that complement the results obtained in the paper.

In Sections 3 and 4, the function F in SOCCP (1.1) is assumed to be of the form $F(x, y, \zeta) = f(x) - y$. This assumption may seem rather restrictive. However, the KKT conditions for SOCP (1.2) can be written as the SOCCP with $F(x, y, \zeta) = f(x) - y$ as follows: In SOCP (1.2), let $z = z' - z''$

with $z' \in \Re^s_+$ and $z'' \in \Re^s_+$, and denote $\hat{z} := \binom{z'}{z''} \in \Re^{2s}_+$, where $\Re^n_+$ is the n-dimensional nonnegative orthant. Moreover, define $\hat{\theta} : \Re^{2s} \to \Re$ by $\hat{\theta}(\hat{z}) = \theta(z' - z'')$, and $\hat{\gamma} : \Re^{2s} \to \Re^t$ by $\hat{\gamma}(\hat{z}) = \gamma(z' - z'')$. Then SOCP (1.2) can be reformulated as

$$\begin{array}{ll} \text{Minimize} & \hat{\theta}(\hat{z}) \\ \text{subject to} & \begin{pmatrix} \hat{\gamma}(\hat{z}) \\ \hat{z} \end{pmatrix} \in \mathcal{K} \times \Re^{2s}_+, \end{array} \tag{6.1}$$

and the KKT conditions for (6.1) are written as

$$\begin{aligned} &\nabla\hat{\theta}(\hat{z}) - \begin{pmatrix} \nabla\hat{\gamma}(\hat{z}) & I \end{pmatrix} \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix} = 0, \\ &\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix} \in \mathcal{K} \times \Re^{2s}_+, \quad \begin{pmatrix} \hat{\gamma}(\hat{z}) \\ \hat{z} \end{pmatrix} \in \mathcal{K} \times \Re^{2s}_+, \quad \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix}^T \begin{pmatrix} \hat{\gamma}(\hat{z}) \\ \hat{z} \end{pmatrix} = 0. \end{aligned} \tag{6.2}$$

Now, let $\hat{\mu}_1 = \hat{\gamma}(\hat{z})$, and notice that Proposition 2.2 holds. Then, (6.2) can be rewritten as

$$\begin{aligned} &\begin{pmatrix} \hat{\gamma}(\hat{z}) \\ \nabla\hat{\theta}(\hat{z}) - \nabla\hat{\gamma}(\hat{z})\hat{\lambda}_1 \end{pmatrix} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\lambda}_2 \end{pmatrix}, \\ &\begin{pmatrix} \hat{\lambda}_1 \\ \hat{z} \end{pmatrix} \in \mathcal{K} \times \Re^{2s}_+, \quad \begin{pmatrix} \hat{\mu}_1 \\ \hat{\lambda}_2 \end{pmatrix} \in \mathcal{K} \times \Re^{2s}_+, \quad \begin{pmatrix} \hat{\lambda}_1 \\ \hat{z} \end{pmatrix}^T \begin{pmatrix} \hat{\mu}_1 \\ \hat{\lambda}_2 \end{pmatrix} = 0. \end{aligned} \tag{6.3}$$

Setting

$$x = \begin{pmatrix} \hat{\lambda}_1 \\ \hat{z} \end{pmatrix}, \quad y = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\lambda}_2 \end{pmatrix}, \quad f(x) = \begin{pmatrix} \hat{\gamma}(\hat{z}) \\ \nabla\hat{\theta}(\hat{z}) - \nabla\hat{\gamma}(\hat{z})\hat{\lambda}_1 \end{pmatrix}, \tag{6.4}$$

the KKT conditions (6.3) for SOCP (1.2) can be reduced to the SOCCP with $F(x, y, \zeta) = f(x) - y$. Note that the KKT conditions (6.3) for SOCP (1.2) contain more variables than the original KKT conditions (1.3). Furthermore, some desirable properties of the functions involved in SOCP (1.2) may be lost. For example, even if θ and γ in SOCP (1.2) are strictly convex, $\hat{\theta}$ and $\hat{\gamma}$ in (6.1) are merely convex. Hence, it would be useful to develop a method that can directly deal with the KKT conditions (1.3), or more generally, SOCCP involving the function $F(x, y, \zeta)$ which is not restricted to be of the form $F(x, y, \zeta) = f(x) - y$.

In Sections 3 and 4, it is also assumed that $\mathcal{K} = \mathcal{K}^n$. For the general case where $\mathcal{K} = \mathcal{K}^{n_1} \times \cdots \times \mathcal{K}^{n_m}$, $H_{\mu,\varepsilon}(x, y)$ and $\nabla H_{\mu,\varepsilon}(x, y)$ are given by

$$\begin{aligned} H_{\mu,\varepsilon}(x, y) &= \begin{pmatrix} \varphi_\mu(x^1, y^1) \\ \vdots \\ \varphi_\mu(x^m, y^m) \\ f_\varepsilon(x) - y \end{pmatrix}, \\ \nabla H_{\mu,\varepsilon}(x, y) &= \begin{pmatrix} I - \operatorname{diag}\{\nabla P_\mu(x^j, y^j)\}_{j=1}^m & \nabla f(x) + \varepsilon I \\ \operatorname{diag}\{\nabla P_\mu(x^j, y^j)\}_{j=1}^m & -I \end{pmatrix}, \end{aligned}$$

where $x = (x^1, \ldots, x^m) \in \Re^{n_1} \times \cdots \times \Re^{n_m}$, $y = (y^1, \ldots, y^m) \in \Re^{n_1} \times \cdots \times \Re^{n_m}$, and $\operatorname{diag}\{\nabla P_\mu(x^j, y^j)\}_{j=1}^m$ denotes the block-diagonal matrix with block elements $\nabla P_\mu(x^j, y^j)$ given by (3.14). Moreover, the definition (4.30) of $\tilde{\lambda}(z)$ is replaced by

$$\tilde{\lambda}(z) := \begin{cases} \min\limits_{(i,j) \in \mathcal{I}(z)} |\lambda_i(z^j)| & (\mathcal{I}(z) \neq \emptyset) \\ 0 & (\mathcal{I}(z) = \emptyset), \end{cases}$$

where $z = (z^1, \ldots, z^m) \in \Re^{n_1} \times \cdots \times \Re^{n_m}$, $\lambda_i(z^j)$ $(i = 1, 2)$ are the spectral values of z^j, and $\mathcal{I}(z) \subseteq \{1, 2\} \times \{1, \ldots, m\}$ is the index set defined by $\mathcal{I}(z) := \{(i, j) \mid \lambda_i(z^j) \neq 0\}$. Then, we can also show that Theorems 3.2 and 4.4 hold for monotone f in a similar way.

In the numerical experiments of Section 5, we generated linear SOCCP (5.1) whose solution set is nonempty, based on the following theorem which shows a sufficient condition for a linear SOCCP to be solvable. This result is a natural extension of the well-known result [7, Theorem 3.1.2] for the linear complementarity problem.

Theorem 6.1 *Let $M \in \Re^{n \times n}$ be a positive semidefinite matrix, and $q \in \Re^n$ be a real-valued vector. If there exists a pair $(\overline{x}, \overline{y})$ such that $\overline{x} \in \operatorname{int} \mathcal{K}$, $\overline{y} \in \operatorname{int} \mathcal{K}$ and $\overline{y} = M\overline{x} + q$, then the SOCCP*

$$x \in \mathcal{K},\ y \in \mathcal{K},\ x^T y = 0,\ y = Mx + q \tag{6.5}$$

has at least one solution.

Proof. First we consider the SOCP:

$$\begin{array}{ll} \text{Minimize} & z^T(Mz + q) \\ \text{subject to} & Mz + q \in \mathcal{K},\ z \in \mathcal{K}\,, \end{array} \tag{6.6}$$

which is strictly feasible by assumption. Since it holds that

$$\xi \in \mathcal{K},\ \eta \in \mathcal{K} \implies \xi^T \eta \geq 0, \tag{6.7}$$

the objective function of SOCP (6.6) is bounded below on the feasible region. Therefore, there exist an optimal solution $z^* \in \Re^n$ and a multiplier vector $\lambda^* \in \Re^n$ satisfying the following KKT conditions:

$$z^* \in \mathcal{K}, \tag{6.8}$$
$$(M + M^T)z^* + q - M^T\lambda^* \in \mathcal{K}, \tag{6.9}$$
$$(z^*)^T((M + M^T)z^* + q - M^T\lambda^*) = 0, \tag{6.10}$$
$$\lambda^* \in \mathcal{K}, \tag{6.11}$$
$$Mz^* + q \in \mathcal{K}, \tag{6.12}$$
$$(\lambda^*)^T(Mz^* + q) = 0. \tag{6.13}$$

From (6.10) we have

$$-(z^*)^T M^T (z^* - \lambda^*) \;=\; (z^*)^T(Mz^* + q) \;\geq\; 0, \tag{6.14}$$

where the inequality follows from (6.8), (6.12) and (6.7). Moreover, we have

$$\begin{aligned} 0 \;&\leq\; (\lambda^*)^T((M + M^T)z^* + q - M^T\lambda^*) \\ &=\; (\lambda^*)^T(Mz^* + q) + (\lambda^*)^T M^T (z^* - \lambda^*) \\ &=\; (\lambda^*)^T M^T (z^* - \lambda^*), \end{aligned} \tag{6.15}$$

where the inequality follows from (6.11), (6.9) and (6.7), and the last equality follows from (6.13). Hence, by (6.14) and (6.15), we have

$$(z^* - \lambda^*)^T M^T (z^* - \lambda^*) \;\leq\; 0. \tag{6.16}$$

Noticing the positive semidefiniteness of M, the inequality in (6.16) is replaced by the equality. Hence,

$$(z^*)^T M^T (z^* - \lambda^*) \;=\; (\lambda^*)^T M^T (z^* - \lambda^*) \;\geq\; 0, \tag{6.17}$$

where the inequality is due to (6.15). By (6.14) and (6.17), we have

$$(z^*)^T(Mz^* + q) \;=\; -(z^*)^T M^T (z^* - \lambda^*) \;=\; 0,$$

which implies $(x, y) := (z^*, Mz^* + q)$ is a solution of SOCCP (6.5). ■

References

[1] C. Chen and O. L. Mangasarian, *A class of smoothing functions for nonlinear and mixed complementarity problems,* Computational Optimization and Applications, 5 (1996), pp. 97–138.

[2] J.-S. Chen, X. Chen and P. Tseng, *Analysis of nonsmooth vector-valued functions associated with second-order cones*, Manuscript, Department of Mathematics, University of Washington, December 2002.

[3] X. Chen, *Smoothing methods for complementarity problems and their applications: A survey*, Journal of the Operations Research Society of Japan, 43 (2000), pp. 32–47.

[4] X. Chen, L. Qi and D. Sun, *Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities*, Mathematics of Computation, 67 (1998), pp. 519–540.

[5] X. D. Chen, D. Sun and J. Sun, *Complementarity functions and numerical experiments for second-order-cone complementarity problems*, Technical Report, http://www.math.nus.edu.sg/~mat sundf/, July 2002.

[6] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

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

[8] A. L. Dontchev and T. Zolezzi, *Well-Posed Optimization Problems*, Lecture Notes in Mathematics 1543, Springer-Verlag, Berlin, 1993.

[9] F. Facchinei and C. Kanzow, *Beyond monotonicity in regularization methods for nonlinear complementarity problems*, SIAM Journal on Control and Optimization, 37 (1999), pp. 1150–1161.

[10] J. Faraut and A. Korányi, *Analysis on Symmetric Cones*, Oxford Mathematical Monographs, Oxford University Press, New York, 1994.

[11] 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.

[12] C. Kanzow and H. Pieper, *Jacobian smoothing methods for general complementarity problems*, SIAM Journal on Optimization, 9 (1999), pp. 342–373.

[13] M. S. Lobo, L. Vandenberghe, S. Boyd and H. Lebret, *Applications of second-order cone programming*, Linear Algebra and Its Applications, 284 (1998), pp. 193–228.

[14] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM Journal on Control and Optimization, 15 (1977), pp. 959–972.

[15] R. D. C. Monteiro and T. Tsuchiya, *Polynomial convergence of primal-dual algorithms for the second-order cone program based on the MZ-family of directions*, Mathematical Programming, 88 (2000), pp. 61–83.

[16] R. S. Palais and C.-L. Terng, *Critical Point Theory and Submanifold Geometry*, Lecture Notes in Mathematics 1353, Springer-Verlag, Berlin, 1988.

[17] J. M. Peng and M. Fukushima, *A hybrid Newton method for solving the variational inequality problem via the D-gap function*, Mathematical Programming, 86 (1999), pp. 367–386.

[18] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Mathematics of Operations Research, 18 (1993), pp. 227–244.

[19] L. Qi and X. Chen, *A globally convergent successive approximation method for severely nonsmooth equations*, SIAM Journal on Control and Optimization, 33 (1995), pp. 402–418.

[20] L. Qi and D. Sun, *A nonsmooth version of Newton's method*, Mathematical Programming, 58 (1993), pp. 353–367.

[21] L. Qi and D. Sun, *Smoothing functions and a smoothing Newton method for complementarity and variational inequality problems*, Journal of Optimization Theory and Applications, 113 (2002), pp. 121–147.

[22] T. Tsuchiya, *A convergence analysis of the scaling-invariant primal-dual path-following algorithms for second-order cone programming*, Optimization Methods and Software, 11 & 12 (1999), pp. 141–182.

[23] N. Yamashita and M. Fukushima, *Modified Newton methods for solving a semismooth reformulation of monotone complementarity problems*, Mathematical Programming, 76 (1997), pp. 469–491.

[24] N. Yamashita and M. Fukushima, *On the level-boundedness of the natural residual function for variational inequality problems*, Technical Report 2002-005, Department of Applied Mathematics and Physics, Kyoto University, March 2002.