Math. Program., Ser. B 104, 293–327 (2005)

Digital Object Identifier (DOI) 10.1007/s10107-005-0617-0

Jein-Shan Chen · Paul Tseng

An unconstrained smooth minimization reformulation of the second-order cone complementarity problem

In honor of Terry Rockafellar on his 70th birthday

Received: July 12, 2004 / Accepted: May 25, 2005
Published online: July 14, 2005 – © Springer-Verlag 2005

Abstract. A popular approach to solving the nonlinear complementarity problem (NCP) is to reformulate it as the global minimization of a certain merit function over $\mathbb{R}^n$. A popular choice of the merit function is the squared norm of the Fischer-Burmeister function, shown to be smooth over $\mathbb{R}^n$ and, for monotone NCP, each stationary point is a solution of the NCP. This merit function and its analysis were subsequently extended to the semidefinite complementarity problem (SDCP), although only differentiability, not continuous differentiability, was established. In this paper, we extend this merit function and its analysis, including continuous differentiability, to the second-order cone complementarity problem (SOCCP). Although SOCCP is reducible to a SDCP, the reduction does not allow for easy translation of the analysis from SDCP to SOCCP. Instead, our analysis exploits properties of the Jordan product and spectral factorization associated with the second-order cone. We also report preliminary numerical experience with solving DIMACS second-order cone programs using a limited-memory BFGS method to minimize the merit function.

Key words. Second-order cone – Complementarity – Merit function – Spectral factorization – Jordan product – Level set – Error bound

1. Introduction

We consider the following conic complementarity problem of finding $x, y \in \mathbb{R}^n$ and $\zeta \in \mathbb{R}^n$ satisfying

$$\langle x, y \rangle = 0, \quad x \in \mathcal{K}, \quad y \in \mathcal{K}, \tag{1}$$

$$x = F(\zeta), \quad y = G(\zeta), \tag{2}$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product, $F : \mathbb{R}^n \to \mathbb{R}^n$ and $G : \mathbb{R}^n \to \mathbb{R}^n$ are smooth (i.e., continuously differentiable) mappings, and $\mathcal{K}$ is a closed convex cone in $\mathbb{R}^n$ that is self-dual in the sense that $\mathcal{K}$ equals its dual cone $\mathcal{K}^* := \{y \mid \langle x, y \rangle \geq 0 \ \forall x \in \mathcal{K}\}$. We will focus on the case where $\mathcal{K}$ is the Cartesian product of second-order cones (SOC), also called Lorentz cones [11]. In other words,

$$\mathcal{K} = \mathcal{K}^{n_1} \times \cdots \times \mathcal{K}^{n_N}, \tag{3}$$

J.-S. Chen: Department of Mathematics, National Taiwan Normal University, Taipei 11677, Taiwan.
e-mail: `jschen@math.ntnu.edu.tw`

P. Tseng: Department of Mathematics, University of Washington, Seattle, Washington 98195, USA.
e-mail: `tseng@math.washington.edu`

Mathematics Subject Classification (1991): 26B05, 26B35, 90C33, 65K05

where $N, n_1, \ldots, n_N \geq 1$, $n_1 + \cdots + n_N = n$, and

$$\mathcal{K}^{n_i} := \{(x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n_i - 1} \mid \|x_2\| \leq x_1\},$$

with $\|\cdot\|$ denoting the Euclidean norm and $\mathcal{K}^1$ denoting the set of nonnegative reals $\mathbb{R}_+$. A special case of (3) is $\mathcal{K} = \mathbb{R}^n_+$, the nonnegative orthant in $\mathbb{R}^n$, which corresponds to $N = n$ and $n_1 = \cdots = n_N = 1$. We will refer to (1), (2), (3) as the *second-order cone complementarity problem* (SOCCP).

An important special case of SOCCP corresponds to $G(\zeta) = \zeta$ for all $\zeta \in \mathbb{R}^n$. Then (1) and (2) reduce to

$$\langle F(\zeta), \zeta \rangle = 0, \quad F(\zeta) \in \mathcal{K}, \quad \zeta \in \mathcal{K}. \tag{4}$$

If $\mathcal{K} = \mathbb{R}^n_+$, then (4) reduces to the nonlinear complementarity problem (NCP) and (1)–(2) reduce to the vertical NCP [9]. The NCP plays a fundamental role in optimization theory and has many applications in engineering and economics; see, e.g., [9, 13–15].

Another important special case of SOCCP corresponds to the Karush-Kuhn-Tucker (KKT) optimality conditions for the convex second-order cone program (CSOCP):

$$\begin{aligned} &\text{minimize } g(x) \\ &\text{subject to } Ax = b, \quad x \in \mathcal{K}, \end{aligned} \tag{5}$$

where $A \in \mathbb{R}^{m \times n}$ has full row rank, $b \in \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}$ is a convex twice continuously differentiable function. When g is linear, this reduces to the SOCP which has numerous applications in engineering design, finance, robust optimization, and includes as special cases convex quadratically constrained quadratic programs and linear programs (LP); see [1, 33] and references therein. The KKT optimality conditions for (5), which are sufficient but not necessary for optimality, are (1) and

$$Ax = b, \quad y = \nabla g(x) - A^T \zeta_d \quad \text{for some } \zeta_d \in \mathbb{R}^m.$$

Choose any $d \in \mathbb{R}^n$ satisfying $Ad = b$. (If no such d exists, then (5) has no feasible solution.) Let $B \in \mathbb{R}^{n \times (n-m)}$ be any matrix whose columns span the null space of A. Then x satisfies $Ax = b$ if and only if $x = d + B\zeta_p$ for some $\zeta_p \in \mathbb{R}^{n-m}$. Thus, the KKT optimality conditions can be written in the form of (1) and (2) with

$$\zeta := (\zeta_p, \zeta_d), \quad F(\zeta) := d + B\zeta_p, \quad G(\zeta) := \nabla g(F(\zeta)) - A^T \zeta_d. \tag{6}$$

Alternatively, since any $\zeta \in \mathbb{R}^n$ can be decomposed into the sum of its orthogonal projection onto the column space of A^T and the null space of A,

$$F(\zeta) := d + (I - A^T(AA^T)^{-1}A)\zeta, \quad G(\zeta) := \nabla g(F(\zeta)) - A^T(AA^T)^{-1}A\zeta \tag{7}$$

can also be used in place of (6). For large problems where A is sparse, (7) has the advantage that the main cost of evaluating the Jacobians ∇F and ∇G lies in inverting AA^T, which can be done efficiently via sparse Cholesky factorization. In contrast, (6) entails multiplication by the matrix B, which can be dense.

There have been proposed various methods for solving CSOCP and SOCCP. They include interior-point methods [2, 3, 33, 36, 37, 42, 52], reformulating SOC constraints as smooth convex constraints [4], (non-interior) smoothing Newton methods [6, 19], and smoothing–regularization methods [22]. These methods require solving a nontrivial system of linear equations at each iteration. For the case where $G \equiv I$ and F is affine with ∇F strictly $\mathcal{K}$-copositive, a matrix splitting method has been proposed [21]. In this paper, we study an alternative approach based on reformulating CSOCP and SOCCP as an unconstrained smooth minimization problem. In particular, we aim to find a smooth function $\psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ such that

$$\psi(x, y) = 0 \quad \Longleftrightarrow \quad (x, y) \text{ satisfies (1).} \tag{8}$$

We call such a ψ a *merit function*. Then SOCCP can be expressed as an unconstrained smooth (global) minimization problem:

$$\min_{\zeta \in \mathbb{R}^n} \; f(\zeta) := \psi(F(\zeta), G(\zeta)). \tag{9}$$

Various gradient methods, such as conjugate gradient methods and (limited-memory) quasi-Newton methods [5, 18, 38], can now be applied to solve (9). They have the advantage of requiring less work per iteration than interior-point methods and non-interior Newton methods. This approach can also be combined with smoothing and nonsmooth Newton methods to improve the efficiency and robustness of the latter, as was done in the case of NCP [7, 8, 12, 17, 24, 27, 30]. For this approach to be effective, the choice of ψ is crucial. In the case of NCP, corresponding to (4) and $\mathcal{K} = \mathbb{R}^n_+$, a popular choice is

$$\psi(x, y) \;=\; \frac{1}{2} \sum_{i=1}^{n} \phi(x_i, y_i)^2$$

for all $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$, where ϕ is the well-known Fischer-Burmeister (FB) NCP-function [16, 17] defined by

$$\phi(x_i, y_i) = \sqrt{x_i^2 + y_i^2} - x_i - y_i.$$

It has been shown that ψ is smooth (even though ϕ is not differentiable) and satisfies (8) [10, 25, 26]. Moreover, when F is monotone or, more generally, a P_0-function, every stationary point of $\zeta \mapsto \psi(F(\zeta), \zeta)$ is a solution of NCP [10, 20]. This is an important property since (i) gradient methods are guaranteed to find stationary points only, and (ii) when an LP is reformulated as an NCP, the resulting F is monotone, but neither strongly monotone nor a uniformly P-function. In contrast, other smooth merit functions for NCP, such as the implicit Lagrangian and the D-gap function [28, 35, 40, 45, 51, 54], require F to be a uniformly P-function in order for stationary points to be solutions of NCP. Thus these other merit functions cannot be used for LP. Subsequently, a number of variants of ψ with additional desirable properties have been proposed, e.g., [6, 10, 29, 31, 34, 41, 47, 49, 53]. A recent discussion of these variants can be found in the paper [47]. Moreover, the above merit function ψ, as well as a related merit function of Yamashita and Fukushima [53], have been extended to the semidefinite complementarity problem (SDCP), which has the form (1), (2), but with x, y being $q \times q$ ($q \geq 1$) real symmetric block-diagonal matrices of fixed block sizes, $\langle \cdot, \cdot \rangle$ being the trace inner product, and $\mathcal{K}$ being the cone of $q \times q$ block-diagonal positive semidefinite matrices of fixed block sizes [50, 53]. However, the analysis in [50] showed ψ to be differentiable, but did not show it to be smooth.[1]

Can the above merit functions for NCP be extended to SOCCP? To our knowledge, this question has not been studied previously. We study it in this paper. We are motivated by previous work on extending merit function from NCP to SDCP [50, 53]. We are further motivated by a recent work [19] showing that the FB function extends from NCP to SOCCP using the Jordan product associated with SOC [11]. Nice properties of the FB function, such as strong semismoothness, are preserved when extended to SOCCP [48]. More specifically, for any $x = (x_1, x_2)$, $y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, we define their *Jordan product* associated with $\mathcal{K}^n$ as

$$x \cdot y \; := \; (\langle x, y \rangle, \; y_1 x_2 + x_1 y_2). \tag{10}$$

The identity element under this product is $e := (1, 0, \ldots, 0)^T \in \mathbb{R}^n$. We write x^2 to mean $x \cdot x$ and write $x + y$ to mean the usual componentwise addition of vectors. It is known that $x^2 \in \mathcal{K}^n$ for all $x \in \mathbb{R}^n$. Moreover, if $x \in \mathcal{K}^n$, then there exists a unique vector in $\mathcal{K}^n$, denoted by $x^{1/2}$, such that $(x^{1/2})^2 = x^{1/2} \cdot x^{1/2} = x$. Then,

$$\phi(x, y) := (x^2 + y^2)^{1/2} - x - y \tag{11}$$

is well defined for all $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ and maps $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^n$. It was shown in [19] that $\phi(x, y) = 0$ if and only if (x, y) satisfies (1). Thus,

$$\psi_{\mathrm{FB}}(x, y) := \frac{1}{2} \sum_{i=1}^{N} \|\phi(x_i, y_i)\|^2, \tag{12}$$

where $x = (x_1, \ldots, x_N)^T$, $y = (y_1, \ldots, y_N)^T \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_N}$, is a merit function for SOCCP. We will show that, like the NCP case, ψ_{FB} is smooth and, when ∇F and $-\nabla G$ are column monotone, every stationary point of (9) solves SOCCP; see Propositions 2 and 3. The same holds for the following analog of the SDCP merit function studied by Yamashita and Fukushima [53]:

$$\psi_{\mathrm{YF}}(x, y) := \psi_0(\langle x, y \rangle) + \psi_{\mathrm{FB}}(x, y), \tag{13}$$

where $\psi_0 : \mathbb{R} \to [0, \infty)$ is any smooth function satisfying

$$\psi_0(t) = 0 \;\; \forall t \leq 0 \qquad \text{and} \qquad \psi_0'(t) > 0 \;\; \forall t > 0; \tag{14}$$

see Proposition 4. In [53], $\psi_0(t) = \frac{1}{4}(\max\{0, t\})^4$ was considered. Analogous to the NCP and SDCP cases, when $\nabla G(\zeta)$ is invertible, a ∇F-free descent direction for

$$f_{\mathrm{FB}}(\zeta) := \psi_{\mathrm{FB}}(F(\zeta), G(\zeta)) \tag{15}$$

and

$$f_{\mathrm{YF}}(\zeta) := \psi_{\mathrm{YF}}(F(\zeta), G(\zeta)) \tag{16}$$

[1] During the revising of this paper, a proof of smoothness is reported in [43].

can be found. The function f_{YF}, compared to f_{FB}, has additional bounded level-set and error bound properties; see Section 5. Our proof of the smoothness of ψ_{FB} in Section 3 is quite technical, but further simplification seems difficult. In particular, neither general properties of the Jordan product associated with symmetric cones [11] nor the strong semismoothness proof for ϕ given in [48] lend themselves readily to a smoothness proof for ψ_{FB}. In Section 6, we report our numerical experience with solving SOCP (5) from the DIMACS library by using a limited-memory BFGS (L-BFGS) method to minimize f_{FB}, with F and G given by (7). On problems with $n \gg m$ and for low-to-medium solution accuracy, L-BFGS appears to be competitive with interior-point methods. We also report our experience with solving CSOCP using a BFGS method to minimize f_{FB}.

It is known that SOCCP can be reduced to an SDCP by observing that, for any $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, we have $x \in \mathcal{K}^n$ if and only if

$$L_x := \begin{bmatrix} x_1 & x_2^T \\ x_2 & x_1 I \end{bmatrix}$$

is positive semidefinite (also see [19, p. 437] and [44]). However, this reduction increases the problem dimension from n to $n(n+1)/2$ and it is not known whether this increase can be mitigated by exploiting the special "arrow" structure of L_x.

Throughout this paper, $\mathbb{R}^n$ denotes the space of n-dimensional real column vectors and ^{T} denotes transpose. For any differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, $\nabla f(x)$ denotes the gradient of f at x. For any differentiable mapping $F = (F_1, ..., F_m)^T : \mathbb{R}^n \to \mathbb{R}^m$, $\nabla F(x) = [\nabla F_1(x) \ \cdots \ \nabla F_m(x)]$ denotes the transpose Jacobian of F at x. For any symmetric matrices $A, B \in \mathbb{R}^{n\times n}$, we write $A \succeq B$ (respectively, $A \succ B$) to mean $A - B$ is positive semidefinite (respectively, positive definite). For nonnegative scalars α and β, we write $\alpha = O(\beta)$ to mean $\alpha \le C\beta$, with C independent of α and β.

2. Jordan product and spectral factorization

It is known that $\mathcal{K}^n$ is a closed convex self-dual cone with nonempty interior given by

$$\text{int}(\mathcal{K}^n) = \{(x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid \|x_2\| < x_1\}.$$

The Jordan product (10), unlike scalar or matrix multiplication, is not associative, which is a main source of complication in the analysis of SOCCP. For any $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, its *determinant* is defined by

$$det(x) := x_1^2 - \|x_2\|^2.$$

In general, $det(x \cdot y) \neq det(x)det(y)$ unless $x_2 = y_2$.

We next recall from [19] that each $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ admits a spectral factorization, associated with $\mathcal{K}^n$, of the form

$$x = \lambda_1 u^{(1)} + \lambda_2 u^{(2)},$$

where λ_1, λ_2 and $u^{(1)}, u^{(2)}$ are the spectral values and the associated spectral vectors of x given by

$$\lambda_i = x_1 + (-1)^i \|x_2\|,$$
$$u^{(i)} = \begin{cases} \frac{1}{2}\left(1,\ (-1)^i \dfrac{x_2}{\|x_2\|}\right) & \text{if } x_2 \neq 0; \\ \frac{1}{2}\left(1,\ (-1)^i w_2\right) & \text{if } x_2 = 0, \end{cases}$$

for $i = 1, 2$, with w_2 being any vector in $\mathbb{R}^{n-1}$ satisfying $\|w_2\| = 1$. If $x_2 \neq 0$, the factorization is unique.

The above spectral factorization of x, as well as x^2 and $x^{1/2}$ and the matrix L_x, have various interesting properties; see [19]. We list four properties that we will use later.

Property 1. For any $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, with spectral values λ_1, λ_2 and spectral vectors $u^{(1)}, u^{(2)}$, the following results hold.

(a) $x^2 = \lambda_1^2 u^{(1)} + \lambda_2^2 u^{(2)} \in \mathcal{K}^n$.
(b) If $x \in \mathcal{K}^n$, then $0 \le \lambda_1 \le \lambda_2$ and $x^{1/2} = \sqrt{\lambda_1}\, u^{(1)} + \sqrt{\lambda_2}\, u^{(2)}$.
(c) If $x \in \mathrm{int}(\mathcal{K}^n)$, then $0 < \lambda_1 \le \lambda_2$, $det(x) = \lambda_1\lambda_2$, and L_x is invertible with

$$L_x^{-1} = \frac{1}{det(x)} \begin{bmatrix} x_1 & -x_2^T \\ -x_2 & \dfrac{det(x)}{x_1} I + \dfrac{1}{x_1} x_2 x_2^T \end{bmatrix}.$$

(d) $x \cdot y = L_x y$ for all $y \in \mathbb{R}^n$, and $L_x \succ 0$ if and only if $x \in \mathrm{int}(\mathcal{K}^n)$.

3. Smoothness property of merit functions

In this section we show that the functions (12) and (13) are smooth functions satisfying (8). For simplicity, we focus on the special case of $N = 1$, i.e.,

$$\psi_{\mathrm{FB}}(x, y) = \frac{1}{2}\|\phi(x, y)\|^2 \tag{17}$$

in this and the next two sections. Extension of our analyses to the general case of $N \ge 1$ is straightforward. We begin with the following result from [19] showing that the FB function ϕ given by (11) has property analogous to the NCP and SDCP cases. Additional properties of ϕ are studied in [19, 48].

Lemma 1. ([19, Proposition 2.1]) *Let* $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ *be given by* (11). *Then*

$$\begin{aligned} \phi(x, y) = 0 &\iff x, y \in \mathcal{K}^n,\ x \cdot y = 0, \\ &\iff x, y \in \mathcal{K}^n,\ \langle x, y\rangle = 0. \end{aligned}$$

Since $x^2, y^2 \in \mathcal{K}^n$ for any $x, y \in \mathbb{R}^n$, we have that $x^2 + y^2 = (\|x\|^2 + \|y\|^2, 2x_1x_2 + 2y_1y_2) \in \mathcal{K}^n$. Thus

$$x^2 + y^2 \notin \mathrm{int}(\mathcal{K}^n) \quad \iff \quad \|x\|^2 + \|y\|^2 = 2\|x_1x_2 + y_1y_2\|. \tag{18}$$

The spectral values of $x^2 + y^2$ are

$$\begin{aligned} \lambda_1 &:= \|x\|^2 + \|y\|^2 - 2\|x_1x_2 + y_1y_2\|, \\ \lambda_2 &:= \|x\|^2 + \|y\|^2 + 2\|x_1x_2 + y_1y_2\|. \end{aligned} \tag{19}$$

Then, by Property 1(b), $z := (x^2 + y^2)^{1/2}$ has the spectral values $\sqrt{\lambda_1}$, $\sqrt{\lambda_2}$ and

$$z = (z_1, z_2) = \left(\frac{\sqrt{\lambda_1} + \sqrt{\lambda_2}}{2}, \ \frac{\sqrt{\lambda_2} - \sqrt{\lambda_1}}{2} w_2 \right), \tag{20}$$

where $w_2 := \dfrac{x_1 x_2 + y_1 y_2}{\|x_1 x_2 + y_1 y_2\|}$ if $x_1 x_2 + y_1 y_2 \neq 0$ and otherwise w_2 is any vector in $\mathbb{R}^{n-1}$ satisfying $\|w_2\| = 1$. The next key lemma, describing special properties of x, y with $x^2 + y^2 \notin \text{int}(\mathcal{K}^n)$, will be used to prove Propositions 1, 2, and Lemma 6.

Lemma 2. *For any* $x = (x_1, x_2)$, $y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ *with* $x^2 + y^2 \notin \text{int}(\mathcal{K}^n)$, *we have*

$$\begin{aligned} x_1^2 &= \|x_2\|^2, \\ y_1^2 &= \|y_2\|^2, \\ x_1 y_1 &= x_2^T y_2, \\ x_1 y_2 &= y_1 x_2. \end{aligned}$$

Proof. By (18), $\|x\|^2 + \|y\|^2 = 2\|x_1 x_2 + y_1 y_2\|$. Thus $\left(\|x\|^2 + \|y\|^2 \right)^2 = 4\|x_1 x_2 + y_1 y_2\|^2$, so that

$$\|x\|^4 + 2\|x\|^2\|y\|^2 + \|y\|^4 = 4(x_1 x_2 + y_1 y_2)^T (x_1 x_2 + y_1 y_2).$$

Notice that $\|x\|^2 = x_1^2 + \|x_2\|^2$ and $\|y\|^2 = y_1^2 + \|y_2\|^2$. Thus,

$$\left(x_1^2 + \|x_2\|^2 \right)^2 + 2\|x\|^2\|y\|^2 + \left(y_1^2 + \|y_2\|^2 \right)^2 = 4x_1^2\|x_2\|^2 + 8x_1 y_1 x_2^T y_2 + 4y_1^2\|y_2\|^2.$$

Simplifying the above expression yields

$$\left(x_1^2 - \|x_2\|^2 \right)^2 + \left(y_1^2 - \|y_2\|^2 \right)^2 + \left(2\|x\|^2\|y\|^2 - 8x_1 y_1 x_2^T y_2 \right) = 0.$$

The first two terms are nonnegative. The third term is also nonnegative because

$$\begin{aligned} \|x\|^2\|y\|^2 &= \left(x_1^2 + \|x_2\|^2 \right)\left(y_1^2 + \|y_2\|^2 \right) \\ &\geq \left(2|x_1|\|x_2\| \right)\left(2|y_1|\|y_2\| \right) \\ &= 4|x_1||y_1|\|x_2\|\|y_2\| \\ &\geq 4x_1 y_1 x_2^T y_2. \end{aligned}$$

Hence

$$x_1^2 = \|x_2\|^2, \qquad y_1^2 = \|y_2\|^2, \qquad 2\|x\|^2\|y\|^2 - 8x_1 y_1 x_2^T y_2 = 0.$$

Substituting $x_1^2 = \|x_2\|^2$ and $y_1^2 = \|y_2\|^2$ into the last equation, the resulting three equations imply $x_1 y_1 = x_2^T y_2$.

It remains to prove that $x_1y_2 = y_1x_2$. If $x_1 = 0$, then $\|x_2\| = |x_1| = 0$ so this relation is true. Symmetrically, if $y_1 = 0$, then this relation is also true. Suppose that $x_1 \neq 0$ and $y_1 \neq 0$. Then $x_2 \neq 0$, $y_2 \neq 0$, and

$$x_1y_1 = x_2^T y_2 = \|x_2\|\|y_2\|\cos\theta = |x_1||y_1|\cos\theta,$$

where θ is the angle between x_2 and y_2. Thus, $\cos\theta \in \{-1, 1\}$, i.e., $y_2 = \alpha x_2$ for some $\alpha \neq 0$. Then

$$x_1y_1 = x_2^T y_2 = \alpha\|x_2\|^2 = \alpha x_1^2,$$

so that $y_1/x_1 = \alpha$. Thus $y_2 = x_2y_1/x_1$.

The next technical lemma shows that two square terms are upper bounded by a quantity that measures how close $x^2 + y^2$ comes to the boundary of $\mathcal{K}^n$ (cf. (18)). This lemma will be used to prove Lemma 4 and Proposition 2.

Lemma 3. *For any $x = (x_1, x_2)$, $y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ with $x_1x_2 + y_1y_2 \neq 0$, we have*

$$\begin{aligned}\left(x_1 - \frac{(x_1x_2 + y_1y_2)^T x_2}{\|x_1x_2 + y_1y_2\|}\right)^2 &\leq \left\| x_2 - x_1\frac{x_1x_2 + y_1y_2}{\|x_1x_2 + y_1y_2\|}\right\|^2 \\ &\leq \|x\|^2 + \|y\|^2 - 2\|x_1x_2 + y_1y_2\|.\end{aligned}$$

Proof. The first inequality can be seen by expanding the square on both sides and using the Cauchy-Schwarz inequality. It remains to prove the second inequality. Let us multiply both sides of this inequality by

$$\|x_1x_2 + y_1y_2\|^2 = x_1^2\|x_2\|^2 + 2x_1y_1x_2^T y_2 + y_1^2\|y_2\|^2$$

and let L and R denote, respectively, the left-hand side and the right-hand side. Since $x_1x_2 + y_1y_2 \neq 0$, the second inequality is equivalent to $R - L \geq 0$. We have

$$\begin{aligned} L &= \left(\|x_2\|^2 - 2x_1\frac{(x_1x_2 + y_1y_2)^T x_2}{\|x_1x_2 + y_1y_2\|} + x_1^2\right)\|x_1x_2 + y_1y_2\|^2 \\ &= \|x_2\|^2\left(x_1^2\|x_2\|^2 + 2x_1y_1x_2^T y_2 + y_1^2\|y_2\|^2\right) \\ &\quad -2x_1\left(x_1\|x_2\|^2 + y_1x_2^T y_2\right)\|x_1x_2 + y_1y_2\| \\ &\quad +x_1^2\left(x_1^2\|x_2\|^2 + 2x_1y_1x_2^T y_2 + y_1^2\|y_2\|^2\right) \\ &= x_1^2\|x_2\|^4 + 2x_1y_1x_2^T y_2\|x_2\|^2 + y_1^2\|x_2\|^2\|y_2\|^2 \\ &\quad -2x_1^2\|x_2\|^2\|x_1x_2 + y_1y_2\| - 2x_1y_1x_2^T y_2\|x_1x_2 + y_1y_2\| \\ &\quad +x_1^4\|x_2\|^2 + 2x_1^3y_1x_2^T y_2 + x_1^2y_1^2\|y_2\|^2, \end{aligned}$$

and

$$
\begin{aligned}
R &= \left(\|x\|^2 + \|y\|^2 - 2\|x_1x_2 + y_1y_2\|\right)\|x_1x_2 + y_1y_2\|^2 \\
&= \left(x_1^2 + \|x_2\|^2 - 2\|x_1x_2 + y_1y_2\|\right)\|x_1x_2 + y_1y_2\|^2 + \|y\|^2\|x_1x_2 + y_1y_2\|^2 \\
&= \left(x_1^2 + \|x_2\|^2 - 2\|x_1x_2 + y_1y_2\|\right)\left(x_1^2\|x_2\|^2 + 2x_1y_1x_2^Ty_2 + y_1^2\|y_2\|^2\right) \\
&\quad + \|y\|^2\|x_1x_2 + y_1y_2\|^2 \\
&= x_1^4\|x_2\|^2 + 2x_1^3y_1x_2^Ty_2 + x_1^2y_1^2\|y_2\|^2 + x_1^2\|x_2\|^4 + 2x_1y_1x_2^Ty_2\|x_2\|^2 \\
&\quad + y_1^2\|x_2\|^2\|y_2\|^2 - 2x_1^2\|x_2\|^2\|x_1x_2 + y_1y_2\| - 4x_1y_1x_2^Ty_2\|x_1x_2 + y_1y_2\| \\
&\quad - 2y_1^2\|y_2\|^2\|x_1x_2 + y_1y_2\| + \|y\|^2\|x_1x_2 + y_1y_2\|^2.
\end{aligned}
$$

Thus, taking the difference and using the Cauchy-Schwarz inequality yields

$$
\begin{aligned}
R - L &= \|y\|^2\|x_1x_2 + y_1y_2\|^2 - 2x_1y_1x_2^Ty_2\|x_1x_2 + y_1y_2\| - 2y_1^2\|y_2\|^2\|x_1x_2 + y_1y_2\| \\
&= y_1^2\|x_1x_2 + y_1y_2\|^2 + \|y_2\|^2\|x_1x_2 + y_1y_2\|^2 \\
&\quad - 2y_1y_2^T(x_1x_2 + y_1y_2)\|x_1x_2 + y_1y_2\| \\
&\geq y_1^2\|x_1x_2 + y_1y_2\|^2 + \|y_2\|^2\|x_1x_2 + y_1y_2\|^2 - 2|y_1|\|y_2\|\,\|x_1x_2 + y_1y_2\|^2 \\
&= \left(|y_1| - \|y_2\|\right)^2\|x_1x_2 + y_1y_2\|^2 \\
&\geq 0.
\end{aligned}
$$

Using Lemmas 1, 2, 3, and [19, Proposition 5.2], we prove our first main result showing that ψ_{FB} is differentiable and its gradient has a computable formula.

Proposition 1. *Let ϕ be given by* (11). *Then ψ_{FB} given by* (17) *has the following properties.*

(a) $\psi_{\text{FB}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ *and satisfies* (8).
(b) ψ_{FB} *is differentiable at every* $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$. *Moreover,* $\nabla_x\psi_{\text{FB}}(0, 0) = \nabla_y\psi_{\text{FB}}(0, 0) = 0$. *If* $(x, y) \neq (0, 0)$ *and* $x^2 + y^2 \in \text{int}(\mathcal{K}^n)$, *then*

$$
\begin{aligned}
\nabla_x\psi_{\text{FB}}(x, y) &= \left(L_xL^{-1}_{(x^2+y^2)^{1/2}} - I\right)\phi(x, y), \\
\nabla_y\psi_{\text{FB}}(x, y) &= \left(L_yL^{-1}_{(x^2+y^2)^{1/2}} - I\right)\phi(x, y). \qquad (21)
\end{aligned}
$$

If $(x, y) \neq (0, 0)$ *and* $x^2 + y^2 \notin \text{int}(\mathcal{K}^n)$, *then* $x_1^2 + y_1^2 \neq 0$ *and*

$$
\nabla_x\psi_{\text{FB}}(x, y) = \left(\frac{x_1}{\sqrt{x_1^2 + y_1^2}} - 1\right)\phi(x, y), \qquad (22)
$$

$$
\nabla_y\psi_{\text{FB}}(x, y) = \left(\frac{y_1}{\sqrt{x_1^2 + y_1^2}} - 1\right)\phi(x, y). \qquad (23)
$$

Proof. (a) This follows from Lemma 1.

(b) Case (1): $x = y = 0$.
For any $h, k \in \mathbb{R}^n$, let $\mu_1 \le \mu_2$ be the spectral values and let $v^{(1)}, v^{(2)}$ be the corresponding spectral vectors of $h^2 + k^2$. Then, by Property 1(b),

$$\begin{aligned}\|(h^2+k^2)^{1/2} - h - k\| &= \|\sqrt{\mu_1}v^{(1)} + \sqrt{\mu_2}v^{(2)} - h - k\| \\ &\le \sqrt{\mu_1}\|v^{(1)}\| + \sqrt{\mu_2}\|v^{(2)}\| + \|h\| + \|k\| \\ &= (\sqrt{\mu_1} + \sqrt{\mu_2})/\sqrt{2} + \|h\| + \|k\|.\end{aligned}$$

Also

$$\begin{aligned}\mu_1 \ \le \ \mu_2 &= \|h\|^2 + \|k\|^2 + 2\|h_1h_2 + k_1k_2\| \\ &\le \|h\|^2 + \|k\|^2 + 2|h_1|\|h_2\| + 2|k_1|\|k_2\| \\ &\le 2\|h\|^2 + 2\|k\|^2.\end{aligned}$$

Combining the above two inequalities yields

$$\begin{aligned}\psi_{\mathrm{FB}}(h,k) - \psi_{\mathrm{FB}}(0,0) &= \|(h^2+k^2)^{1/2} - h - k\|^2 \\ &\le \Big((\sqrt{\mu_1} + \sqrt{\mu_2})/\sqrt{2} + \|h\| + \|k\|\Big)^2 \\ &\le \Big(2\sqrt{2\|h\|^2 + 2\|k\|^2}/\sqrt{2} + \|h\| + \|k\|\Big)^2 \\ &= O(\|h\|^2 + \|k\|^2).\end{aligned}$$

This shows that ψ_{FB} is differentiable at $(0, 0)$ with

$$\nabla_x \psi_{\mathrm{FB}}(0,0) = \nabla_y \psi_{\mathrm{FB}}(0,0) = 0.$$

Case (2): $(x, y) \ne (0, 0)$ and $x^2 + y^2 \in \mathrm{int}(\mathcal{K}^n)$.

Since $x^2 + y^2 \in \mathrm{int}(\mathcal{K}^n)$, Proposition 5.2 of [19] implies that ϕ is continuously differentiable at (x, y). Since ψ_{FB} is the composition of ϕ with $x \mapsto \frac{1}{2}\|x\|^2$, then ψ_{FB} is continuously differentiable at (x, y). The expressions (21) for $\nabla_x \psi_{\mathrm{FB}}(x, y)$ and $\nabla_y \psi_{\mathrm{FB}}(x, y)$ follow from the chain rule for differentiation and the expression for the Jacobian of ϕ given in [19, Proposition 5.2] (also see [19, Corollary 5.4]).

Case (3): $(x, y) \ne (0, 0)$ and $x^2 + y^2 \notin \mathrm{int}(\mathcal{K}^n)$.
By (18), $\|x\|^2 + \|y\|^2 = 2\|x_1x_2 + y_1y_2\|$. Since $(x, y) \ne (0, 0)$, this also implies $x_1x_2 + y_1y_2 \ne 0$, so Lemmas 2 and 3 are applicable. By (20),

$$(x^2+y^2)^{1/2} = \left(\frac{\sqrt{\lambda_1} + \sqrt{\lambda_2}}{2},\ \frac{\sqrt{\lambda_2} - \sqrt{\lambda_1}}{2} w_2\right),$$

where λ_1, λ_2 are given by (19) and $w_2 := \dfrac{x_1x_2 + y_1y_2}{\|x_1x_2 + y_1y_2\|}$. Thus $\lambda_1 = 0$ and $\lambda_2 > 0$. Since $x_1x_2 + y_1y_2 \ne 0$, we have $x_1'x_2' + y_1'y_2' \ne 0$ for all $(x', y') \in \mathbb{R}^n \times \mathbb{R}^n$ sufficiently near to (x, y). Moreover,

$$
\begin{aligned}
2\psi_{\mathrm{FB}}(x', y') &= \|(x'^2 + y'^2)^{1/2} - x' - y'\|^2 \\
&= \|(x'^2 + y'^2)^{1/2}\|^2 + \|x' + y'\|^2 - 2\langle (x'^2 + y'^2)^{1/2}, x' + y'\rangle \\
&= \|x'\|^2 + \|y'\|^2 + \|x' + y'\|^2 - 2\langle (x'^2 + y'^2)^{1/2}, x' + y'\rangle,
\end{aligned}
$$

where the third equality uses the observation that $\|z\|^2 = \langle z^2, e\rangle$ for any $z \in \mathbb{R}^n$. Since $\|x'\|^2 + \|y'\|^2 + \|x' + y'\|^2$ is clearly differentiable in (x', y'), it suffices to show that

$$
\begin{aligned}
&2\langle (x'^2 + y'^2)^{1/2}, x' + y'\rangle \\
&= (\sqrt{\mu_2} + \sqrt{\mu_1})(x'_1 + y'_1) + (\sqrt{\mu_2} - \sqrt{\mu_1}) \frac{(x'_1 x'_2 + y'_1 y'_2)^T (x'_2 + y'_2)}{\|x'_1 x'_2 + y'_1 y'_2\|} \\
&= \sqrt{\mu_2} \left(x'_1 + y'_1 + \frac{(x'_1 x'_2 + y'_1 y'_2)^T (x'_2 + y'_2)}{\|x'_1 x'_2 + y'_1 y'_2\|} \right) \\
&\quad + \sqrt{\mu_1} \left(x'_1 + y'_1 - \frac{(x'_1 x'_2 + y'_1 y'_2)^T (x'_2 + y'_2)}{\|x'_1 x'_2 + y'_1 y'_2\|} \right)
\end{aligned}
\tag{24}
$$

is differentiable at $(x', y') = (x, y)$, where μ_1, μ_2 are the spectral values of $x'^2 + y'^2$, i.e., $\mu_i = \|x'\|^2 + \|y'\|^2 + 2(-1)^i \|x'_1 x'_2 + y'_1 y'_2\|$. Since $\lambda_2 > 0$, we see that the first term on the right-hand side of (24) is differentiable at $(x', y') = (x, y)$. We claim that the second term on the right-hand side of (24) is $o(\|h\| + \|k\|)$ with $h := x' - x$, $k := y' - y$, i.e., it is differentiable with zero gradient. To see this, notice that $x_1 x_2 + y_1 y_2 \neq 0$, so that $\mu_1 = \|x'\|^2 + \|y'\|^2 - 2\|x'_1 x'_2 + y'_1 y'_2\|$, viewed as a function of (x', y'), is differentiable at $(x', y') = (x, y)$. Moreover, $\mu_1 = \lambda_1 = 0$ when $(x', y') = (x, y)$. Thus, first-order Taylor's expansion of μ_1 at (x, y) yields

$$
\mu_1 \;=\; O(\|x' - x\| + \|y' - y\|) \;=\; O(\|h\| + \|k\|).
$$

Also, since $x_1 x_2 + y_1 y_2 \neq 0$, by the product and quotient rules for differentiation, the function

$$
x'_1 + y'_1 - \frac{(x'_1 x'_2 + y'_1 y'_2)^T (x'_2 + y'_2)}{\|x'_1 x'_2 + y'_1 y'_2\|}
\tag{25}
$$

is differentiable at $(x', y') = (x, y)$. Moreover, the function (25) has value 0 at $(x', y') = (x, y)$. This is because

$$
x_1 + y_1 - \frac{(x_1 x_2 + y_1 y_2)^T (x_2 + y_2)}{\|x_1 x_2 + y_1 y_2\|} = x_1 - w_2^T x_2 \;+\; y_1 - w_2^T y_2 \;=\; 0 + 0,
$$

where $w_2 := (x_1 x_2 + y_1 y_2)/\|x_1 x_2 + y_1 y_2\|$ and the last equality uses the fact that, by Lemma 3 and $\|x\|^2 + \|y\|^2 = 2\|x_1 x_2 + y_1 y_2\|$, we have $w_2^T x_2 = x_1$, $w_2^T y_2 = y_1$. (By symmetry, Lemma 3 still holds when x and y are switched.) Thus, the function (25) is $O(\|h\| + \|k\|)$ in magnitude. This together with $\mu_1 = O(\|h\| + \|k\|)$ shows that the second term on the right of (24) is $O((\|h\| + \|k\|)^{3/2}) = o(\|h\| + \|k\|)$.

Thus, we have shown that ψ_{FB} is differentiable at (x, y). Moreover, the preceding argument shows that $2\nabla\psi_{\mathrm{FB}}(x, y)$ is the sum of the gradient of $\|x'\|^2 + \|y'\|^2 + \|x' + y'\|^2$

and the gradient of the first term on the right of (24), evaluated at $(x', y') = (x, y)$. The gradient of $\|x'\|^2 + \|y'\|^2 + \|x' + y'\|^2$ with respect to x', evaluated at $(x', y') = (x, y)$, is $4x + 2y$. Using the product and quotient rules for differentiation, the gradient of the first term on the right of (24) with respect to x_1', evaluated at $(x', y') = (x, y)$, works out to be

$$\begin{aligned} &\frac{x_1 + w_2^T x_2}{\sqrt{\lambda_2}} \left(x_1 + y_1 + w_2^T (x_2 + y_2)\right) \\ &\quad + \sqrt{\lambda_2} \left(1 + \frac{x_2^T (x_2 + y_2)}{\|x_1 x_2 + y_1 y_2\|} - \frac{w_2^T (x_2 + y_2)}{\|x_1 x_2 + y_1 y_2\|} w_2^T x_2\right) \\ &= \frac{2x_1(x_1 + y_1)}{\sqrt{x_1^2 + y_1^2}} + 2\sqrt{x_1^2 + y_1^2}, \end{aligned}$$

where $w_2 := (x_1 x_2 + y_1 y_2)/\|x_1 x_2 + y_1 y_2\|$ and the equality uses Lemma 2 and the fact that, by Lemma 3 and $\|x\|^2 + \|y\|^2 = 2\|x_1 x_2 + y_1 y_2\|$, we have $w_2^T x_2 = x_1$, $w_2^T y_2 = y_1$. Similarly, the gradient of the first term on the right of (24) with respect to x_2', evaluated at $(x', y') = (x, y)$, works out to be

$$\begin{aligned} &\frac{x_2 + w_2 x_1}{\sqrt{\lambda_2}} \left(x_1 + y_1 + w_2^T (x_2 + y_2)\right) \\ &\quad + \sqrt{\lambda_2} \left(\frac{2x_1 x_2 + (x_1 + y_1) y_2}{\|x_1 x_2 + y_1 y_2\|} - \frac{w_2^T (x_2 + y_2)}{\|x_1 x_2 + y_1 y_2\|} w_2 x_1\right) \\ &= 2\, \frac{2x_1 x_2 + (x_1 + y_1) y_2}{\sqrt{x_1^2 + y_1^2}}. \end{aligned}$$

In particular, the equality uses the fact that, by Lemma 2, we have $x_1 y_2 = y_1 x_2$ and $\|x_1 x_2 + y_1 y_2\| = x_1^2 + y_1^2$, so that $w_2 x_1 = x_2$ and $\lambda_2 = 4(x_1^2 + y_1^2)$. Thus, combining the preceding gradient expressions yields

$$2\nabla_x \psi_{\mathrm{FB}}(x, y) = 4x + 2y - \begin{bmatrix} 2\sqrt{x_1^2 + y_1^2} \\ 0 \end{bmatrix} - \frac{2}{\sqrt{x_1^2 + y_1^2}} \begin{bmatrix} x_1(x_1 + y_1) \\ 2x_1 x_2 + (x_1 + y_1) y_2 \end{bmatrix}.$$

Using $\|x_1 x_2 + y_1 y_2\| = x_1^2 + y_1^2$ and $\lambda_2 = 4(x_1^2 + y_1^2)$, we can also write

$$(x^2 + y^2)^{1/2} = \left(\sqrt{x_1^2 + y_1^2},\ \frac{x_1 x_2 + y_1 y_2}{\sqrt{x_1^2 + y_1^2}}\right),$$

so that

$$\phi(x, y) = \left(\sqrt{x_1^2 + y_1^2} - (x_1 + y_1),\ \frac{x_1 x_2 + y_1 y_2}{\sqrt{x_1^2 + y_1^2}} - (x_2 + y_2)\right). \tag{26}$$

Using the fact that $x_1 y_2 = y_1 x_2$, we can rewrite the above expression for $\nabla_x \psi_{\mathrm{FB}}(x, y)$ in the form of (22). By symmetry, (23) also holds.

Proposition 1 gives a formula for $\nabla\psi_{\rm FB}$ when $\psi_{\rm FB}$ is given by (11), (12). Using Lemma 3, we have the following lemma on the uniform boundedness of the matrices in (21). This will be used to prove the smoothness of $\psi_{\rm FB}$ at $(0,0)$.

Lemma 4. *There exists a scalar constant $C>0$ such that*

$$\|L_x L^{-1}_{(x^2+y^2)^{1/2}}\|_F \le C, \qquad \|L_y L^{-1}_{(x^2+y^2)^{1/2}}\|_F \le C$$

for all $(x,y)\neq(0,0)$ satisfying $x^2+y^2\in \mathrm{int}(\mathcal{K}^n)$. ($\|A\|_F$ denotes the Frobenius norm of $A\in\mathbb{R}^{n\times n}$.)

Proof. Consider any $(x,y)\neq(0,0)$ satisfying $x^2+y^2\in\mathrm{int}(\mathcal{K}^n)$. Let λ_1,λ_2 be the spectral values of x^2+y^2 and let $z:=(x^2+y^2)^{1/2}$. Then, z is given by (20), i.e.,

$$z_1=\frac{\sqrt{\lambda_1}+\sqrt{\lambda_2}}{2}, \qquad z_2=\frac{\sqrt{\lambda_2}-\sqrt{\lambda_1}}{2}w_2,$$

with λ_1,λ_2 given by (19), and $w_2:=\dfrac{x_1x_2+y_1y_2}{\|x_1x_2+y_1y_2\|}$ if $x_1x_2+y_1y_2\neq 0$; otherwise w_2 is any vector satisfying $\|w_2\|=1$. Using Property 1(c), we have that

$$\begin{aligned}
&L_xL_z^{-1}\\
&=\frac{1}{det(z)}\begin{bmatrix} x_1z_1-x_2^Tz_2 & -x_1z_2^T+\frac{det(z)}{z_1}x_2^T+\frac{x_2^Tz_2}{z_1}z_2^T\\ x_2z_1-x_1z_2 & -x_2z_2^T+\frac{x_1det(z)}{z_1}I+\frac{x_1}{z_1}z_2z_2^T\end{bmatrix}\\
&=\frac{1}{\sqrt{\lambda_1}\sqrt{\lambda_2}}\begin{bmatrix} \frac{\sqrt{\lambda_1}+\sqrt{\lambda_2}}{2}x_1+\frac{\sqrt{\lambda_1}-\sqrt{\lambda_2}}{2}x_2^Tw_2 & \frac{\sqrt{\lambda_1}-\sqrt{\lambda_2}}{2}x_1w_2^T+\frac{2\sqrt{\lambda_1}\sqrt{\lambda_2}}{\sqrt{\lambda_1}+\sqrt{\lambda_2}}x_2^T\\ & +\frac{(\sqrt{\lambda_1}-\sqrt{\lambda_2})^2}{2(\sqrt{\lambda_1}+\sqrt{\lambda_2})}x_2^Tw_2w_2^T\\ \frac{\sqrt{\lambda_1}+\sqrt{\lambda_2}}{2}x_2+\frac{\sqrt{\lambda_1}-\sqrt{\lambda_2}}{2}x_1w_2 & \frac{\sqrt{\lambda_1}-\sqrt{\lambda_2}}{2}x_2w_2^T+\frac{2\sqrt{\lambda_1}\sqrt{\lambda_2}}{\sqrt{\lambda_1}+\sqrt{\lambda_2}}x_1I\\ & +\frac{(\sqrt{\lambda_1}-\sqrt{\lambda_2})^2}{2(\sqrt{\lambda_1}+\sqrt{\lambda_2})}x_1w_2w_2^T\end{bmatrix}\\
&=\begin{bmatrix} \frac{(x_1+x_2^Tw_2)}{2\sqrt{\lambda_2}}+\frac{(x_1-x_2^Tw_2)}{2\sqrt{\lambda_1}} & \left(\frac{x_1w_2^T}{2\sqrt{\lambda_2}}-\frac{x_1w_2^T}{2\sqrt{\lambda_1}}\right)+\frac{2x_2^T}{\sqrt{\lambda_1}+\sqrt{\lambda_2}}\\ & +\frac{\frac{\sqrt{\lambda_2}}{\sqrt{\lambda_1}}-2+\frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}}}{2(\sqrt{\lambda_1}+\sqrt{\lambda_2})}x_2^Tw_2w_2^T\\ \frac{(x_2+x_1w_2)}{2\sqrt{\lambda_2}}+\frac{(x_2-x_1w_2)}{2\sqrt{\lambda_1}} & \left(\frac{x_2w_2^T}{2\sqrt{\lambda_2}}-\frac{x_2w_2^T}{2\sqrt{\lambda_1}}\right)+\frac{2x_1I}{\sqrt{\lambda_1}+\sqrt{\lambda_2}}\\ & +\frac{\frac{\sqrt{\lambda_2}}{\sqrt{\lambda_1}}-2+\frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}}}{2(\sqrt{\lambda_1}+\sqrt{\lambda_2})}x_1w_2w_2^T\end{bmatrix}. \qquad (27)
\end{aligned}$$

Since $\lambda_2\ge\|x\|^2$, we see that $\sqrt{\lambda_2}\ge|x_1|$ and $\sqrt{\lambda_2}\ge\|x_2\|$. Also, $\|w_2\|=1$. Thus, terms that involve dividing x_1 or x_2 or x_1w_2 or $x_2^Tw_2$ or $x_1w_2w_2^T$ or $x_2^Tw_2w_2^T$ by $\sqrt{\lambda_2}$ or $\sqrt{\lambda_1}+\sqrt{\lambda_2}$ are uniformly bounded. Also, $\sqrt{\lambda_1}/\sqrt{\lambda_2}\le 1$. Thus

$$
\begin{aligned}
&L_x L_z^{-1}\\
&= \begin{bmatrix} O(1) + \dfrac{(x_1 - x_2^T w_2)}{2\sqrt{\lambda_1}} & O(1) - \dfrac{x_1 w_2^T}{2\sqrt{\lambda_1}} + \dfrac{\frac{\sqrt{\lambda_2}}{\sqrt{\lambda_1}}}{2(\sqrt{\lambda_1}+\sqrt{\lambda_2})} x_2^T w_2 w_2^T \\ O(1) + \dfrac{(x_2 - x_1 w_2)}{2\sqrt{\lambda_1}} & O(1) - \dfrac{x_2 w_2^T}{2\sqrt{\lambda_1}} + \dfrac{\frac{\sqrt{\lambda_2}}{\sqrt{\lambda_1}}}{2(\sqrt{\lambda_1}+\sqrt{\lambda_2})} x_1 w_2 w_2^T \end{bmatrix}\\
&= \begin{bmatrix} O(1) + \dfrac{(x_1 - x_2^T w_2)}{2\sqrt{\lambda_1}} & O(1) - \dfrac{x_1 w_2^T}{2(\sqrt{\lambda_1}+\sqrt{\lambda_2})} - \dfrac{\sqrt{\lambda_2}(x_1 - x_2^T w_2)}{2(\sqrt{\lambda_1}+\sqrt{\lambda_2})\sqrt{\lambda_1}} w_2^T \\ O(1) + \dfrac{(x_2 - x_1 w_2)}{2\sqrt{\lambda_1}} & O(1) - \dfrac{x_2 w_2^T}{2(\sqrt{\lambda_1}+\sqrt{\lambda_2})} - \dfrac{\sqrt{\lambda_2}(x_2 - x_1 w_2)}{2(\sqrt{\lambda_1}+\sqrt{\lambda_2})\sqrt{\lambda_1}} w_2^T \end{bmatrix},
\end{aligned}
$$

where $O(1)$ denote terms that are uniformly bounded, with bound independent of (x, y). By Lemma 3, if $x_1x_2 + y_1y_2 \neq 0$, then $|x_1 - x_2^T w_2| \leq \|x_2 - x_1 w_2\| \leq \sqrt{\lambda_1}$. If $x_1x_2 + y_1y_2 = 0$, then $\lambda_1 = \|x\|^2 + \|y\|^2$ so that, by choosing w_2 to further satisfy $x_2^T w_2 = 0$ (in addition to $\|w_2\| = 1$), we obtain

$$|x_1 - x_2^T w_2| \leq \|x_2 - x_1 w_2\| = \|x\| \leq \sqrt{\lambda_1}.$$

Thus, all terms in $L_x L_z^{-1}$ are uniformly bounded.

Using Lemmas 2, 3, 4 and Proposition 1, we now prove the smoothness of $\psi_{\rm FB}$. This has been proven for the NCP case [10, 25, 26] but not for the SOCCP case. A proof for the SDP case was only recently reported in [43]. Our proof for the SOCCP case is fairly involved due to the structure of the SOC and its associated Jordan product.

Proposition 2. *Let ϕ be given by* (11). *Then $\psi_{\rm FB}$ given by* (17) *is smooth everywhere on* $\mathbb{R}^n \times \mathbb{R}^n$.

Proof. By Proposition 1, $\psi_{\rm FB}$ is differentiable everywhere on $\mathbb{R}^n \times \mathbb{R}^n$. We will show that $\nabla\psi_{\rm FB}$ is continuous at every $(a, b) \in \mathbb{R}^n \times \mathbb{R}^n$. By the symmetry between x and y in $\nabla\psi_{\rm FB}$, it suffices to show that $\nabla_x\psi_{\rm FB}$ is continuous at every $(a, b) \in \mathbb{R}^n \times \mathbb{R}^n$.

Case (1): $a = b = 0$.
By Proposition 1, $\nabla_x\psi_{\rm FB}(0, 0) = 0$. Thus, we need to show that $\nabla_x\psi_{\rm FB}(x, y) \to 0$ as $(x, y) \to (0, 0)$. We consider two subcases: (i) $(x, y) \neq (0, 0)$ and $x^2+y^2 \in \operatorname{int}(\mathcal{K}^n)$ and (ii) $(x, y) \neq (0, 0)$ and $x^2+y^2 \notin \operatorname{int}(\mathcal{K}^n)$. In subcase (i), we have from Proposition 1 that $\nabla_x\psi_{\rm FB}(x, y)$ is given by the expression (21). By Lemma 4, $L_x L^{-1}_{(x^2+y^2)^{1/2}}$ is uniformly bounded, with bound independent of (x, y). Also, ϕ given by (11) is continuous at $(0, 0)$ so that $\phi(x, y) \to 0$ as $(x, y) \to (a, b)$. It follows from (21) that $\nabla_x\psi_{\rm FB}(x, y) \to 0$ as $(x, y) \to (a, b)$ in subcase (i). In subcase (ii), we have from Proposition 1 that $\nabla_x\psi_{\rm FB}(x, y)$ is given by the expression (22). Clearly $x_1/\sqrt{x_1^2 + y_1^2}$ is uniformly bounded, with bound independent of (x, y). Also, $\phi(x, y) \to 0$ as $(x, y) \to (a, b)$. It follows from (22) that $\nabla_x\psi_{\rm FB}(x, y) \to 0$ as $(x, y) \to (a, b)$ in subcase (ii).

Case (2): $(a, b) \neq (0, 0)$ and $a^2 + b^2 \in \operatorname{int}(\mathcal{K}^n)$.
It was already shown in the proof of Proposition 1 that $\psi_{\rm FB}$ is continuously differentiable at (a, b).

Case (3): $(a,b) \neq (0,0)$ and $a^2 + b^2 \notin \mathrm{int}(\mathcal{K}^n)$.
By (18), $\|a\|^2 + \|b\|^2 = 2\|a_1 a_2 + b_1 b_2\|$. By Proposition 1, we have $a_1^2 + b_1^2 > 0$ and

$$\nabla_x \psi_{\mathrm{FB}}(a,b) \;=\; \left(\frac{a_1}{\sqrt{a_1^2 + b_1^2}} - 1 \right) \phi(a,b).$$

We need to show that $\nabla_x \psi_{\mathrm{FB}}(x,y) \to \nabla_x \psi_{\mathrm{FB}}(a,b)$. We consider two cases: (i) $(x,y) \neq (0,0)$ and $x^2 + y^2 \in \mathrm{int}(\mathcal{K}^n)$ and (ii) $(x,y) \neq (0,0)$ and $x^2 + y^2 \notin \mathrm{int}(\mathcal{K}^n)$. In subcase (ii), we have from Proposition 1 that $\nabla_x \psi_{\mathrm{FB}}(x,y)$ is given by the expression (22). This expression is continuous at (a,b). Thus $\nabla_x \psi_{\mathrm{FB}}(x,y) \to \nabla_x \psi_{\mathrm{FB}}(a,b)$ as $(x,y) \to (a,b)$ in subcase (ii). The remainder of our proof treats subcase (i). In subcase (i), we have from Proposition 1 that $\nabla_x \psi_{\mathrm{FB}}(x,y)$ is given by the expression (21), i.e.,

$$\begin{aligned}
\nabla_x \psi_{\mathrm{FB}}(x,y) &= \left(L_x L^{-1}_{(x^2+y^2)^{1/2}} - I \right) \phi(x,y) \\
&= L_x L^{-1}_{(x^2+y^2)^{1/2}} (x^2 + y^2)^{1/2} - L_x L^{-1}_{(x^2+y^2)^{1/2}} (x+y) - \phi(x,y) \\
&= x - L_x L^{-1}_{(x^2+y^2)^{1/2}} (x+y) - \phi(x,y).
\end{aligned}$$

Also, by Lemma 2, we have $\|a_1 a_2 + b_1 b_2\| = \frac{1}{2}\|a\|^2 + \frac{1}{2}\|b\|^2 = a_1^2 + b_1^2$ and $a_1 b_2 = b_1 a_2$, implying that (see (19), (20))

$$\begin{aligned}
\frac{a_1}{\sqrt{a_1^2 + b_1^2}} (a^2 + b^2)^{1/2} &= \frac{a_1}{\sqrt{a_1^2 + b_1^2}} \left(\sqrt{a_1^2 + b_1^2}, \frac{a_1 a_2 + b_1 b_2}{\sqrt{a_1^2 + b_1^2}} \right) \\
&= \left(a_1, \frac{a_1^2 a_2 + a_1 b_1 b_2}{a_1^2 + b_1^2} \right) \\
&= \left(a_1, \frac{a_1^2 a_2 + b_1^2 a_2}{a_1^2 + b_1^2} \right) \\
&= (a_1, a_2) \\
&= a.
\end{aligned}$$

This together with (22) yields

$$\begin{aligned}
\nabla_x \psi_{\mathrm{FB}}(a,b) &= \left(\frac{a_1}{\sqrt{a_1^2 + b_1^2}} - 1 \right) \phi(a,b) \\
&= \frac{a_1}{\sqrt{a_1^2 + b_1^2}} \left((a^2 + b^2)^{1/2} - (a+b) \right) - \phi(a,b) \\
&= \frac{a_1}{\sqrt{a_1^2 + b_1^2}} (a^2 + b^2)^{1/2} - \frac{a_1}{\sqrt{a_1^2 + b_1^2}} (a+b) - \phi(a,b) \\
&= a - \frac{a_1}{\sqrt{a_1^2 + b_1^2}} (a+b) - \phi(a,b).
\end{aligned}$$

Since ϕ is continuous, to prove $\nabla_x \psi_{\text{FB}}(x, y) \to \nabla_x \psi_{\text{FB}}(a, b)$ as $(x, y) \to (a, b)$, it suffices to show that

$$L_x L^{-1}_{(x^2+y^2)^{1/2}} x \to \frac{a_1}{\sqrt{a_1^2 + b_1^2}}\, a \qquad \text{as } (x, y) \to (a, b), \tag{28}$$

$$L_x L^{-1}_{(x^2+y^2)^{1/2}} y \to \frac{a_1}{\sqrt{a_1^2 + b_1^2}}\, b \qquad \text{as } (x, y) \to (a, b). \tag{29}$$

Since $\|a\|^2 + \|b\|^2 = 2\|a_1 a_2 + b_1 b_2\|$ and $(a, b) \neq (0, 0)$, then $a_1 a_2 + b_1 b_2 \neq 0$. Thus, by taking (x, y) sufficiently near to (a, b), we can assume that $x_1 x_2 + y_1 y_2 \neq 0$. Let $z := (x^2 + y^2)^{1/2}$. Then z is given by (20) with λ_1, λ_2 given by (19) and $w_2 := \dfrac{x_1 x_2 + y_1 y_2}{\|x_1 x_2 + y_1 y_2\|}$. In addition, $det(z) = z_1^2 - \|z_2\|^2 = \sqrt{\lambda_1 \lambda_2}$. Let $(\zeta_1, \zeta_2) := L_x L_z^{-1} x$. Then (28) reduces to

$$\zeta_1 \to \frac{a_1^2}{\sqrt{a_1^2 + b_1^2}} \quad \text{and} \quad \zeta_2 \to \frac{a_1}{\sqrt{a_1^2 + b_1^2}}\, a_2 \qquad \text{as } (x, y) \to (a, b). \tag{30}$$

We prove (30) below. By Lemma 2, as $(x, y) \to (a, b)$,

$$\lambda_1 \to 0, \quad \lambda_2 \to \|a\|^2 + \|b\|^2 + 2\|a_1 a_2 + b_1 b_2\| = 4(a_1^2 + b_1^2), \quad z_1 \to \sqrt{a_1^2 + b_1^2}. \tag{31}$$

Using (27), we calculate the first component of $L_x L_z^{-1} x$ to be

$$\begin{aligned}
\zeta_1 &:= \frac{1}{det(z)} \left(x_1^2 z_1 - x_2^T z_2 x_1 - x_1 z_2^T x_2 + \frac{det(z)}{z_1} \|x_2\|^2 + \frac{(x_2^T z_2)^2}{z_1} \right), \\
&= \frac{\|x_2\|^2}{z_1} + \frac{1}{z_1 det(z)} \left(x_1^2 z_1^2 - 2 x_2^T z_2 x_1 z_1 + (x_2^T z_2)^2 \right) \\
&= \frac{\|x_2\|^2}{z_1} + \frac{(x_1 z_1 - x_2^T z_2)^2}{z_1 det(z)}.
\end{aligned}$$

Also, using Lemma 2 and (31),

$$\frac{\|x_2\|^2}{z_1} \to \frac{\|a_2\|^2}{\sqrt{a_1^2 + b_1^2}} = \frac{a_1^2}{\sqrt{a_1^2 + b_1^2}}.$$

Thus, to prove the first relation in (30), it suffices to show that

$$\frac{(x_1 z_1 - x_2^T z_2)^2}{z_1 det(z)} \to 0 \qquad \text{as } (x, y) \to (a, b).$$

We have that

$$\begin{aligned}\frac{(x_1z_1 - x_2^T z_2)^2}{z_1 det(z)} &= \frac{1}{z_1\sqrt{\lambda_1\lambda_2}}\left(x_1\frac{\sqrt{\lambda_1}+\sqrt{\lambda_2}}{2} + \frac{\sqrt{\lambda_1}-\sqrt{\lambda_2}}{2}x_2^T w_2\right)^2 \\ &= \frac{1}{z_1\sqrt{\lambda_1\lambda_2}}\left(x_1\sqrt{\lambda_1} + \frac{\sqrt{\lambda_2}-\sqrt{\lambda_1}}{2}\left(x_1 - x_2^T w_2\right)\right)^2 \\ &= \frac{1}{z_1\sqrt{\lambda_2}}\left(x_1^2\sqrt{\lambda_1} + x_1\left(\sqrt{\lambda_2}-\sqrt{\lambda_1}\right)\left(x_1 - x_2^T w_2\right)\right. \\ &\quad \left. + \frac{(\sqrt{\lambda_2}-\sqrt{\lambda_1})^2}{4\sqrt{\lambda_1}}\left(x_1 - x_2^T w_2\right)^2\right). \end{aligned} \tag{32}$$

We also have from (31) that $\lambda_1 \to 0, \sqrt{\lambda_2} \to 2\sqrt{a_1^2+b_1^2} > 0$, and $z_1 \to \sqrt{a_1^2+b_1^2} > 0$. Moreover, by Lemma 3 and $w_2 = \dfrac{x_1x_2 + y_1y_2}{\|x_1x_2 + y_1y_2\|}$,

$$\frac{\left(x_1 - x_2^T w_2\right)^2}{\sqrt{\lambda_1}} \to 0 \qquad \text{as } (x, y) \to (a, b).$$

Thus the right-hand side of (32) tends to zero as $(x, y) \to (a, b)$. This proves the first relation in (30).

Using (27), we calculate the last $n-1$ components of $L_x L_z^{-1} x$ to be

$$\begin{aligned}\zeta_2 &:= \frac{1}{det(z)}\left(x_1x_2z_1 - x_1^2z_2 - x_2^T z_2x_2 + \frac{x_1det(z)}{z_1}x_2 + \frac{x_1}{z_1}z_2z_2^T x_2\right) \\ &= \frac{x_1}{z_1}x_2 + \frac{1}{det(z)}\left((x_1z_1 - x_2^T z_2)x_2 + x_1\left(\frac{x_2^T z_2}{z_1} - x_1\right)z_2\right) \\ &= \frac{x_1}{z_1}x_2 + \frac{(x_1z_1 - x_2^T z_2)}{det(z)}\left(x_2 - \frac{x_1}{z_1}z_2\right).\end{aligned}$$

Also, by (31),

$$\frac{x_1}{z_1}x_2 \;\to\; \frac{a_1}{\sqrt{a_1^2+b_1^2}}\,a_2.$$

Thus, to prove the second relation in (30), it suffices to show that

$$\frac{(x_1z_1 - x_2^T z_2)}{det(z)}\left(x_2 - \frac{x_1}{z_1}z_2\right) \to 0 \qquad \text{as } (x, y) \to (a, b).$$

First, $\dfrac{(x_1 z_1 - x_2^T z_2)}{det(z)}$ is bounded as $(x, y) \to (a, b)$ because, by (20),

$$\begin{aligned}\frac{(x_1 z_1 - x_2^T z_2)}{det(z)} &= \frac{1}{\sqrt{\lambda_1 \lambda_2}}\left(x_1 \frac{\sqrt{\lambda_1}+\sqrt{\lambda_2}}{2} - \frac{\sqrt{\lambda_2}-\sqrt{\lambda_1}}{2} x_2^T w_2\right) \\ &= \frac{1}{\sqrt{\lambda_1 \lambda_2}}\left(x_1\sqrt{\lambda_1} + \frac{\sqrt{\lambda_2}-\sqrt{\lambda_1}}{2}\left(x_1 - x_2^T w_2\right)\right) \\ &= \frac{x_1}{\sqrt{\lambda_2}} + \frac{\sqrt{\lambda_2}-\sqrt{\lambda_1}}{2\sqrt{\lambda_1\lambda_2}}\left(x_1 - x_2^T w_2\right) \\ &= \frac{x_1}{\sqrt{\lambda_2}} + \frac{1-\sqrt{\lambda_1}/\sqrt{\lambda_2}}{2}\,\frac{\left(x_1 - x_2^T w_2\right)}{\sqrt{\lambda_1}},\end{aligned}$$

and the first term on the right-hand side converges to $a_1/\sqrt{4(a_1^2+b_1^2)}$ (see (31)) while the second term is bounded by (19) and Lemma 3. Second, $x_2 - \dfrac{x_1}{z_1} z_2 \to 0$ as $(x, y) \to (a, b)$ because, by (20) and (31),

$$\begin{aligned} x_2 - \frac{x_1}{z_1} z_2 &\to a_2 - \frac{a_1}{\sqrt{a_1^2+b_1^2}}\,\frac{\sqrt{4(a_1^2+b_1^2)}}{2}\,\frac{a_1 a_2 + b_1 b_2}{\|a_1 a_2 + b_1 b_2\|} \\ &= a_2 - \frac{a_1^2 a_2 + a_1 b_1 b_2}{\|a_1 a_2 + b_1 b_2\|} \\ &= a_2 - \frac{a_1^2 a_2 + b_1^2 a_2}{a_1^2 + b_1^2} \\ &= a_2 - a_2 \\ &= 0,\end{aligned}$$

where the second equality is due to Lemma 2, so that $a_1 b_2 = b_1 a_2$ and $\|a_1 a_2 + b_1 b_2\| = a_1^2 + b_1^2$. This proves the second relation in (30).

Thus, we have proven (28). An analogous argument can be used to prove (29), which we omit for simplicity. This shows that $\nabla_x \psi_{\mathrm{FB}}(x, y) \to \nabla_x \psi_{\mathrm{FB}}(a, b)$ as $(x, y) \to (a, b)$ in subcase (i).

It follows from Proposition 2 that the merit function ψ_{YF} given by (13), with ψ_0 a smooth function, is also smooth.

4. Stationary points of merit functions for monotone SOCCP

In this section we consider the case where SOCCP has a monotonicity property and show that every stationary point of (9) is a solution of the SOCCP. As in the previous section, we focus our analysis on the case of $N = 1$ for simplicity. We first need the following technical lemma from [19].

Lemma 5. ([19, Proposition 3.4]) *For any* $x, y \in \mathbb{R}^n$ *and* $w \in \mathcal{K}^n$ *such that* $w^2 - x^2 - y^2 \in \mathcal{K}^n$, *we have* $L_w^2 \succeq L_x^2 + L_y^2$.

Using Lemmas 1, 2, 5 and Proposition 1, we prove the following key properties of $\nabla\psi_{\text{FB}}$. Similar properties have been proven for the case of NCP [10, 20, 34] and SDCP [50, 53]. However, our proof is quite different from these other proofs due to the different structures of SOC and its associated Jordan product.

Lemma 6. *Let* ϕ *be given by* (11) *and let* ψ_{FB} *be given by* (17). *For any* $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$, *we have the following results.*

(a)

$$\langle x, \nabla_x \psi_{\text{FB}}(x, y)\rangle + \langle y, \nabla_y \psi_{\text{FB}}(x, y)\rangle = \|\phi(x, y)\|^2 . \tag{33}$$

(b)

$$\langle \nabla_x \psi_{\text{FB}}(x, y), \nabla_y \psi_{\text{FB}}(x, y)\rangle \;\geq\; 0, \tag{34}$$

with equality holding if and only if $\phi(x, y) = 0$.

Proof. Case (1): $x = y = 0$.
By Proposition 1, $\nabla_x \psi_{\text{FB}}(x, y) = \nabla_y \psi_{\text{FB}}(x, y) = 0$, so the proposition is true.

Case (2): $(x, y) \neq (0, 0)$ and $x^2 + y^2 \in \text{int}(\mathcal{K}^n)$.
By Proposition 1, we have

$$\begin{aligned}
\nabla_x \psi_{\text{FB}}(x, y) &= \left(L_x L_z^{-1} - I\right)\phi(x, y),\\
\nabla_y \psi_{\text{FB}}(x, y) &= \left(L_y L_z^{-1} - I\right)\phi(x, y),
\end{aligned}$$

where we let $z := (x^2 + y^2)^{1/2}$. For simplicity, we will write $\phi(x, y)$ as ϕ. Thus,

$$\begin{aligned}
\langle x, \nabla_x \psi_{\text{FB}}(x, y)\rangle + \langle y, \nabla_y \psi_{\text{FB}}(x, y)\rangle &= \langle x, (L_x L_z^{-1} - I)\phi\rangle + \langle y, (L_y L_z^{-1} - I)\phi\rangle\\
&= \langle (L_z^{-1} L_x - I)x, \phi\rangle + \langle (L_z^{-1} L_y - I)y, \phi\rangle\\
&= \langle L_z^{-1} L_x x + L_z^{-1} L_y y - x - y, \phi\rangle\\
&= \langle L_z^{-1}(x^2 + y^2) - x - y, \phi\rangle\\
&= \langle L_z^{-1} z^2 - x - y, \phi\rangle\\
&= \langle z - x - y, \phi\rangle\\
&= \|\phi\|^2,
\end{aligned}$$

where the next-to-last equality follows from $L_z z = z^2$, so that $L_z^{-1} z^2 = z$. This proves (33). Similarly,

$$\begin{aligned}
\langle \nabla_x \psi_{\text{FB}}(x, y),\ \nabla_y \psi_{\text{FB}}(x, y)\rangle &= \langle (L_x L_z^{-1} - I)\phi,\ (L_y L_z^{-1} - I)\phi\rangle\\
&= \langle (L_x - L_z) L_z^{-1}\phi,\ (L_y - L_z) L_z^{-1}\phi\rangle\\
&= \langle (L_y - L_z)(L_x - L_z) L_z^{-1}\phi,\ L_z^{-1}\phi\rangle. \qquad (35)
\end{aligned}$$

Let S be the symmetric part of $(L_y - L_z)(L_x - L_z)$. Then

$$\begin{aligned} S &= \frac{1}{2}\Big((L_y - L_z)(L_x - L_z) + (L_x - L_z)(L_y - L_z)\Big) \\ &= \frac{1}{2}\Big(L_x L_y + L_y L_x - L_z(L_x + L_y) - (L_x + L_y)L_z + 2L_z^2\Big) \\ &= \frac{1}{2}(L_z - L_x - L_y)^2 + \frac{1}{2}(L_z^2 - L_x^2 - L_y^2). \end{aligned}$$

Since $z \in \mathcal{K}^n$ and $z^2 = x^2 + y^2$, Lemma 5 yields $L_z^2 - L_x^2 - L_y^2 \succeq O$. Then (35) yields

$$\begin{aligned} &\langle \nabla_x \psi_{\text{FB}}(x, y),\ \nabla_y \psi_{\text{FB}}(x, y)\rangle \\ &\quad = \langle S L_z^{-1}\phi, L_z^{-1}\phi\rangle \\ &\quad = \frac{1}{2}\langle (L_z - L_x - L_y)^2 L_z^{-1}\phi, L_z^{-1}\phi\rangle + \frac{1}{2}\langle (L_z^2 - L_x^2 - L_y^2) L_z^{-1}\phi, L_z^{-1}\phi\rangle \\ &\quad \geq \frac{1}{2}\langle (L_z - L_x - L_y)^2 L_z^{-1}\phi, L_z^{-1}\phi\rangle \\ &\quad = \frac{1}{2}\|L_\phi L_z^{-1}\phi\|^2, \end{aligned}$$

where the last equality uses $L_z - L_x - L_y = L_{z-x-y} = L_\phi$. This proves (34).

If the inequality in (34) holds with equality, then the above relation yields $\|L_\phi L_z^{-1}\phi\|^2 = 0$ and, by Property 1(d),

$$\phi \cdot (L_z^{-1}\phi) = L_\phi L_z^{-1}\phi = 0.$$

Then, the definition of Jordan product (10) yields

$$\langle \phi,\ L_z^{-1}\phi\rangle = 0.$$

Since $z = (x^2 + y^2)^{1/2} \in \text{int}(\mathcal{K}^n)$ so that $L_z^{-1} \succ O$ (see Property 1(d)), this implies $\phi = 0$. Conversely, if $\phi = 0$, then it follows from (21) that

$$\langle \nabla_x \psi_{\text{FB}}(x, y), \nabla_y \psi_{\text{FB}}(x, y)\rangle = 0.$$

Case (3): $(x, y) \neq (0, 0)$ and $x^2 + y^2 \notin \text{int}(\mathcal{K}^n)$.
By Proposition 1, we have

$$\begin{aligned} \nabla_x \psi_{\text{FB}}(x, y) &= \left(\frac{x_1}{\sqrt{x_1^2 + y_1^2}} - 1\right)\phi(x, y), \\ \nabla_y \psi_{\text{FB}}(x, y) &= \left(\frac{y_1}{\sqrt{x_1^2 + y_1^2}} - 1\right)\phi(x, y). \end{aligned}$$

Thus,

$$
\begin{aligned}
&\langle x, \nabla_x \psi_{\text{FB}}(x, y)\rangle + \langle y, \nabla_y \psi_{\text{FB}}(x, y)\rangle \\
&= \left(\frac{x_1}{\sqrt{x_1^2 + y_1^2}} - 1 \right) \langle x, \phi(x, y)\rangle + \left(\frac{y_1}{\sqrt{x_1^2 + y_1^2}} - 1 \right) \langle y, \phi(x, y)\rangle \\
&= \left\langle \left(\frac{x_1}{\sqrt{x_1^2 + y_1^2}} - 1 \right) x + \left(\frac{y_1}{\sqrt{x_1^2 + y_1^2}} - 1 \right) y, \phi(x, y) \right\rangle \\
&= \left\langle \frac{x_1 x + y_1 y}{\sqrt{x_1^2 + y_1^2}} - x - y, \phi(x, y) \right\rangle \\
&= \langle \phi(x, y), \phi(x, y)\rangle,
\end{aligned}
$$

where the last equality uses (26). This proves (33). Similarly,

$$
\begin{aligned}
\langle \nabla_x \psi_{\text{FB}}(x, y),\ \nabla_y \psi_{\text{FB}}(x, y)\rangle &= \left(\frac{x_1}{\sqrt{x_1^2 + y_1^2}} - 1 \right) \left(\frac{y_1}{\sqrt{x_1^2 + y_1^2}} - 1 \right) \|\phi(x, y)\|^2 \\
&\geq 0.
\end{aligned}
$$

This proves (34). If the inequality in (34) holds with equality, then either $\phi(x, y) = 0$ or $\frac{x_1}{\sqrt{x_1^2+y_1^2}} = 1$ or $\frac{y_1}{\sqrt{x_1^2+y_1^2}} = 1$. In the second case, we have $y_1 = 0$ and $x_1 \geq 0$, so that Lemma 2 yields $y_2 = 0$ and $x_1 = \|x_2\|$. In the third case, we have $x_1 = 0$ and $y_1 \geq 0$, so that Lemma 2 yields $x_2 = 0$ and $y_1 = \|y_2\|$. Thus, in these two cases, we have $x \cdot y = 0,\ x \in \mathcal{K}^n,\ y \in \mathcal{K}^n$. Then, by Lemma 1, $\phi(x, y) = 0$.

Below we assume that

$$
\nabla F(\zeta), -\nabla G(\zeta) \text{ are column monotone} \qquad \forall \zeta \in \mathbb{R}^n; \tag{36}
$$

see [9, p. 1014], [34, p. 222].[2] In the case of (4), corresponding to $\nabla G(\zeta) = I$, (36) is equivalent to F being monotone. More generally, if $\nabla G(\zeta)$ is invertible, then (36) is equivalent to $\nabla G(\zeta)^{-1} \nabla F(\zeta) \succeq O$ for all $\zeta \in \mathbb{R}^n$. In the case of (6), (36) is satisfied always. To see this, note that $\nabla F(\zeta) = [B\ 0]^T$ and $\nabla G(\zeta) = [B\ 0]^T \nabla^2 g(F(\zeta)) - [0\ A^T]^T$. Hence $\nabla F(\zeta) u - \nabla G(\zeta) v = 0$ is equivalent to

$$
\begin{bmatrix} B^T \\ 0 \end{bmatrix} u - \begin{bmatrix} B^T \\ 0 \end{bmatrix} \nabla^2 g(F(\zeta)) v + \begin{bmatrix} 0 \\ A \end{bmatrix} v = 0
$$

for any $u, v \in \mathbb{R}^n$. This yields $B^T u = B^T \nabla^2 g(F(\zeta)) v$ and $Av = 0$. The second equation implies $v = Bw$ for some $w \in \mathbb{R}^{n-m}$, so that multiplying the first equation on the left by w^T and using $\nabla^2 g(F(\zeta)) \succeq 0$ (since g is convex) yields

$$
w^T B^T u = v^T u = v^T \nabla^2 g(F(\zeta)) v \geq 0.
$$

[2] $M, N \in \mathbb{R}^{n \times n}$ are column monotone if, for any $u, v \in \mathbb{R}^n$, $Mu + Nv = 0 \Rightarrow u^T v \geq 0$.

In the case of (7), (36) is also satisfied always, as can be argued similarly. Moreover, the argument extends to the more general problem where ∇g is replaced by any differentiable monotone mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$.

Using Lemma 5(b), we prove below the first main result of this section, based on the merit function ψ_{FB}. Analogous results have been proven for the NCP case [10, 20] and the SDCP case [50].

Proposition 3. *Let ϕ be given by* (11) *and let ψ_{FB} be given by* (17). *Let f_{FB} be given by* (15)*, where F and G are differentiable mappings from $\mathbb{R}^n$ to $\mathbb{R}^n$ satisfying* (36). *Then, for every $\zeta \in \mathbb{R}^n$, either (i) $f_{\mathrm{FB}}(\zeta) = 0$ or (ii) $\nabla f_{\mathrm{FB}}(\zeta) \neq 0$. In case (ii), if $\nabla G(\zeta)$ is invertible, then $\langle d_{\mathrm{FB}}(\zeta), \nabla f_{\mathrm{FB}}(\zeta)\rangle < 0$, where*

$$d_{\mathrm{FB}}(\zeta) := -(\nabla G(\zeta)^{-1})^T \nabla_x \psi_{\mathrm{FB}}(F(\zeta), G(\zeta)).$$

Proof. Fix any $\zeta \in \mathbb{R}^n$. By Proposition 2, ψ_{FB} is smooth, so the chain rule for differentiation yields

$$\nabla f_{\mathrm{FB}}(\zeta) = \nabla F(\zeta)\nabla_x \psi_{\mathrm{FB}}(F(\zeta), G(\zeta)) + \nabla G(\zeta)\nabla_y \psi_{\mathrm{FB}}(F(\zeta), G(\zeta)).$$

Suppose $\nabla f_{\mathrm{FB}}(\zeta) = 0$. The column monotone property of $\nabla F(\zeta), -\nabla G(\zeta)$ yields

$$\langle \nabla_x \psi_{\mathrm{FB}}(F(\zeta), G(\zeta)), \nabla_y \psi_{\mathrm{FB}}(F(\zeta), G(\zeta))\rangle \leq 0.$$

By Lemma 6(b), the above inequality must hold with equality and hence $\phi(F(\zeta), G(\zeta)) = 0$. Thus $f_{\mathrm{FB}}(\zeta) = \frac{1}{2}\|\phi(F(\zeta), G(\zeta))\|^2 = 0$.

Suppose $\nabla f_{\mathrm{FB}}(\zeta) \neq 0$ and $\nabla G(\zeta)$ is invertible. Then (dropping the argument "(ζ)" for simplicity),

$$\begin{aligned}
\langle d_{\mathrm{FB}}, \nabla f_{\mathrm{FB}}\rangle &= \langle -(\nabla G^{-1})^T \nabla_x \psi_{\mathrm{FB}}(F, G),\ \nabla F \nabla_x \psi_{\mathrm{FB}}(F, G) + \nabla G \nabla_y \psi_{\mathrm{FB}}(F, G)\rangle \\
&= -\langle \nabla_x \psi_{\mathrm{FB}}(F, G),\ (\nabla G^{-1}\nabla F)\nabla_x \psi_{\mathrm{FB}}(F, G) + \nabla_y \psi_{\mathrm{FB}}(F, G)\rangle \\
&= -\langle \nabla_x \psi_{\mathrm{FB}}(F, G),\ (\nabla G^{-1}\nabla F)\nabla_x \psi_{\mathrm{FB}}(F, G)\rangle \\
&\quad -\langle \nabla_x \psi_{\mathrm{FB}}(F, G),\ \nabla_y \psi_{\mathrm{FB}}(F, G)\rangle \\
&\leq -\langle \nabla_x \psi_{\mathrm{FB}}(F, G),\ \nabla_y \psi_{\mathrm{FB}}(F, G)\rangle,
\end{aligned}$$

where the inequality follows from $\nabla G^{-1}\nabla F \succeq 0$. By Lemma 6(b), the right-hand side is non-positive and equals zero if and only if $\phi(F, G) = 0$, i.e., ζ is a global minimum of f_{FB}. Since $\nabla f_{\mathrm{FB}} \neq 0$, the right-hand side cannot equal zero, so it must be negative.

The direction $d_{\mathrm{FB}}(\zeta)$ has the advantage that, unlike $-\nabla f_{\mathrm{FB}}(\zeta)$, it does not require $\nabla F(\zeta)$ for its evaluation. However, for CSOCP (6) or (7), $\nabla G(\zeta)$ is not invertible, so this direction cannot be used. Using Lemma 5, we prove below the second main result of this section, based on the merit function ψ_{YF} given by (13). Similar results have been proven for the NCP case [34] and the SDCP case [53].

Proposition 4. *Let ϕ be given by* (11)*, let ψ_{FB} be given by* (17)*, and let ψ_{YF} be given by* (13)*, with $\psi_0 : \mathbb{R} \to [0, \infty)$ being any smooth function satisfying* (14). *Let f_{YF} be given by* (16)*, where F and G are differentiable mappings from $\mathbb{R}^n$ to $\mathbb{R}^n$ satisfying*

(36). *Then, for every* $\zeta \in \mathbb{R}^n$, *either (i)* $f_{\mathrm{YF}}(\zeta) = 0$ *or (ii)* $\nabla f_{\mathrm{YF}}(\zeta) \neq 0$. *In case (ii), if* $\nabla G(\zeta)$ *is invertible, then* $\langle d_{\mathrm{YF}}(\zeta), \nabla f_{\mathrm{YF}}(\zeta)\rangle < 0$, *where*

$$d_{\mathrm{YF}}(\zeta) := -(\nabla G(\zeta)^{-1})^T \Big(\psi_0'(\langle F(\zeta), G(\zeta)\rangle) G(\zeta) + \nabla_x \psi_{\mathrm{FB}}(F(\zeta), G(\zeta)) \Big).$$

Proof. Fix any $\zeta \in \mathbb{R}^n$. By Proposition 2, ψ_{FB} is smooth. Since ψ_0 is smooth, (13) shows that ψ_{YF} is smooth. Then the chain rule for differentiation yields

$$\begin{aligned}\nabla f_{\mathrm{YF}}(\zeta) = \alpha \Big(&\nabla F(\zeta) G(\zeta) + \nabla G(\zeta) F(\zeta) \Big) \\ &+ \nabla F(\zeta) \nabla_x \psi_{\mathrm{FB}}(F(\zeta), G(\zeta)) + \nabla G(\zeta) \nabla_y \psi_{\mathrm{FB}}(F(\zeta), G(\zeta)),\end{aligned}$$

where we let $\alpha := \psi_0'\Big(\langle F(\zeta), G(\zeta)\rangle \Big)$.

Suppose $\nabla f_{\mathrm{YF}}(\zeta) = 0$. Then, dropping the argument "(ζ)" for simplicity, we have

$$\alpha \Big(\nabla F G + \nabla G F \Big) + \nabla F \nabla_x \psi_{\mathrm{FB}}(F, G) + \nabla G \nabla_y \psi_{\mathrm{FB}}(F, G) = 0.$$

The column monotone property of $\nabla F, -\nabla G$ yields

$$\langle \alpha G + \nabla_x \psi_{\mathrm{FB}}(F, G), \alpha F + \nabla_y \psi_{\mathrm{FB}}(F, G)\rangle \leq 0.$$

Upon collecting terms on the left-hand side, we have

$$\begin{aligned}\alpha^2 \langle F, G\rangle &+ \alpha \big(\langle F, \nabla_x \psi_{\mathrm{FB}}(F, G)\rangle + \langle G, \nabla_y \psi_{\mathrm{FB}}(F, G)\rangle \big) \\ &+ \langle \nabla_x \psi_{\mathrm{FB}}(F, G), \nabla_y \psi_{\mathrm{FB}}(F, G)\rangle \leq 0.\end{aligned}$$

Our assumption (14) on ψ_0 implies the first term is nonnegative. By Lemma 6, the second and the third terms are also nonnegative. Thus, the third term must be zero, so Lemma 6(b) implies $\phi(F, G) = 0$. Thus $f_{\mathrm{FB}}(\zeta) = \frac{1}{2}\|\phi(F(\zeta), G(\zeta))\|^2 = 0$.

Suppose $\nabla f_{\mathrm{YF}}(\zeta) \neq 0$ and $\nabla G(\zeta)$ is invertible. Again, we drop the argument "(ζ)" for simplicity. Then,

$$\begin{aligned}\langle d_{\mathrm{YF}}, \nabla f_{\mathrm{YF}}\rangle &= \Big\langle -(\nabla G^{-1})^T (\alpha G + \nabla_x \psi_{\mathrm{FB}}(F, G)),\ \nabla F(\alpha G + \nabla_x \psi_{\mathrm{FB}}(F, G)) \\ &\qquad + \nabla G(\alpha F + \nabla_y \psi_{\mathrm{FB}}(F, G)) \Big\rangle \\ &= -\Big\langle \alpha G + \nabla_x \psi_{\mathrm{FB}}(F, G),\ \nabla G^{-1} \nabla F(\alpha G + \nabla_x \psi_{\mathrm{FB}}(F, G)) \Big\rangle \\ &\quad - \Big\langle \alpha G + \nabla_x \psi_{\mathrm{FB}}(F, G),\ \alpha F + \nabla_y \psi_{\mathrm{FB}}(F, G) \Big\rangle \\ &\leq -\Big\langle \alpha G + \nabla_x \psi_{\mathrm{FB}}(F, G),\ \alpha F + \nabla_y \psi_{\mathrm{FB}}(F, G) \Big\rangle \\ &= -\alpha^2 \langle F, G\rangle - \alpha \Big(\langle F,\ \nabla_x \psi_{\mathrm{FB}}(F, G)\rangle + \langle G,\ \nabla_y \psi_{\mathrm{FB}}(F, G)\rangle \Big) \\ &\quad - \langle \nabla_x \psi_{\mathrm{FB}}(F, G),\ \nabla_y \psi_{\mathrm{FB}}(F, G)\rangle,\end{aligned}$$

where the first inequality follows from $\nabla G^{-1}\nabla F \succeq 0$. We argued earlier that all three terms on the right-hand side are non-positive. Moreover, by Lemma 6(b), the third term is zero if and only if $\phi(F, G) = 0$, i.e., ζ is a global minimum of f_{YF} and hence a stationary point of f_{YF}. Since $\nabla f_{\mathrm{YF}}(\zeta) \neq 0$, the right-hand side cannot equal zero, so it must be negative.

5. Bounded level sets and error bounds for f_{YF}

In this section, we consider the merit function f_{YF} given by (16). We show that, analogous to the NCP and SDCP cases [34, 53], if F and G have a joint monotonicity property and a strictly feasible solution exists, then f_{YF} has bounded level sets. If F and G have a joint strong monotonicity property, then f_{YF} has bounded level sets and provides a global error bound on the distance to a solution of SOCCP. In contrast, the merit function f_{FB} given by (15) lacks these properties due to the absence of the term $\psi_0(\langle F(\zeta), G(\zeta)\rangle)$.

As in the previous two sections, we focus our analysis on the case of $N = 1$ (i.e., $\mathcal{K} = \mathcal{K}^n$) for simplicity. In what follows, for each $x \in \mathbb{R}^n$, x_+ denotes the nearest-point (in the Euclidean norm) projection of x onto $\mathcal{K}^n$. We begin with the following lemma.

Lemma 7. *Let $\mathcal{K}$ be any closed convex cone in $\mathbb{R}^n$. For each $x \in \mathbb{R}^n$, let $x_{\mathcal{K}}^+$ and $x_{\mathcal{K}}^-$ denote the nearest-point (in the Euclidean norm) projection of x onto $\mathcal{K}$ and $-\mathcal{K}^*$, respectively. The following results hold.*

(a) For any $x \in \mathbb{R}^n$, we have $x = x_{\mathcal{K}}^+ + x_{\mathcal{K}}^-$ and $\|x\|^2 = \|x_{\mathcal{K}}^+\|^2 + \|x_{\mathcal{K}}^-\|^2$.
(b) For any $x \in \mathbb{R}^n$ and $y \in \mathcal{K}$, we have $\langle x, y\rangle \leq \langle x_{\mathcal{K}}^+, y\rangle$.
(c) If $\mathcal{K}$ is self-dual, then for any $x \in \mathbb{R}^n$ and $y \in \mathcal{K}$, we have $\left\|(x+y)_{\mathcal{K}}^+\right\| \geq \left\|x_{\mathcal{K}}^+\right\|$.
(d) For any $x \in \mathcal{K}^n$, $y \in \mathbb{R}^n$ with $x^2 - y^2 \in \mathcal{K}^n$, we have $x - y \in \mathcal{K}^n$.

Proof. (a). These are well-known results in convex geometry on representing x as the sum of its projection onto $\mathcal{K}$ and its polar $-\mathcal{K}^*$.

(b). Since $x_{\mathcal{K}}^- \in -\mathcal{K}^*$ and $y \in \mathcal{K}$, $\langle x_{\mathcal{K}}^-, y\rangle \leq 0$. By (a), $\langle x, y\rangle = \langle x_{\mathcal{K}}^+, y\rangle + \langle x_{\mathcal{K}}^-, y\rangle \leq \langle x_{\mathcal{K}}^+, y\rangle$.

(c). Since $\mathcal{K}$ is self-dual, we have $y \in \mathcal{K}^*$. Then $(x+y)_{\mathcal{K}}^- - y \in -\mathcal{K}^*$. Since $x_{\mathcal{K}}^-$ is the nearest-point projection of x onto $-\mathcal{K}^*$, this implies

$$\|x_{\mathcal{K}}^- - x\| \leq \|((x+y)_{\mathcal{K}}^- - y) - x\|.$$

By (a), this simplifies to $\|x_{\mathcal{K}}^+\| \leq \|(x+y)_{\mathcal{K}}^+\|$.

(d) This is Proposition 3.4 of [19].

Lemma 7(c) generalizes [53, Lemma 2.4]. Using Lemma 7, we obtain the following two lemmas that are analogs of [53, Lemmas 2.5, 2.6] for SDCP.

Lemma 8. *Let ψ_{FB} be given by* (11) *and* (17). *For any $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$, we have*

$$4\psi_{\mathrm{FB}}(x, y) \geq 2\left\|\phi(x, y)_+\right\|^2 \geq \left\|(-x)_+\right\|^2 + \left\|(-y)_+\right\|^2.$$

Proof. The first inequality follows from Lemma 7(a). It remains to show the second inequality. By Lemma 7(d), $(x^2+y^2)^{1/2}-x \in \mathcal{K}^n$. Since $\mathcal{K}^n$ is self-dual, then Lemma 7(c) yields

$$\left\| \left((x^2+y^2)^{1/2}-x-y\right)_+ \right\|^2 \geq \left\| (-y)_+ \right\|^2.$$

By a symmetric argument,

$$\left\| \left((x^2+y^2)^{1/2}-x-y\right)_+ \right\|^2 \geq \left\| (-x)_+ \right\|^2.$$

Adding the above two inequalities yields the desired second inequality.

Lemma 9. *Let $\psi_{\rm FB}$ be given by* (11) *and* (17). *For any $\{(x^k, y^k)\}_{k=1}^\infty \subseteq \mathbb{R}^n \times \mathbb{R}^n$, let $\lambda_1^k \leq \lambda_2^k$ and $\mu_1^k \leq \mu_2^k$ denote the spectral values of x^k and y^k, respectively. Then the following results hold.*

(a) If $\lambda_1^k \to -\infty$ or $\mu_1^k \to -\infty$, then $\psi_{\rm FB}(x^k, y^k) \to \infty$.
(b) Suppose that $\{\lambda_1^k\}$ and $\{\mu_1^k\}$ are bounded below. If $\lambda_2^k \to \infty$ or $\mu_2^k \to \infty$, then $\langle x, x^k\rangle + \langle y, y^k\rangle \to \infty$ for any $x, y \in {\rm int}(\mathcal{K}^n)$.

Proof. (a). This follows from Lemma 8 and the fact that

$$2\|(-x^k)_+\|^2 = \sum_{i=1}^2 \left(\max\{0, -\lambda_i^k\}\right)^2$$

and similarly for $\|(-y^k)_+\|^2$; see [19, Property 2.2 and Proposition 3.3].

(b). Fix any $x=(x_1,x_2)$, $y=(y_1,y_2) \in \mathbb{R}\times\mathbb{R}^{n-1}$ with $\|x_2\|<x_1$, $\|y_2\|<y_1$. Using the spectral decomposition

$$x^k = \left(\frac{\lambda_1^k+\lambda_2^k}{2}, \ \frac{\lambda_2^k-\lambda_1^k}{2} w_2^k\right) \quad \text{with } \|w_2^k\| = 1,$$

we have

$$\langle x, x^k\rangle = \left(\frac{\lambda_1^k+\lambda_2^k}{2}\right) x_1 + \left(\frac{\lambda_2^k-\lambda_1^k}{2}\right) x_2^T w_2^k = \frac{\lambda_1^k}{2}(x_1 - x_2^T w_2^k) + \frac{\lambda_2^k}{2}(x_1 + x_2^T w_2^k). \tag{37}$$

Since $\|w_2^k\|=1$, we have $x_1 - x_2^T w_2^k \geq x_1 - \|x_2\| > 0$ and $x_1 + x_2^T w_2^k \geq x_1 - \|x_2\| > 0$. Since $\{\lambda_1^k\}$ is bounded below, the first term on the right-hand side of (37) is bounded below. If $\{\lambda_2^k\} \to \infty$, then the second term on the right-hand side of (37) tends to infinity. Hence, $\langle x, x^k\rangle \to \infty$. A similar argument shows that $\langle y, y^k\rangle$ is bounded below. Thus, $\langle x, x^k\rangle + \langle y, y^k\rangle \to \infty$. If $\{\mu_2^k\} \to \infty$, the argument is symmetric to the one above.

In what follows, we say that F and G are *jointly monotone* if

$$\langle F(\zeta) - F(\xi), G(\zeta) - G(\xi)\rangle \geq 0 \quad \forall \zeta,\ \xi \in \mathbb{R}^n.$$

Similarly, F and G are *jointly strongly monotone* if there exists $\rho > 0$ such that

$$\langle F(\zeta) - F(\xi), G(\zeta) - G(\xi)\rangle \geq \rho\|\zeta - \xi\|^2 \quad \forall \zeta,\ \xi \in \mathbb{R}^n.$$

In the case where $G(\zeta) = \zeta$ for all $\zeta \in \mathbb{R}^n$, the above notions are equivalent to the well-known notion of F being, respectively, monotone and strongly monotone [9, Section 2.3]. Since F is differentiable, F being monotone is equivalent to $\nabla F(\zeta) \succeq O$ for all $\zeta \in \mathbb{R}^n$; see, e.g., [9, Proposition 2.3.2].[3] It can be seen that F, G given by (6) or (7) are jointly monotone, but not jointly strongly monotone. It is not difficult to see that if F, G are jointly strongly monotone, then SOCCP has at most one solution. Sufficient conditions for SOCCP to have a solution are given in, e.g., [9, Sections 2.2, 2.4], [23, Chapter 6], as well as Proposition 6.

Using Lemmas 7(b) and 8, we obtain the following global error bound results for SOCCP that is an analog of [53, Theorem 4.2] for SDCP. The proof, based on Lemmas 7(b) and 8, is similar to the proof of [34, Theorem 3.4] and [53, Theorem 4.2] and is included for completeness.

Proposition 5. *Suppose that F and G are jointly strongly monotone mappings from $\mathbb{R}^n$ to $\mathbb{R}^n$. Also, suppose that SOCCP has a solution ζ^*. Then there exists a scalar $\tau > 0$ such that*

$$\tau\|\zeta - \zeta^*\|^2 \leq \max\{0, \langle F(\zeta), G(\zeta)\rangle\} + \|(-F(\zeta))_+\| + \|(-G(\zeta))_+\| \quad \forall \zeta \in \mathbb{R}^n. \tag{38}$$

Moreover,

$$\tau\|\zeta - \zeta^*\|^2 \leq \psi_0^{-1}\left(f_{\rm YF}(\zeta)\right) + 2\sqrt{2} f_{\rm YF}(\zeta)^{1/2} \quad \forall \zeta \in \mathbb{R}^n, \tag{39}$$

where $f_{\rm YF}$ is given by (13), (16), (17), *$\psi_0 : \mathbb{R} \to [0, \infty)$ is a smooth function satisfying* (14), *and ψ_0^{-1} denotes the inverse function of ψ_0 on* $[0, \infty)$.[4]

Proof. Since F and G are jointly strongly monotone, there exists a scalar $\rho > 0$ such that, for any $\zeta \in \mathbb{R}^n$,

$$\begin{aligned}
\rho\|\zeta - \zeta^*\|^2 &\leq \langle F(\zeta) - F(\zeta^*),\ G(\zeta) - G(\zeta^*)\rangle \\
&= \langle F(\zeta), G(\zeta)\rangle + \langle -F(\zeta), G(\zeta^*)\rangle + \langle F(\zeta^*), -G(\zeta)\rangle \\
&\leq \max\{0, \langle F(\zeta), G(\zeta)\rangle\} + \langle (-F(\zeta))_+, G(\zeta^*)\rangle + \langle F(\zeta^*), (-G(\zeta))_+\rangle \\
&\leq \max\{0, \langle F(\zeta), G(\zeta)\rangle\} + \|(-F(\zeta))_+\|\ \|G(\zeta^*)\| + \|F(\zeta^*)\|\ \|(-G(\zeta))_+\| \\
&\leq \max\{1, \|F(\zeta^*)\|, \|G(\zeta^*)\|\} \\
&\quad \times \Big(\max\{0, \langle F(\zeta), G(\zeta)\rangle\} + \|(-F(\zeta))_+\| + \|(-G(\zeta))_+\| \Big),
\end{aligned}$$

[3] However, F and G being jointly monotone seems not equivalent to $\nabla F(\zeta), -\nabla G(\zeta)$ being column monotone for all $\zeta \in \mathbb{R}^n$.

[4] ψ_0^{-1} is well defined since, by (14), ψ_0 is strictly increasing on $[0, \infty)$.

where the second inequality uses Lemma 7(b). Setting $\tau := \dfrac{\rho}{\max\{1, \|F(\zeta^*)\|, \|G(\zeta^*)\|\}}$ yields (38).

Using (13), (14) and (16), we have

$$\max\{0, \langle F(\zeta), G(\zeta)\rangle\} \le \psi_0^{-1}\left(f_{\mathrm{YF}}(\zeta)\right) \quad \text{and} \quad \psi_{\mathrm{FB}}(F(\zeta), G(\zeta)) \le f_{\mathrm{YF}}(\zeta).$$

Using Lemma 8 and the second inequality, we have

$$\begin{aligned} \|(-F(\zeta))_+\| + \|(-G(\zeta))_+\| &\le \sqrt{2}\left(\|(-F(\zeta))_+\|^2 + \|(-G(\zeta))_+\|^2\right)^{1/2} \\ &\le 2\sqrt{2}\,\psi_{\mathrm{FB}}(F(\zeta), G(\zeta))^{1/2} \\ &\le 2\sqrt{2}\, f_{\mathrm{YF}}(\zeta)^{1/2}. \end{aligned}$$

Thus,

$$\max\{0, \langle F(\zeta), G(\zeta)\rangle\} + \|(-F(\zeta))_+\| + \|(-G(\zeta))_+\| \le \psi_0^{-1}\left(f_{\mathrm{YF}}(\zeta)\right) + 2\sqrt{2} f_{\mathrm{YF}}(\zeta)^{1/2}.$$

This together with (38) yields (39).

If in addition F is continuous and $G(\zeta) = \zeta$ for all $\zeta \in \mathbb{R}^n$, then the assumption that the SOCCP has a solution can be dropped from Proposition 5; see, e.g., [9, Proposition 2.2.7]. Also, the exponent 2 in the definition of joint strong monotonicity can be replaced by any $q > 1$, and Proposition 5 would generalize accordingly.

By using Lemma 9 and Proposition 4, we have the following analog of [53, Theorem 4.1] on solution existence and boundedness of the level sets of f_{YF}.

Proposition 6. *Suppose that F and G are differentiable, jointly monotone mappings from $\mathbb{R}^n$ to $\mathbb{R}^n$ satisfying*

$$\lim_{\|\zeta\| \to \infty} \|F(\zeta)\| + \|G(\zeta)\| = \infty. \tag{40}$$

Suppose also that SOCCP is strictly feasible, i.e., there exists $\bar{\zeta} \in \mathbb{R}^n$ such that $F(\bar{\zeta}), G(\bar{\zeta}) \in \mathrm{int}(\mathcal{K}^n)$. Then the level set

$$\mathcal{L}(\gamma) := \{\zeta \in \mathbb{R}^n \mid f_{\mathrm{YF}}(\zeta) \le \gamma\}$$

is bounded for all $\gamma \ge 0$, where f_{YF} is given by (13), (16), (17), *and $\psi_0 : \mathbb{R} \to [0, \infty)$ is a smooth function satisfying* (14). *If in addition F, G satisfy* (36), *then $\mathcal{L}(\gamma) \ne \emptyset$ for all $\gamma \ge 0$.*

Proof. For any $\gamma \ge 0$, if $\{\zeta^k\}_{k=1}^{\infty} \subseteq \mathcal{L}(\gamma)$, then $\{f_{\mathrm{YF}}(\zeta^k)\}$ is bounded and the joint monotonicity of F and G yields

$$\langle F(\zeta^k), G(\bar{\zeta})\rangle + \langle F(\bar{\zeta}), G(\zeta^k)\rangle \le \langle F(\zeta^k), G(\zeta^k)\rangle + \langle F(\bar{\zeta}), G(\bar{\zeta})\rangle, \quad k = 1, 2, \ldots$$

Using this together with Lemma 9 and an argument analogous to the proof of [53, Theorem 4.1], we obtain that $\{\|F(\zeta^k)\| + \|G(\zeta^k)\|\}$ is bounded. Then (40) implies $\{\zeta^k\}$ is bounded. This shows that $\mathcal{L}(\gamma)$ is bounded.

The proof of $\mathcal{L}(\gamma) \ne \emptyset$ uses Proposition 4 and is nearly identical to the proof of [53, Theorem 4.1].

It is straighforward to verify that F, G given by (6) or (7) are jointly monotone. Also, we saw in Section 4 that they satisfy (36). If g is linear or, more generally, $\lim_{\|x\|\to\infty} \|\nabla g(x)\|/\|x\| = 0$, then (40) holds and, by Proposition 6, CSOCP has nonempty bounded optimal primal and dual solution sets whenever it has strictly feasible primal and dual solutions. This result in fact extends to the more general problem where ∇g is replaced by any differentiable monotone mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$. This result also holds when F is differentiable monotone and $G(\zeta) = \zeta$ for all $\zeta \in \mathbb{R}^n$.

6. Preliminary numerical experience

Propositions 2 and 3 show that SOCP and, more generally, CSOCP (5) may be reformulated as the unconstrained minimization of the smooth merit function f_{FB} (or f_{YF}), with F, G given by either (6) or (7). In particular, the merit function has a stationary point if and only if both primal and dual optimal solutions of the CSOCP exist and there is no duality gap. And each stationary point yields primal and dual optimal solutions. Thus, we can solve the CSOCP by applying any unconstrained minimization method to the merit function. In contrast to primal-dual interior-point methods for SOCP, this approach does not require the SOCP or its dual to have an interior feasible solution, and it opens SOCP to solution by unconstrained optimization methods. It also allows non-interior starting points. In this section, we report our preliminary experience with solving SOCP from the DIMACS library and randomly generated CSOCP by this approach. In our tests, we use the merit function f_{FB}. Comparable results are expected with f_{YF}.

We consider F, G given by (7). We evaluate F, G using the Cholesky factorization of AA^T, which is efficient when A is sparse.[5] In particular, given such a factorization $LL^T = AA^T$, we can compute $x = F(\zeta)$ and $y = G(\zeta)$ for each ζ via two (sparse) matrix-vector multiplications and two forward/backward solves:

$$Lu = A\zeta, \quad L^T v = u, \quad w = A^T v, \quad x = d + \zeta - w, \quad y = \nabla g(x) - w.$$

In contrast to interior-point methods, the Cholesky factorization needs to be computed only once, thus allowing f_{FB} and its gradient to be efficiently evaluated. All computer codes are written in Matlab, except for the evaluation of $\phi(x, y)$ and $\nabla\psi_{\mathrm{FB}}(x, y)$, which are more efficiently written in Fortran and called from Matlab as Mex files (since their evaluations require looping N times through each SOC). In fact, coding these evaluations in Fortran instead of Matlab reduced the overall cpu time by a factor of about 10, despite some loss in accuracy which results in higher iteration counts. Cholesky factorization is computed using the Matlab routine `chol`. For the vector d satisfying $Ad = b$, which is effectively the initial x (see below), we compute it as a solution of $\min_d \|Ad - b\|$ using Matlab's least square solver. It would be worthwhile to explore other choices.

For the unconstrained optimization method

$$\zeta^{\mathrm{new}} := \zeta + \alpha\Delta,$$

[5] We also experimented with a version that uses pre-conditioned conjugate gradient method instead of Cholesky factorization, but it did not seem to improve the cpu time significantly. Precomputing and storing the $n \times n$ matrix $A^T(AA^T)^{-1}A$ also did not improve the cpu time, even on problems with dense A.

we compute the direction Δ by either the conjugate gradient (CG) method (using Polak-Ribiere or Fletcher-Reeve updates) or the BFGS method or the limited-memory BFGS (L-BFGS) method, and we compute the stepsize α by the Armijo rule (with 1 as the initial trial stepsize, which is typically accepted) [5, 18, 38]. We do not enforce the Wolfe condition [18, Chapter 2] since it is expensive, requiring an extra gradient evaluation per stepsize. To ensure convergence, we revert to the steepest descent direction $-\nabla f_{\text{FB}}(\zeta)$ whenever the current direction Δ fails to satisfy the sufficient descent condition

$$\nabla f_{\text{FB}}(\zeta)^T \Delta \leq -10^{-5} \|\nabla f_{\text{FB}}(\zeta)\| \|\Delta\|.$$

The initial point is chosen to be $\zeta^{\text{init}} = 0$, so that $x^{\text{init}} = d$ and $y^{\text{init}} = c$. It may be worthwhile to explore other choices. The method terminates when

$$\max\{f_{\text{FB}}(\zeta), |x^T y|\} \leq \texttt{accur}, \tag{41}$$

where `accur` is a user-specified solution accuracy. (The duality gap $|x^T y|$ is added to facilitate comparison with interior-point methods.) The method requires 1 gradient evaluation and at least 1 function evaluation per iteration. This is the dominant computation for CG and L-BFGS.

6.1. Solving SOCP with sparse A

We consider the special case of CSOCP where A is sparse and $g(x) = c^T x$ for some $c \in \mathbb{R}^n$. The test problems are drawn from the DIMACS Implementation Challenge library [39], a collection of nontrivial medium-to-large SOCP arising from applications. In our tests, L-BFGS is found to be clearly superior to CG and BFGS. Thus we focus on L-BFGS from here on. The recommended memory length of 5 [38, Section 9.1] is found to work the best. However, for the scaling matrix $H^0 = \gamma I$, the choice

$$\gamma = \frac{1}{p^T q \cdot q^T q}$$

is found to work better than the four choices used by Liu and Nocedal [32], including the recommended choice of $\gamma = p^T q / q^T q$ [38, p. 226], where $p := \zeta - \zeta^{\text{old}}$ and $q := \nabla f_{\text{FB}}(\zeta) - \nabla f_{\text{FB}}(\zeta^{\text{old}})$. We do not have a good explanation for this. We will refer to the above method as L-BFGS-Merit. We also tested an alternative implementation whereby the public-domain Fortran L-BFGS code of Nocedal (1990 version) [32], with default value of H^0, is called by Matlab as Mex files. Nocedal's code uses a stepsize procedure of Moré and Thuente, which enforces a curvature condition as well as sufficient descent. However, on the DIMACS problems, this alternative implementation requires more iterations and cpu time than L-BFGS-Merit to reach the same solution accuracy.

In our tests on the DIMACS problems, we find that L-BFGS-Merit can solve SOCP to low-medium accuracy (`ac\-cur` $\leq$ 1e-5) fairly fast on problems where n is much bigger than m (in particular, `nb`, `nb_L2`, `nb_L2_bessel`). This can be seen from the cpu times reported in Table 1, comparing L-BFGS-Merit with SeDuMi (Version 1.05) by Jos Sturm [46] with varying termination accuracy. SeDuMi is a primal-dual interior-point code that, in the benchmarking of Mittelmann [36, p. 424], is within a factor of 2

Table 1. Performance of SeDuMi and L-BFGS-Merit on three DIMACS problems. (cpu times are in seconds on a Linux PC cluster, running Matlab 6.1)

Problem		**SeDuMi**	**L-BFGS-Merit**
name	m, n	**iter/cpu** (pars.eps)	**iter/cpu/minxy** (accur)
`nb`	123, 2383	18/12.5 (1e-4)	67/2.0/-4e-4 (1e-4)
		19/13.7 (1e-5)	1042/33.5/-2e-4 (1e-5)
		20/14.2 (1e-6)	> 5000 iters (1e-6)
`nb_L2`	123, 4195	10/14.7 (1e-4)	279/18.0/-8e-5 (1e-4)
		11/16.2 (1e-5)	330/19.7/-9e-6 (1e-5)
		12/17.1 (1e-6)	343/21.6/-5e-7 (1e-6)
`nb_L2_bessel`	123, 2641	9/7.5 (1e-4)	65/2.3/-4e-4 (1e-4)
		11/9.0 (1e-5)	108/3.9/-4e-5 (1e-5)
		13/11.8 (1e-6)	108/3.9/-4e-5 (1e-6)
		15/13.2 (1e-7)	197/6.6/-5e-7 (1e-7)

of being the fastest at solving these problems. `pars.eps` is the user-specified solution accuracy for SeDuMi. Since L-BFGS-Merit does not maintain x and y to be in $\mathcal{K}$, we also report the minimum spectral value of, x and y on termination (minxy). As shown in Table 1, L-BFGS-Merit requires more iterations than SeDuMi but less cpu time per iteration. For accuracy below 1e-6, L-BFGS-Merit is competitive with SeDuMi, but not at higher accuracy. The number of L-BFGS iterations is generally reasonable compared to those reported in [32, 38]. Figure 1 plots the merit function value versus iteration number on the problem `nb`. On the remaining DIMACS problems for which $n < 4m$, L-BFGS-Merit converges, but very slowly. For example, on `nb_L1` (with $m = 915, n = 3176$), the left-hand side of (41) is still at 0.8 after 5000 L-BFGS iterations. Improving the convergence rate of L-BFGS on such problems is a topic for future study.

The above results, though limited, suggests that, for SOCP with $n \gg m$ and low-to-medium solution accuracy, a merit function based method like L-BFGS-Merit *might* provide a viable alternative to interior-point methods. The merit function can also be

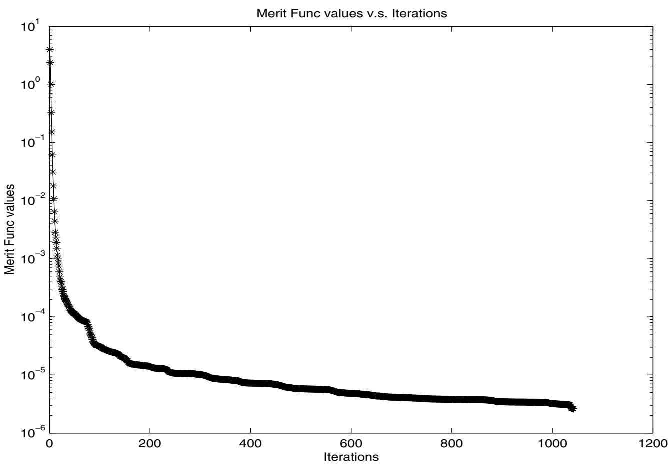

Fig. 1. Plot of Merit function value versus Iteration for L-BFGS-Merit on nb

combined with smoothing and nonsmooth Newton methods to improve the efficiency and robustness of the latter, as was done in the case of NCP [7, 8, 12, 17, 24, 27, 30].

6.2. *Solving CSOCP with dense A*

We consider the special case of CSOCP where A is dense. As we know of no benchmark CSOCP, we generated our own test problems. To make the problems more realistic, we consider a sum-of-norms problem [33, Section 2.2] with a convex regularization term added:

$$\min_{w \geq 0} \sum_{i=1}^{M} \|A_i w - b_i\| + h(w),$$

where $A_i \in \mathbb{R}^{m_i \times \ell}$, $b_i \in \mathbb{R}^{m_i}$, and $h : \mathbb{R}^\ell \to \mathbb{R}$ is a convex twice continuously differentiable function. We transform this problem into the following CSOCP:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{M} z_i + h(w) \\ \text{subject to} & A_i w + s_i = b_i, \quad (z_i, s_i) \in \mathcal{K}^{m_i+1}, \quad i = 1, ..., M, \quad w \geq 0. \end{array}$$

In our tests, we generate each m_i randomly from $\{2, 3, ..., r\}$ ($r \geq 2$), and generate each entry of A_i and b_i randomly according to a uniform distribution from the interval $[-1, 1]$ and $[-5, 5]$, respectively. Thus, the constraint matrix is dense if, say, $\ell \geq m = m_1 + \cdots + m_M$. We use either linear $h(w) = c^T w$ with $c = (1, ..., 1)^T$ (an SOCP) or cubic $h(w) = c^T w + \frac{1}{3}\|w\|_3^3$, where $\|\cdot\|_3$ denotes the 3-norm.

The problem parameters and the performance of L-BFGS-Merit are reported in Table 2. For comparison, we also report the performances of SeDuMi for linear h and of the BFGS and CG methods, referred to as CG-Merit and BFGS-Merit, for cubic h. For termination, `pars.eps` in SeDuMi and `accur` in L-BFGS-Merit, BFGS-Merit, CG-Merit are both set to 1e-3. From Table 2, we see that L-BFGS-Merit is consistently faster than BFGS-Merit and CG-Merit. We also ran the methods with `accur` set to 1e-6, and the same trend is observed, with iteration count and cpu time for L-BFGS-Merit increasing by at most a factor of 2. Although BFGS-Merit has fewer iterations on some problems, its cpu time is higher due to the expensive BFGS update. Interestingly,

Table 2. Performance of SeDuMi, L-BFGS-Merit, BFGS-Merit, CG-Merit on regularized sum-of-norms problems. (cpu times are in seconds on an HP DL360 workstation, running Matlab 6.5.1 under Red Hat Linux 3.3).

		linear h		**cubic** h		
Problem		**SeDuMi**	**L-BFGS-Merit**	**BFGS-Merit**	**CG-Merit**	**L-BFGS-Merit**
ℓ, M, r	m, n	**iter/cpu**	**iter/cpu**	**iter/cpu**	**iter/cpu**	**iter/cpu**
250,10,10	64,324	11/0.3	789/2.8	256/6.1	1344/4.7	427/1.6
250,50,10	312,612	9/2.1	1005/12.2	1197/108.1	13722/186.3	491/8.2
250,10,50	318,578	10/2.2	2144/27.5	1004/84.4	783/112.2	206/3.5
500,10,10	56,566	11/0.5	2548/11.1	352/24.6	1703/6.6	497/2.4
500,50,10	283,833	11/3.8	636/8.6	546/85.1	3173/69.0	700/12.4
500,10,50	246,756	12/3.3	283/3.2	272/36.3	1290/23.0	371/5.6
1000,10,100	611,1621	14/31.1	332/18.0	343/207.8	7561/550.9	317/24.8

L-BFGS-Merit has faster convergence for nonlinear h than for linear h. Perhaps the added cubic term further pushes some components of w towards zero and thus accelerates convergence. For linear h, L-BFGS-Merit is slower than SeDuMi except on the last two problems where ℓ and r are largest. We do not have a good explanation for this. Perhaps this depends on the number of SOC constraints that are active at an optimal solution. Further studies are needed.

In general, the merit function approach seems to be practical for solving CSOCP, especially when g is nonlinear (for which few practical methods exist) and low-accuracy solutions suffice.

7. Conclusions and final remarks

We have shown that, analogous to the NCP case, the SOCCP (1), (2), (3) can be reformulated as an unconstrained smooth minimization problem using the merit function f_{FB} given by (12), (15) or f_{YF} given by (13), (16). Moreover, analogous to the NCP and SDCP cases, if $\nabla F(\zeta)$ and $-\nabla G(\zeta)$ are column monotone, then either ζ is a global minimum of f_{FB} and f_{YF} or it is not a stationary point. In the latter case, if $\nabla G(\zeta)$ is invertible, then a ∇F-free descent direction at ζ can be found. In addition, we give conditions under which f_{YF} has bounded level sets or provides a global error bound on the distance to a solution. Preliminary numerical experience with solving SOCP and CSOCP is reported. As a direction for future research, it would be interesting to extend to SOCCP other NCP merit functions and associated solution methods, such as those surveyed in [47].

For the CSOCP (5), an alternative merit function to f_{FB}, as suggested by one referee, is

$$\hat{f}(x, \lambda) := \psi_{\text{FB}}(x, \nabla g(x) - A^T\lambda) + \frac{1}{2}\|Ax - b\|^2.$$

A drawback of this merit function is that the variables have dimension $n + m$ instead of n. It is also more sensitive to scaling of A and b. Interestingly, this merit function has a similar property as f_{FB} in that every stationary point is a least-square solution of CSOCP. In particular, if (x, λ) is a stationary point of $\hat{f}$, then

$$\begin{aligned} 0 &= \nabla_x \hat{f}(x, \lambda) = \nabla_x \psi_{\text{FB}}(x, \hat{y}) + \nabla^2 g(x)\nabla_y \psi_{\text{FB}}(x, \hat{y}) + A^T(Ax - b), \\ 0 &= \nabla_\lambda \hat{f}(x, \lambda) = A\nabla_y \psi_{\text{FB}}(x, \hat{y}), \end{aligned}$$

where $\hat{y} = \nabla g(x) - A^T\lambda$. Multiplying the first equation on the left by $\nabla_y \psi_{\text{FB}}(x, \hat{y})^T$ yields

$$0 = \nabla_y \psi_{\text{FB}}(x, \hat{y})^T \nabla_x \psi_{\text{FB}}(x, \hat{y}) + \nabla_y \psi_{\text{FB}}(x, \hat{y})^T \nabla^2 g(x)\nabla_y \psi_{\text{FB}}(x, \hat{y}).$$

Using $\nabla^2 g(x) \succeq 0$ and Lemma 6(b), this implies $\psi_{\text{FB}}(x, \hat{y}) = 0$ and hence $\nabla_x \psi_{\text{FB}}(x, \hat{y}) = \nabla_y \psi_{\text{FB}}(x, \hat{y}) = 0$. Then the first equation yields $A^T(Ax - b) = 0$, so $\|Ax - b\|^2$ is at minimum value. Thus, if $Ax = b$ is consistent, then every stationary point of $\hat{f}$ is a primal-dual optimal solution pair of CSOCP.

Finally, since $x \in \mathcal{K}^n$ if and only if $L_x \succeq O$, we might ask whether (1) with $\mathcal{K} = \mathcal{K}^n$ is equivalent to

$$L_x \succeq O, \quad L_y \succeq O, \quad \langle L_x, L_y \rangle = 0, \tag{42}$$

where $\langle A, B \rangle = \mathrm{tr}[A^T B]$ for $A, B \in \mathbb{R}^{n \times n}$. If this were true, then we can construct new merit functions for SOCCP by composing merit functions for SDCP [50, 53] with the *linear* mapping $(x, y) \mapsto (L_x, L_y)$. However, it can be seen that (42) is equivalent to

$$x_1 \geq \|x_2\|, \quad y_1 \geq \|y_2\|, \quad nx_1 y_1 + 2x_2^T y_2 = 0.$$

The two inequalities imply $x_1 y_1 \geq 0$ and $x_1 y_1 + x_2^T y_2 \geq 0$. Then the equality $0 = (n-2)x_1 y_1 + 2(x_1 y_1 + x_2^T y_2)$ yields, for $n > 2$, $x_1 y_1 = x_2^T y_2 = 0$. Thus, for $n > 2$, (42) implies (1) but not conversely. In particular, $x = (1, 1, 0)^T$, $y = (1, -1, 0)^T$ satisfy (1) but not (42).

Acknowledgements. We thank two anonymous referees for their careful reading of this paper and their helpful comments.

References

1. Alizadeh, F., Goldfarb, D.: Second-order cone programming. Math. Program. **95**, 3–51 (2003)
2. Alizadeh, F., Schmieta, S.: Symmetric cones, potential reduction methods, and word-by-word extensions. In: Wolkowicz, H., Saigal, R., Vandenberghe, L., (eds.), Handbook of Semidefinite Programming, Kluwer, Boston, 2000, pp. 195–233
3. Andersen, E.D., Roos, C., Terlaky, T.: On implementing a primal-dual interior-point method for conic quadratic optimization. Math. Program. Ser. B, **95**, 249–277 (2003)
4. Benson, H.Y., Vanderbei, R.J.: Solving problems with semidefinite and related constraints using interior-point methods for nonlinear programming. Math. Program. Ser. B, **95**, 279–302 (2003)
5. Bertsekas, D.P.: Nonlinear Programming. 2nd ed., Athena Scientific, Belmont, 1999
6. Chen, X.-D., Sun, D., Sun, J.: Complementarity functions and numerical experiments for second-order cone complementarity problems. Comput. Optim. Appl. **25**, 39–56 (2003)
7. De Luca, T., Facchinei, F., Kanzow, C.: A semismooth equation approach to the solution of nonlinear complementarity problems. Math. Program. **75**, 407–439 (1996)
8. Facchinei, F., Kanzow, C.: A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems. Math. Program. **76**, 493–512 (1997)
9. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, Volumes I and II. Springer-Verlag, New York, 2003
10. Facchinei, F., Soares, J.: A new merit function for nonlinear complementarity problems and a related algorithm. SIAM J. Optim. **7**, 225–247 (1997)
11. Faraut, U., Korányi, A.: Analysis on Symmetric Cones. Oxford Mathematical Monographs, Oxford University Press, New York, 1994
12. Ferris, M.C., Kanzow, C., Munson, T.S.: Feasible descent algorithms for mixed complementarity problems. Math. Program. **86**, 475–497 (1999)
13. Ferris, M.C., Pang, J.-S., Engineering and economic applications of complementarity problems. SIAM Rev. **39**, 669–713 (1997)
14. Ferris, M.C., Pang, J.-S., (eds.): Complementarity and Variational Problems: State of the Art. SIAM Publications, Philadelphia, 1996
15. Ferris, M.C., Mangasarian, O.L., Pang, J.-S., eds.: Complementarity: Applications, Algorithms and Extensions. Kluwer Academic Publishers, Dordrecht, 2001
16. Fischer, A.: A special Newton-type optimization methods. Optim. **24**, 269–284 (1992)
17. Fischer, A.: Solution of the monotone complementarity problem with locally Lipschitzian functions. Math. Program. **76**, 513–532 (1997)
18. Fletcher, R.: Practical Methods of Optimization. 2nd ed., Wiley-Interscience, Chichester, 1987

19. Fukushima, M., Luo, Z.-Q., Tseng, P.: Smoothing functions for second-order cone complementarity problems. SIAM J. Optim. **12**, 436–460 (2002)
20. Geiger, C., Kanzow, C.: On the resolution of monotone complentarity problems. Comput. Optim. Appl. **5**, 155–173 (1996)
21. Hayashi, S., Yamaguchi, T., Yamashita, N., Fukushima, M.: A matrix splitting method for symmetric affine second-order cone complementarity problems. Report, Department of Applied Mathematics and Physics, Kyoto University, Kyoto, Japan, June 2003; revised February 2004; to appear in J. Comput. Appl. Math.
22. Hayashi, S., Yamashita, N., Fukushima, M.: A combined smoothing and regularization method for monotone second-order cone complementarity problems. SIAM J. Optim. **15**, 593–615 (2005)
23. Isac, G.: Complementarity Problems. Springer-Verlag, Berlin, 1992
24. Jiang, H., Qi, L.: A new nonsmooth equations approach to nonlinear complementarities. SIAM J. Control Optim. **35**, 178–193 (1997)
25. Kanzow, C.: An unconstrained optimization technique for large scale linearly constrained convex minimization problems. Comput. **53**, 101–117 (1994)
26. Kanzow, C.: Nonlinear complementarity as unconstrained optimization. J. Optim. Theory Appl. **88**, 139–155 (1996)
27. Kanzow, C.: Global optimization techniques for mixed complementarity problems. J. Global Optim. **16**, 1–21 (2000)
28. Kanzow, C., Fukushima, M.: Solving box constrained variational inequalities by using the natural residual with D-gap function globalization. Oper. Res. Letters **23**, 45–51 (1998)
29. Kanzow, C., Kleinmichel, H.: A class of Newton-type methods for equality and inequality constrained optimization. Optim. Methods Softw. **5**, 173–198 (1995)
30. Kanzow, C., Pieper, H.: Jacobian smoothing methods for nonlinear complementarity problems. SIAM J. Optim. **9**, 342–373 (1999)
31. Kanzow, C., Yamashita, Y., Fukushima, M.: New NCP functions and their properties. J. Optim. Theory Appl. **97**, 115–135 (1997)
32. Liu, D.C., Nocedal, J.: On the limited memory BFGS method for large scale optimization. Math. Program. **45**, 503–528 (1989)
33. Lobo, M.S., Vandenberghe, L., Boyd, S., Lebret, H.: Application of second-order cone programming. Lin. Algeb. Appl. **284**, 193–228 (1998)
34. Luo, Z.-Q., Tseng, P.: A new class of merit functions for the nonlinear complementarity problem. In: Ferris, M.C., Pang, J.-S., (eds.), Complementarity and Variational Problems: State of the Art, SIAM Publications, Philadelphia, 1997, pp. 204–225
35. Mangasarian, O.L., Solodov, M.V.: Nonlinear complementarity as unconstrained and constrained minimization. Math. Program. **62**, 277–297 (1993)
36. Mittelmann, H.D.: An independent benchmarking of SDP and SOCP solvers. Math. Program. **95**, 407–430 (2003)
37. Monteiro, R.D.C., Tsuchiya, T.: Polynomial convergence of primal-dual algorithms for the second-order cone programs based on the MZ-family of directions. Math. Program. **88**, 61–83 (2000)
38. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer-Verlag, New York, 1999
39. Pataki, G., Schmieta, S.: The DIMACS library of semidefinite-quadratic-linear programs. Preliminary draft, Computational Optimization Research Center, Columbia University, New York, July 2002. http://dimacs.rutgers.edu/Challenges/Seventh/Instances/
40. Peng, J.-M.: Equivalence of variational inequality problems to unconstrained minimization. Math. Program. **78**, 347–355 (1997)
41. Qi, L.: Regular pseudo-smooth NCP and BVIP functions and globally and quadratically convergent generalized Newton methods for complementarity and variational inequality problems. Math. Oper. Res. **24**, 440–471 (1999)
42. Schmieta, S., Alizadeh, F.: Associative and Jordan algebras, and polynomial time interior-point algorithms for symmetric cones. Math. Oper. Res. **26**, 543–564 (2001)
43. Sim, C.-K., Sun, J., Ralph, D.: A note on the Lipschitz continuity of the gradient of the squared norm of the matrix-valued Fischer-Burmeister function. Report, Department of Mathematics, National University of Singapore, Singapore, November 2004; submitted to Math. Program.
44. Sim, C.-K., Zhao, G.: A note on treating a second order cone program as a special case of a semidefinite program. Math. Program. **102**, 609–613 (2005)
45. Solodov, M.V.: Implicit Lagrangian. In: Floudas, C., Pardalos, P., (eds.), Encyclopedia of Optimization. Kluwer Academic Publishers, Dordrecht, 1999
46. Sturm, J.F.: Using Sedumi 1.02, A Matlab* toolbox for optimization over symmetric cones (updated for Version 1.05). Report, Department of Econometrics, Tilburg University, Tilburg, The Netherlands, August 1998–October 2001

47. Sun, D., Qi, L.: On NCP functions. Comput. Optim. Appl. **13**, 201–220 (1999)
48. Sun, D., Sun, J.: Strong semismoothness of Fischer-Burmeister SDC and SOC functions. Math. Program. **103**, to appear (2005)
49. Sun, D., Womersley, R.S.: A new unconstrained differentiable merit function for box constrained variational inequality problems and a damped Gauss-Newton method. SIAM J. Optim. **9**, 388–413 (1999)
50. Tseng, P.: Merit function for semidefinite complementarity problems. Math. Program. **83**, 159–185 (1998)
51. Tseng, P., Yamashita, N., Fukushima, M.: Equivalence of complementarity problems to differentiable minimization: a unified approach. SIAM J. Optim. **6**, 446–460 (1996)
52. Tsuchiya, T.: A convergence analysis of the scaling-invariant primal-dual path-following algorithms for second-order cone programming. Optim. Methods Softw. **11**, 141–182 (1999)
53. Yamashita, N., Fukushima, M.: A new merit function and a descent method for semidefinite complementarity problems. In: Fukushima, M., Qi, L., (eds.), Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, Kluwer Academic Publishers, Boston, 1999, pp. 405–420
54. Yamashita, N., Taji, K., Fukushima, M.: Unconstrained optimization reformulations of variational inequality problems. J. Optim. Theory Appl. **92**, 439–456 (1997)