SIAM J. OPTIM.
Vol. 19, No. 2, pp. 883–910

A CLASS OF INTERIOR PROXIMAL-LIKE ALGORITHMS FOR CONVEX SECOND-ORDER CONE PROGRAMMING*

SHAOHUA PAN† AND JEIN-SHAN CHEN‡

Abstract. We propose a class of interior proximal-like algorithms for the second-order cone program, which is to minimize a closed proper convex function subject to general second-order cone constraints. The class of methods uses a distance measure generated by a twice continuously differentiable strictly convex function on $(0, +\infty)$, and includes as a special case the entropy-like proximal algorithm [Eggermont, *Linear Algebra Appl.*, 130 (1990), pp. 25–42], which was originally proposed for minimizing a convex function subject to nonnegative constraints. Particularly, we consider an approximate version of these methods, allowing the inexact solution of subproblems. Like the entropy-like proximal algorithm for convex programming with nonnegative constraints, we, under some mild assumptions, establish the global convergence expressed in terms of the objective values for the proposed algorithm, and we show that the sequence generated is bounded, and every accumulation point is a solution of the considered problem. Preliminary numerical results are reported for two approximate entropy-like proximal algorithms, and numerical comparisons are also made with the merit function approach [Chen and Tseng, *Math. Program.*, 104 (2005), pp. 293–327], which verify the effectiveness of the proposed method.

Key words. proximal method, measure of distance, second-order cone, second-order cone-convexity

AMS subject classifications. 65K05, 90C30

DOI. 10.1137/070685683

1. Introduction.

We consider the following convex second-order cone programming (CSOCP):

$$\begin{array}{ll} \min & f(\zeta) \\ \text{subject to (s.t.)} & A\zeta + b \succeq_{\mathcal{K}} 0, \end{array} \tag{1}$$

where $f : \mathbb{R}^m \to (-\infty, +\infty]$ is a closed proper convex function; A is an $n \times m$ matrix, with $n \geq m$; b is a vector in $\mathbb{R}^n$; $x \succeq_{\mathcal{K}} 0$ means $x \in \mathcal{K}$; and $\mathcal{K}$ is the Cartesian product of second-order cones (SOCs), also called Lorentz cones [14]. In other words,

$$\mathcal{K} = \mathcal{K}^{n_1} \times \mathcal{K}^{n_2} \times \cdots \times \mathcal{K}^{n_N}, \tag{2}$$

where $N, n_1, \ldots, n_N \geq 1$, $n_1 + n_2 + \cdots + n_N = n$, and

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

with $\|\cdot\|$ denoting the Euclidean norm and $\mathcal{K}^1$ denoting the set of nonnegative reals $\mathbb{R}_+$. The CSOCP, as an extension of the standard second-order cone programming, has a wide range of applications from engineering, control, and finance to robust optimization and combinatorial optimization; see [1, 21, 23] and references therein.

*Received by the editors March 19, 2007; accepted for publication (in revised form) April 25, 2008; published electronically August 13, 2008.
http://www.siam.org/journals/siopt/19-2/68568.html

†School of Mathematical Sciences, South China University of Technology Guangzhou 510640, China (shhpan@cut.edu.cn). This author's work is partially supported by the Doctoral Starting-up Foundation (B13B6050640) of GuangDong Province.

‡Department of Mathematics, National Taiwan Normal University Taipei 11677, Taiwan (jschen@math.ntnu.edu.tw). Member of Mathematics Division, National Center for Theoretical Sciences, Taipei Office. This author's work is partially supported by the National Science Council of Taiwan.

Recently, the second-order cone programming (SOCP) and the SOC complementarity problem have received much attention in optimization. There exist many methods for solving the CSOCP, including the smoothing methods [10, 15], the smoothing-regularization method [17], the semismooth Newton method [22], and the merit function approach [8]. All of these methods are proposed by using some SOC complementarity function or merit function to reformulate the KKT optimality conditions of the CSOCP as a nonsmooth (or smoothing) system of equations or an unconstrained minimization problem. Notice that the CSOCP is a typical convex programming problem which has extensive applications. But, to the best of our knowledge, there are few convex programming methods developed for (or extended to) the CSOCP except the interior point method [33]. Hence, it is worthy to explore other types of convex programming methods for the CSOCP which are different from the aforementioned methods.

One such method is the proximal point algorithm for minimizing a convex function $f(\zeta)$ over $\mathbb{R}^m$, which generates a sequence $\{\zeta^k\}$ by the following iterative scheme:

$$\zeta^k = \operatorname*{argmin}_{\zeta \in \mathbb{R}^m} \left\{ f(\zeta) + \frac{1}{2\mu_k} \|\zeta - \zeta^{k-1}\|^2 \right\}, \tag{3}$$

where μ_k is a sequence of positive numbers. The method was originally introduced by Martinet [24] with the Moreau proximal approximation of f (see [25]), and then further developed by Rockafellar [30, 31]. Later, some researchers [5, 13, 32] proposed and studied nonquadratic proximal point algorithms by replacing the quadratic distance in (3) with a Bregman distance or an entropy-like distance.

The entropy-like proximal algorithm was designed for minimizing a convex function $f(\zeta)$ subject to nonnegative constraints $\zeta \geq 0$. In [12], Eggermont first introduced the Kullback–Leibler relative entropy, defined by

$$[1] d(\zeta, \xi) = \sum_{i=1}^{m} \zeta_i \ln(\zeta_i/\xi_i) + \zeta_i - \xi_i \quad \forall \zeta \geq 0,\ \xi > 0,$$

and established the following entropy-like proximal point algorithm:

$$\begin{cases} \zeta^0 > 0, \\ \zeta^k = \operatorname*{argmin}_{\zeta > 0} \left\{ f(\zeta) + {\mu_k}^{-1} d(\zeta^{k-1}, \zeta) \right\}. \end{cases} \tag{4}$$

Later, Teboulle [32] proposed to replace the usual Kullback–Leibler relative entropy with a new type of distance-like function, called φ-divergence, to define the entropy-like proximal map. Let $\varphi : \mathbb{R} \to (-\infty, +\infty]$ be a closed proper convex function satisfying certain conditions (see [18, 32]). The φ-divergence induced by φ is defined as

$$d_\varphi(\zeta, \xi) := \sum_{i=1}^{m} \xi_i \varphi(\zeta_i/\xi_i). \tag{5}$$

Based on the φ-divergence, Isume et al. [18, 19] generalized Eggermont's algorithm as

$$\begin{cases} \zeta^0 > 0, \\ \zeta^k = \operatorname*{argmin}_{\zeta > 0} \left\{ f(\zeta) + {\mu_k}^{-1} d_\varphi(\zeta, \zeta^{k-1}) \right\}, \end{cases} \tag{6}$$

[1] The convention of $0 \ln 0 = 0$ is used throughout this paper.

and they obtained the convergence theorems under weaker assumptions. Clearly, when

$$\varphi(t) = -\ln t + t - 1 \quad (t > 0),$$

we have that $d_\varphi(\zeta, \xi) = d(\xi, \zeta)$, and consequently the algorithm reduces to Eggermont's.

Observing that the proximal-like algorithm (6) associated with $\varphi(t) = -\ln t + t - 1$ inherits the features of the interior point method as well as the proximal point method, Auslender [2] extended the algorithm to general linearly constrained convex minimization problems and variational inequalities on polyhedra. Then, is it possible to extend the algorithm to nonpolyhedra symmetric conic optimization problems and establish the corresponding convergence results? In this paper, we will explore its extension to the setting of SOCs and establish a class of interior proximal-like algorithms for the CSOCP. We should mention that the algorithm (6) with the entropy function $t \ln t - t + 1$ $(t \geq 0)$ was recently extended to convex semidefinite programming [11].

For simplicity, in the rest of this paper, we focus on the case where $\mathcal{K} = \mathcal{K}^n$. All of the analysis can be carried over to the general case where $\mathcal{K}$ has the direct product structure as (2). It is known that $\mathcal{K}^n$ is a closed convex cone with the interior given by

$$\mathrm{int}(\mathcal{K}^n) := \left\{(x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid x_1 > \|x_2\|\right\}.$$

For any x, y in $\mathbb{R}^n$, we write $x \succeq_{\mathcal{K}^n} y$ if $x - y \in \mathcal{K}^n$; and write $x \succ_{\mathcal{K}^n} y$ if $x - y \in \mathrm{int}(\mathcal{K}^n)$. In other words, we have that $x \succeq_{\mathcal{K}^n} 0$ if and only if $x \in \mathcal{K}^n$ and $x \succ_{\mathcal{K}^n} 0$ if and only if $x \in \mathrm{int}(\mathcal{K}^n)$. We denote $\mathcal{F}$ by the constraint set of the CSOCP, i.e.,

$$\mathcal{F} := \Big\{\zeta \in \mathbb{R}^m \mid A\zeta + b \succeq_{\mathcal{K}^n} 0\Big\}. \tag{7}$$

It is not difficult to verify that $\mathcal{F}$ is convex, and its interior $\mathrm{int}(\mathcal{F})$ is given by

$$\mathrm{int}(\mathcal{F}) := \Big\{\zeta \in \mathbb{R}^m \mid A\zeta + b \succ_{\mathcal{K}^n} 0\Big\}.$$

The proximal-like algorithm that we propose for the CSOCP is defined as follows:

$$\begin{cases} \zeta^0 \in \mathrm{int}(\mathcal{F}), \\ \zeta^k = \operatorname*{argmin}_{\zeta \in \mathrm{int}(\mathcal{F})} \left\{ f(\zeta) + \mu_k^{-1} D(A\zeta + b, A\zeta^{k-1} + b) \right\}, \end{cases} \tag{8}$$

where $D : \mathbb{R}^n \times \mathbb{R}^n \to (-\infty, +\infty]$ is a closed proper convex function generated by a class of twice continuously differentiable strictly convex functions on $(0, +\infty)$, and the specific expression is given in section 3. The class of distance measures, as will be shown in section 3, includes as a special case the natural extension of $d_\varphi(x, y)$, with $\varphi(t) = -\ln t + t - 1$ to the SOCs. For the proximal-like algorithm (8), we particularly consider an approximate version which allows an inexact minimization of the subproblem (8) and establish its global convergence results under some mild assumptions. Numerical results are reported for two approximate entropy-like proximal algorithms, which verify the effectiveness of the proximal method proposed. In addition, numerical comparisons with the merit function approach [8] indicate that the condition number of the Hessian matrix $\nabla^2 f(\zeta)$ has a great influence on the numerical performance of the proximal-like algorithm and the merit function approach, but the former seems to have no direct relation with the dense degree of test problems, but the latter tends to more function evaluations as the density increases.

The outline of this paper is as follows. In section 2, we review some basic concepts and properties associated with SOCs. In section 3, we state the definition of $D(x, y)$ and present some specific examples. Some favorable properties of $D(x, y)$ are investigated in section 4. In section 5, we describe an approximate proximal-like algorithm allowing inexact minimization in (8) and establish the global convergence of the algorithm. In section 6, we report our numerical experiences for the proposed proximal-like algorithm by solving some convex SOCPs. Finally, we conclude this paper in section 7.

Throughout this paper, I represents an identity matrix of suitable dimension, and $\mathbb{R}^n$ denotes the space of n-dimensional real column vectors. For a differentiable function h on $\mathbb{R}$, we denote h', h'', and h''' by its first, second, and third derivative, respectively. Given a set S, we denote $\bar{S}$, $\mathrm{int}(S)$, and $\mathrm{bd}(S)$ by the closure, the interior and the boundary of S, respectively. Note that a function is closed if and only if it is lower semicontinuous, and a function is proper if $f(\zeta) < \infty$ for at least one $\zeta \in \mathbb{R}^m$ and $f(\zeta) > -\infty$ for all $\zeta \in \mathbb{R}^m$. For a closed proper convex function $f : \mathbb{R}^m \to (-\infty, +\infty]$, we denote its domain by $\mathrm{dom} f := \{\ \zeta \in \mathbb{R}^m \mid f(\zeta) < \infty\}$ and the subdifferential of f at $\widehat{\zeta}$ by

$$\partial f(\widehat{\zeta}) := \left\{ w \in \mathbb{R}^m \mid f(\zeta) \geq f(\widehat{\zeta}) + \langle w, \zeta - \widehat{\zeta} \rangle \ \forall \zeta \in \mathbb{R}^m \right\}.$$

If f is differentiable at ζ, the notation $\nabla f(\zeta)$ represents the gradient at ζ of f.

2. Preliminaries. This section recalls some basic concepts and preliminary results related to SOCs that will be used in the subsequent analysis. For any $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ and $y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, we define their *Jordan product* as

$$x \circ y := (\langle x, y \rangle,\ y_1 x_2 + x_1 y_2). \tag{9}$$

We write x^2 to mean $x \circ x$ and write $x + y$ to mean the usual componentwise addition of vectors. Then $\circ$, $+$, and $e = (1, 0, \ldots, 0)^T \in \mathbb{R}^n$ have the following basic properties (see [14, 15]): (1) $e \circ x = x$ for all $x \in \mathbb{R}^n$. (2) $x \circ y = y \circ x$ for all $x, y \in \mathbb{R}^n$. (3) $x \circ (x^2 \circ y) = x^2 \circ (x \circ y)$ for all $x, y \in \mathbb{R}^n$. (4) $(x + y) \circ z = x \circ z + y \circ z$ for all $x, y, z \in \mathbb{R}^n$. The Jordan product is not associative. For example, for $n = 3$, let $x = (1, -1, 1)$ and $y = z = (1, 0, 1)$, then we have that $(x \circ y) \circ z = (4, -1, 4) \neq x \circ (y \circ z) = (4, -2, 4)$. However, it is power associated, i.e., $x \circ (x \circ x) = (x \circ x) \circ x$ for all $x \in \mathbb{R}^n$. Thus, we may, without fear of ambiguity, write x^m for the product of m copies of x and $x^{m+n} = x^m \circ x^n$ for all positive integers m and n. We stipulate that $x^0 = e$. Besides, $\mathcal{K}^n$ is not closed under Jordan product. For example, $x = (1, 1, 0), y = (2, -1, 3) \in \mathcal{K}^n$, but $x \circ y = (1, 1, 3) \notin \mathcal{K}^n$.

For each $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, the *determinant* and the *trace* of x are defined by

$$\det(x) = x_1^2 - \|x_2\|^2, \quad \mathrm{tr}(x) = 2x_1. \tag{10}$$

In general, $\det(x \circ y) \neq \det(x)\det(y)$ unless $x_2 = \alpha y_2$ for some $\alpha \in \mathbb{R}$. A vector $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ is said to be *invertible* if $\det(x) \neq 0$. If x is invertible, then there exists a unique $y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ satisfying $x \circ y = y \circ x = e$. We call this y the inverse of x and denote it by x^{-1}. In fact, we have that

$$x^{-1} = \frac{1}{x_1^2 - \|x_2\|^2}(x_1, -x_2) = \frac{1}{\det(x)}(\operatorname{tr}(x)e - x). \tag{11}$$

Hence, $x \in \operatorname{int}(\mathcal{K}^n)$ if and only if $x^{-1} \in \operatorname{int}(\mathcal{K}^n)$, and $(x^k)^{-1}$ is well-defined if $x \in \operatorname{int}(\mathcal{K}^n)$.

In the following, we recall from [15] 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

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

where $\lambda_i(x)$ and $u_x^{(i)}$ for $i = 1, 2$ are the spectral values and the associated spectral vectors of x, respectively, given by

$$\begin{aligned} &\lambda_i(x) = x_1 + (-1)^i \|x_2\|, \\ &u_x^{(i)} = \begin{cases} \frac{1}{2}\left(1,\ (-1)^i \frac{x_2}{\|x_2\|}\right) & \text{if } x_2 \neq 0; \\ \frac{1}{2}\left(1,\ (-1)^i \bar{x}_2\right) & \text{if } x_2 = 0, \end{cases} \end{aligned} \tag{12}$$

with $\bar{x}_2$ being any vector in $\mathbb{R}^{n-1}$ such that $\|\bar{x}_2\| = 1$. If $x_2 \neq 0$, then the factorization is unique. The spectral decomposition along with the Jordan algebra associated with SOC has some basic properties, whose proofs can be found in [14, 15]. Here, we list four of them that will often be used in the subsequent sections.

PROPERTY 2.1. *For any $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ with the spectral values $\lambda_1(x), \lambda_2(x)$ and spectral vectors $u_x^{(1)}, u_x^{(2)}$ given as in* (12)*, the following results hold:*

(a) *$u_x^{(1)}$ and $u_x^{(2)}$ are orthogonal under Jordan product and have length $1/\sqrt{2}$, i.e.,*

$$u_x^{(1)} \circ u_x^{(2)} = 0, \quad \|u_x^{(1)}\| = \|u_x^{(2)}\| = 1/\sqrt{2}.$$

(b) *$u_x^{(1)}$ and $u_x^{(2)}$ are idempotent under Jordan product, i.e., $u_x^{(i)} \circ u_x^{(i)} = u_x^{(i)}$ for $i = 1, 2$.*

(c) *The determinant, the trace, and the Euclidean norm of x can be denoted by $\lambda_1(x), \lambda_2(x)$:*

$$\det(x) = \lambda_1(x)\lambda_2(x), \quad \operatorname{tr}(x) = \lambda_1(x) + \lambda_2(x), \quad \|x\|^2 = \frac{[\lambda_1(x)]^2 + [\lambda_2(x)]^2}{2}.$$

(d) *$\lambda_1(x)$ are nonnegative (positive) if and only if $x \in \mathcal{K}^n$ ($x \in \operatorname{int}(\mathcal{K}^n)$).*

LEMMA 2.1.

(a) *For any $x \in \mathbb{R}^n$, $x \succeq_{\mathcal{K}^n} 0 \Longleftrightarrow \langle x, y\rangle \geq 0$ for any $y \succeq_{\mathcal{K}^n} 0$.*

(b) *For any $x \in \mathbb{R}^n$, $x \succ_{\mathcal{K}^n} 0 \Longleftrightarrow \langle x, y\rangle > 0$ for any $y \succeq_{\mathcal{K}^n} 0$ and $y \neq 0$.*

(c) *For any $x, y \in \mathbb{R}^n$, let $\lambda_i(x)$ and $\lambda_i(y)$ for $i = 1, 2$ be their spectral values. Then,*

$$\lambda_1(x)\lambda_2(y) + \lambda_2(x)\lambda_1(y) \leq \operatorname{tr}(x \circ y) \leq \lambda_1(x)\lambda_1(y) + \lambda_2(x)\lambda_2(y).$$

Proof. Part (a) is direct by the self-duality of $\mathcal{K}^n$, and we next consider parts (b) and (c).

(b) Let $x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$. The necessity follows from

$$\langle x, y\rangle = x_1y_1 + x_2^Ty_2 \geq x_1y_1 - \|x_2\|\|y_2\| \geq x_1y_1 - y_1\|x_2\| = y_1(x_1 - \|x_2\|) > 0,$$

where the first inequality is by Cauchy–Schwartz, the second is due to $y \succeq_{\mathcal{K}^n} 0$, and the third is since $x \succ_{\mathcal{K}^n} 0$ and $y \neq 0,\ y \succeq_{\mathcal{K}^n} 0$. Next, we prove the sufficiency. First, from $\langle x, y\rangle > 0$ for any $y \succeq_{\mathcal{K}^n} 0$ and $y \neq 0$, we deduce that $x_1 > 0$ by setting $y = e$. If $x_2 = 0$, then the conclusion follows. If $x_2 \neq 0$, then we set $y = (1, -\frac{x_2}{\|x_2\|})$. Clearly, $y \succeq_{\mathcal{K}^n} 0,\ y \neq 0$, and $0 < \langle x, y\rangle = x_1 - \|x_2\| = \lambda_1(x)$. By Property 2.1 (d), we then have $x \succ_{\mathcal{K}^n} 0$.

(c) For any $x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, by (12) we can compute that

$$\lambda_1(x)\lambda_2(y) + \lambda_2(x)\lambda_1(y) = 2x_1y_1 - 2\|x_2\|\|y_2\| \leq 2(x_1y_1 + x_2^Ty_2) = \mathrm{tr}(x \circ y),$$
$$\lambda_1(x)\lambda_1(y) + \lambda_2(x)\lambda_2(y) = 2x_1y_1 + 2\|x_2\|\|y_2\| \geq 2(x_1y_1 + x_2^Ty_2) = \mathrm{tr}(x \circ y)\ .$$

Combining with the two inequalities above then yields the desired result. ◻

For any $h : \mathbb{R} \to \mathbb{R}$, the following vector-valued function was considered in [6, 15]:

$$h^{\mathrm{soc}}(x) = h[\lambda_1(x)] \cdot u_x^{(1)} + h[\lambda_2(x)] \cdot u_x^{(2)} \quad \forall x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}. \tag{13}$$

If h is defined only on a subset of $\mathbb{R}$, then h^{soc} is defined on the corresponding subset of $\mathbb{R}^n$. The definition in (13) is unambiguous whether $x_2 \neq 0$ or $x_2 = 0$. For the vector-valued function h^{soc} induced by h, we have the following results.

LEMMA 2.2. *Given a function $h : \mathrm{I}_{\mathbb{R}} \to \mathbb{R}$, let $h^{\mathrm{soc}} : S \to \mathbb{R}^n$ be the vector-valued function induced by h as in (13), where $\mathrm{I}_{\mathbb{R}} \subseteq \mathbb{R}$ and $S \subseteq \mathbb{R}^n$. Then, the following results hold:*

(a) *For any $x \in S$, $\lambda_i[h^{\mathrm{soc}}(x)] = h[\lambda_i(x)]$ for $i = 1, 2$ and $\mathrm{tr}[h^{\mathrm{soc}}(x)] = \sum_{i=1}^2 h[\lambda_i(x)]$.*

(b) *If h is continuously differentiable on $\mathrm{I}_{\mathbb{R}}$, then h^{soc} is continuously differentiable on the set S, and its transposed Jacobian at $x = (x_1, x_2) \in S$ is given by the formula*

$$\nabla h^{\mathrm{soc}}(x) = h'(x_1)I \tag{14}$$

if $x_2 = 0$, and otherwise

$$\nabla h^{\mathrm{soc}}(x) = \begin{bmatrix} b & c\dfrac{x_2^T}{\|x_2\|} \\ c\dfrac{x_2}{\|x_2\|} & aI + (b-a)\dfrac{x_2x_2^T}{\|x_2\|^2} \end{bmatrix}, \tag{15}$$

where

$$a = \frac{h[\lambda_2(x)] - h[\lambda_1(x)]}{\lambda_2(x) - \lambda_1(x)},\ b = \frac{h'[\lambda_2(x)] + h'[\lambda_1(x)]}{2},\ c = \frac{h'[\lambda_2(x)] - h'[\lambda_1(x)]}{2}.$$

(c) *If h is continuously differentiable on $\mathrm{I}_{\mathbb{R}}$, then $\mathrm{tr}[h^{\mathrm{soc}}(x)]$ is continuously differentiable on the set S, and its gradient $\nabla \mathrm{tr}[h^{\mathrm{soc}}(x)] = 2\nabla h^{\mathrm{soc}}(x) \cdot e = 2(h')^{\mathrm{soc}}(x)$.*

(d) *If h is (strictly) convex on $\mathrm{I}_{\mathbb{R}}$, then $\mathrm{tr}[h^{\mathrm{soc}}(x)]$ is (strictly) convex on the set S.*

Proof. (a) The proof is direct by the definition of h^{soc} and the spectral value.

(b) The conclusion follows directly from [15, Propostion 5.2] or [6, Proposition 4].

(c) Since $\mathrm{tr}[h^{\mathrm{soc}}(x)] = 2\langle h^{\mathrm{soc}}(x), e\rangle$, by part (b) $\mathrm{tr}[h^{\mathrm{soc}}(x)]$ is obviously continuously differentiable. Applying the chain rule for the inner product of two functions yields

$$\nabla \mathrm{tr}[h^{\mathrm{soc}}(x)] = 2\nabla h^{\mathrm{soc}}(x) \cdot e,$$

where $\nabla h^{\rm soc}(x)$ is given by (14)–(15). By a simple computation, it is easy to verify that

$$\nabla h^{\rm soc}(x)\cdot e = h'[\lambda_1(x)]u_x^{(1)} + h'[\lambda_2(x)]u_x^{(2)} = (h')^{\rm soc}(x).$$

Combining the last two equalities immediately gives the second part of the conclusions.

(d) The proof is similar to that of [26, Lemma 3.2 (d)], and so we omit it. ∎

To close this section, we review the definition of SOC-convexity and SOC-monotonicity. The two concepts, such as the matrix-convexity and the matrix-monotonicity in the semidefinite programming, play an important role in the solution methods of SOCPs.

DEFINITION 2.1 (see [7]). *Given a function $h : {\rm I}_{\mathbb{R}} \to \mathbb{R}$, let $h^{\rm soc} : S \to \mathbb{R}^n$ be the vector-valued function defined as in* (13)*, where ${\rm I}_{\mathbb{R}} \subseteq \mathbb{R}$ and $S \subseteq \mathbb{R}^n$. Then,*

(a) *h is said to be SOC-monotone of order n on ${\rm I}_{\mathbb{R}}$ if for any $x, y \in S$,*

$$x \succeq_{\mathcal{K}^n} y \Longrightarrow h^{\rm soc}(x) \succeq_{\mathcal{K}^n} h^{\rm soc}(y).$$

(b) *h is said to be SOC-convex of order n on ${\rm I}_{\mathbb{R}}$ if for any $x, y \in S$ and $0 \le \beta \le 1$,*

$$h^{\rm soc}\Big(\beta x + (1-\beta)y\Big) \preceq_{\mathcal{K}^n} \beta h^{\rm soc}(x) + (1-\beta)h^{\rm soc}(y). \tag{16}$$

We say that h is SOC-convex (respectively, SOC-monotone) on ${\rm I}_{\mathbb{R}}$ if h is SOC-convex of all order n (respectively, SOC-monotone of all order n) on ${\rm I}_{\mathbb{R}}$. A function h is said to be SOC-concave on ${\rm I}_{\mathbb{R}}$ whenever $-h$ is SOC-convex on ${\rm I}_{\mathbb{R}}$. When h is continuous on ${\rm I}_{\mathbb{R}}$, the condition in (16) can be replaced by the more special condition:

$$h^{\rm soc}\left(\frac{x+y}{2}\right) \preceq_{\mathcal{K}^n} \frac{1}{2}\Big(h^{\rm soc}(x) + h^{\rm soc}(y)\Big). \tag{17}$$

Obviously, the set of SOC-monotone functions and the set of SOC-convex functions are both closed under positive linear combinations and under pointwise limits.

3. Distance-like functions in SOCs.

In this section, we present the definition of the distance-like function $D(x, y)$ involved in the proximal-like algorithm (8) and some specific examples. Let $\phi : \mathbb{R} \to (-\infty, +\infty]$ be a closed proper convex function with ${\rm dom}\phi = [0, +\infty)$ and assume that

(C.1) ϕ is strictly convex on its domain.

(C.2) ϕ is twice continuously differentiable on ${\rm int}({\rm dom}\phi)$, with $\lim_{t\to 0^+} \phi''(t) = +\infty$.

(C.3) $\phi'(t)t - \phi(t)$ is convex on ${\rm int}({\rm dom}\phi)$.

(C.4) ϕ' is SOC-concave on ${\rm int}({\rm dom}\phi)$.

In what follows, we denote by Φ the class of functions satisfying Conditions C.1–C.4.

Given a $\phi \in \Phi$, let $\phi^{\rm soc}$ and $(\phi')^{\rm soc}$ be the vector-valued function given as in (13). We define $D(x, y)$ involved in the proximal-like algorithm (8) by

$$D(x,y) := \begin{cases} {\rm tr}\Big[\phi^{\rm soc}(y) - \phi^{\rm soc}(x) - (\phi')^{\rm soc}(x)\circ(y-x)\Big] & \forall x \in {\rm int}(\mathcal{K}^n), y \in \mathcal{K}^n, \\ +\infty & \text{otherwise.} \end{cases} \tag{18}$$

The function, as will be shown in the next section, possesses some favorable properties. Particularly, $D(x, y) \ge 0$ for any $x, y \in {\rm int}(\mathcal{K}^n)$, and $D(x, y) = 0$ if and only if $x = y$. Hence, $D(x, y)$ can be used to measure the distance between the two points in ${\rm int}(\mathcal{K}^n)$.

In the following, we concentrate on the examples of the distance-like function $D(x,y)$. For this purpose, we first give another characterization for Condition C.3.

LEMMA 3.1. *Let* $\phi:\mathbb{R}\to(-\infty,+\infty]$ *be a closed proper function with* $\mathrm{dom}\phi=[0,+\infty)$. *If* ϕ *is thrice continuously differentiable on* $\mathrm{int}(\mathrm{dom}\phi)$*, then* ϕ *satisfies Condition* C.3 *if and only if its derivative function* ϕ' *is exponentially convex,*[2] *or*

$$\phi'(t_1t_2)\le\frac{1}{2}\Big(\phi'(t_1^2)+\phi'(t_2^2)\Big)\quad\forall t_1,t_2>0. \tag{19}$$

Proof. Since the function ϕ is thrice continuously differentiable on $\mathrm{int}(\mathrm{dom}\phi)$, ϕ satisfies Condition C.3 if and only if

$$\phi''(t)+t\phi'''(t)\ge 0\quad(\forall t>0).$$

Observe that the inequality is also equivalent to

$$t\phi''(t)+t^2\phi'''(t)\ge 0\quad(\forall t>0),$$

and hence substituting by $t=\exp(\theta)$ for $\theta\in\mathbb{R}$ into the inequality yields that

$$\exp(\theta)\phi''(\exp(\theta))+\exp(2\theta)\phi'''(\exp(\theta))\ge 0\quad\forall\theta\in\mathbb{R}.$$

Since the left-hand side of this inequality is exactly $[\phi'(\exp(\theta))]''$, it means that $\phi'(\exp(\cdot))$ is convex on $\mathbb{R}$. Consequently, the first part of the conclusions follows.

Note that the convexity of $\phi'(\exp(\cdot))$ on $\mathbb{R}$ is equivalent to saying, for any $\theta_1,\theta_2\in\mathbb{R}$,

$$\phi'(\exp(r\theta_1+(1-r)\theta_2))\le r\phi'(\exp(\theta_1))+(1-r)\phi'(\exp(\theta_2)),\quad r\in[0,1],$$

which, by letting $t_1=\exp(\theta_1)$ and $t_2=\exp(\theta_2)$, can be rewritten as

$$\phi'(t_1^rt_2^{1-r})\le r\phi'(t_1)+(1-r)\phi'(t_2)\ \ \forall t_1,t_2>0 \text{ and } r\in[0,1].$$

This is clearly equivalent to the statement in (19) due to the continuity of ϕ'. ∎

Remark 3.1. The exponential convexity was also used in the definition of the *self-regular* function [27] in which the authors denote Ω by the set of functions whose elements are twice continuously differentiable and exponentially convex on $(0,+\infty)$. By Lemma 3.1, clearly, if $h\in\Omega$, then the function $\int_0^t h(\theta)d\theta$ necessarily satisfies Condition C.3. For example, $\ln t$ belongs to Ω, and so $\int_0^t\ln\theta d\theta=t\ln t$ satisfies Condition C.3.

For the characterizations of the SOC-concavity, interested readers may refer to [7, 9]. Here, we present a lemma which states that the composition of two SOC-concave functions is SOC-concave under some conditions. By this lemma, we may conveniently obtain some new SOC-concave functions from the existing ones.

LEMMA 3.2. *Let* $g:\mathrm{J}_{\mathbb{R}}\to\mathbb{R}$ *and* $h:\mathrm{I}_{\mathbb{R}}\to\mathrm{J}_{\mathbb{R}}$*, where* $\mathrm{J}_{\mathbb{R}}\subseteq\mathbb{R}$ *and* $\mathrm{I}_{\mathbb{R}}\subseteq\mathbb{R}$. *If* g *is SOC-concave and SOC-monotone on* $\mathrm{J}_{\mathbb{R}}$ *and* h *is SOC-concave on* $\mathrm{I}_{\mathbb{R}}$*, then their composition* $g(h(\cdot))$ *is also SOC-concave on* $\mathrm{I}_{\mathbb{R}}$. *If, in addition,* h *is SOC-monotone on* $\mathrm{I}_{\mathbb{R}}$*, then* $g(h(\cdot))$ *is also SOC-monotone on* $\mathrm{I}_{\mathbb{R}}$.

Proof. For the sake of notation, let $g^{\mathrm{soc}}:\widehat{S}\to\mathbb{R}^n$ and $h^{\mathrm{soc}}:S\to\widehat{S}$ be the vector-valued functions associated with g and h, respectively, where $S\subseteq\mathbb{R}^n$ and $\widehat{S}\subseteq\mathbb{R}^n$.

[2] Which means the function $\phi'(\exp(\cdot)):\mathbb{R}\to\mathbb{R}$ is convex on $\mathbb{R}$,

Define $\widehat{g}(t) = g(h(t))$. Then, for any $x \in S$, it follows from (11) and (13) that

$$\begin{aligned} g^{\rm soc}(h^{\rm soc}(x)) &= g^{\rm soc}\Big[h(\lambda_1(x))u_x^{(1)} + h(\lambda_2(x))u_x^{(2)}\Big] \\ &= g[h(\lambda_1(x))]u_x^{(1)} + g[h(\lambda_2(x))]u_x^{(2)} \\ &= \widehat{g}^{\rm soc}(x). \end{aligned} \tag{20}$$

We next prove that $\widehat{g}(t)$ is SOC-concave on $\mathrm{I}_{\mathbb{R}}$. For any $x, y \in S$ and $0 \le \beta \le 1$, from the SOC-concavity of $h(t)$ it follows that

$$h^{\rm soc}(\beta x + (1-\beta)y) \succeq_{\mathcal{K}^n} \beta h^{\rm soc}(x) + (1-\beta)h^{\rm soc}(y).$$

Using the SOC-monotonicity and SOC-concavity of g, we then obtain that

$$\begin{aligned} g^{\rm soc}\Big[h^{\rm soc}(\beta x + (1-\beta)y)\Big] &\succeq_{\mathcal{K}^n} g^{\rm soc}\Big[\beta h^{\rm soc}(x) + (1-\beta)h^{\rm soc}(y)\Big] \\ &\succeq_{\mathcal{K}^n} \beta g^{\rm soc}[h^{\rm soc}(x)] + (1-\beta)g^{\rm soc}[h^{\rm soc}(y)]. \end{aligned}$$

This together with (20) implies that for any $x, y \in S$ and $0 \le \beta \le 1$,

$$\widehat{g}^{\rm soc}\Big(\beta x + (1-\beta)y\Big) \succeq_{\mathcal{K}^n} \beta\widehat{g}^{\rm soc}(x) + (1-\beta)\widehat{g}^{\rm soc}(y).$$

Consequently, the function $\widehat{g}(t)$, i.e., $g(h(\cdot))$ is SOC-concave on $\mathrm{I}_{\mathbb{R}}$. The second part of the conclusions is obvious. ☐

PROPOSITION 3.1. (a) *The function $h(t) = t^r$, with $0 \le r \le 1$ is both SOC-concave and SOC-monotone on* $[0, +\infty)$.

(b) $h(t) = -t^{-r}$, *with* $0 \le r \le 1$ *is SOC-concave and SOC-monotone on* $(0, +\infty)$.

(c) *For all $u \le 0$, $h(t) = \frac{1}{u-t}$ is SOC-concave as well as SOC-monotone on* $(0, +\infty)$.

(d) *The function* $\ln t$ *is SOC-concave and SOC-monotone on* $(0, +\infty)$.

Proof. (a) The proof has been given by [7, Proposition 3.7], and we here omit it.

(b) The conclusion follows directly from [9, Corollary 4.2].

(c) Let $g(t) = -1/t$ and $\widehat{h}(t) = t - u$. Then, $h(t) = 1/(u-t)$ is exactly the composition of the two functions, i.e., $h(t) = g(\widehat{h}(t))$. From part (b), $g(t)$ is SOC-monotone and SOC-concave on $(0, +\infty)$; whereas by [7, Proposition 3.1 (b)] $\widehat{h}(t)$ is SOC-monotone and SOC-concave on $(0, +\infty)$. Thus, applying Lemma 3.2, we readily obtain the conclusion.

(d) The proof can be found in [9]. In view of the importance of $\ln t$, we here present a different proof by following the same line as [3]. Noting that

$$\ln t = \int_{-\infty}^{0} \left[\frac{1}{u-t} - \frac{u}{u^2+1}\right] du \quad (t > 0),$$

we have for any $x \in \mathrm{int}(\mathcal{K}^n)$ that

$$\ln x = \int_{-\infty}^{0} \left[(ue - x)^{-1} - \frac{u}{u^2+1}e\right] du. \tag{21}$$

For any $x = (x_1, x_2),\ y = (y_1, y_2) \in \mathrm{int}(\mathcal{K}^n)$ and any $0 \le \beta \le 1$, let

$$w = \ln(\beta x + (1-\beta)y) - \beta \ln x - (1-\beta)\ln y.$$

Then, by the definition of SOC-concavity, proving the SOC-concavity of $\ln t$ on $(0,+\infty)$ is equivalent to showing that $w \in \mathcal{K}^n$. From (21) and (11), it follows that

$$\begin{aligned}
w &= \int_{-\infty}^{0} \left[(ue-\beta x-(1-\beta)y)^{-1} - \beta(ue-x)^{-1} - (1-\beta)(ue-y)^{-1}\right] du \\
&= \begin{pmatrix} \displaystyle\int_{-\infty}^{0} \left[\frac{u-\beta x_1-(1-\beta)y_1}{\det(ue-\beta x-(1-\beta)y)} - \frac{\beta(u-x_1)}{\det(ue-x)} - \frac{(1-\beta)(u-y_1)}{\det(ue-y)}\right] du \\ \displaystyle\int_{-\infty}^{0} \left[\frac{\beta x_2+(1-\beta)y_2}{\det(ue-\beta x-(1-\beta)y)} - \frac{\beta x_2}{\det(ue-x)} - \frac{(1-\beta)y_2}{\det(ue-y)}\right] du \end{pmatrix} \\
&:= \begin{pmatrix} w_1 \\ w_2 \end{pmatrix},
\end{aligned}$$

where $w_1 \in \mathbb{R}$ and $w_2 \in \mathbb{R}^{n-1}$. However, by Proposition 3.1 (c) and Definition 2.1,

$$\Big(ue-\beta x-(1-\beta)y\Big)^{-1} - \beta(ue-x)^{-1} - (1-\beta)(ue-y)^{-1} \in \mathcal{K}^n,$$

which implies that

$$\frac{u-\beta x_1-(1-\beta)y_1}{\det(ue-\beta x-(1-\beta)y)} - \frac{\beta(u-x_1)}{\det(ue-x)} - \frac{(1-\beta)(u-y_1)}{\det(ue-y)} \geq 0$$

and

$$\begin{aligned}
&\left\|\frac{\beta x_2+(1-\beta)y_2}{\det(ue-\beta x-(1-\beta)y)} - \frac{\beta x_2}{\det(ue-x)} - \frac{(1-\beta)y_2}{\det(ue-y)}\right\| \\
&\leq \frac{u-\beta x_1-(1-\beta)y_1}{\det(ue-\beta x-(1-\beta)y)} - \frac{\beta(u-x_1)}{\det(ue-x)} - \frac{(1-\beta)(u-y_1)}{\det(ue-y)}.
\end{aligned}$$

As a consequence,

$$\begin{aligned}
w_1 &= \int_{-\infty}^{0} \left[\frac{u-\beta x_1-(1-\beta)y_1}{\det(ue-\beta x-(1-\beta)y)} - \frac{\beta(u-x_1)}{\det(ue-x)} - \frac{(1-\beta)(u-y_1)}{\det(ue-y)}\right] du \\
&\geq 0
\end{aligned}$$

and

$$\begin{aligned}
\|w_2\| &\leq \int_{-\infty}^{0} \left\|\left[\frac{\beta x_2+(1-\beta)y_2}{\det(ue-\beta x-(1-\beta)y)} - \frac{\beta x_2}{\det(ue-x)} - \frac{(1-\beta)y_2}{\det(ue-y)}\right]\right\| du \\
&\leq \int_{-\infty}^{0} \left[\frac{u-\beta x_1-(1-\beta)y_1}{\det(ue-\beta x-(1-\beta)y)} - \frac{\beta(u-x_1)}{\det(ue-x)} - \frac{(1-\beta)(u-y_1)}{\det(ue-y)}\right] du \\
&= w_1.
\end{aligned}$$

This shows that $w \in \mathcal{K}^n$, and consequently $\ln t$ is SOC-concave on $(0,+\infty)$. By a similar argument, we can prove that $\ln t$ is SOC-monotone on $(0,+\infty)$. ∎

From Lemma 3.2 and Proposition 3.1, we may obtain the following corollary, which particularly shows that the modified logarithmic barrier function is SOC-concave.

COROLLARY 3.1. (a) *The modified logarithmic barrier function* $\ln(\alpha+t)$ *for* $\alpha > 0$ *is both SOC-concave and SOC-monotone on* $(-\alpha,+\infty)$.

(b) *For any* $\alpha > 0$ *and* $\beta > 0$, *the functions* $\ln(\alpha+\beta t^r)$, *with* $0 \leq r \leq 1$ *are SOC-concave and SOC-monotone on* $[0,+\infty)$.

(c) *For any* $u > 0$, *the functions* $\frac{t}{u+t}$ *are SOC-concave and SOC-monotone on* $(0, +\infty)$.

(d) *For all* $u > 0$, *the functions* $\frac{-1}{\sqrt{u+t}}$ *are SOC-concave and SOC-monotone on* $(-u, +\infty)$.

Proof. (a) The proof is due to Proposition 3.1(d), [7, Proposition 3.1], and Lemma 3.2 by letting $g : (0, +\infty) \to \mathbb{R}$ be $g(t) = \ln t$, and $h : (-a, +\infty) \to (0, +\infty)$ be $h(t) = a + t$.

(b) Let $g : (0, +\infty) \to \mathbb{R}$ be $g(t) = \ln t$, and $h : (0, +\infty) \to (0, +\infty)$ be $h(t) = a + \beta t^r$. The result follows from Proposition 3.1(a), Proposition 3.1(d), and Lemma 3.2.

(c) Let $g : (-1, 0) \to (0, 1)$ be $g(t) = 1 + t$, and $h : (0, +\infty) \to (-1, 0)$ be $h(t) = -u/(u+t)$. Then, we obtain the result from Proposition 3.1(c), [7, Proposition 3.1], and Lemma 3.2. The result also extends the conclusion of [7, Proposition 3.4].

(d) Let $g : (0, +\infty) \to (0, +\infty)$ be $g(t) = \sqrt{t}$, and $h : (-u, +\infty) \to (0, +\infty)$ be $h(t) = u + t$. Then, from Lemma 3.2 it follows that $g(h(t)) = \sqrt{u + t}$ is SOC-concave and SOC-monotone on $(-u, +\infty)$. Using Lemma 3.2 again with $g(t) = -1/t$ and $h(t) = \sqrt{u + t}$, we obtain the desired result. ∎

Now we present serval examples of $D(x, y)$ to close this section. From these examples, we may see that the conditions required by $\phi \in \Phi$ are not so strict, and the construction of the distance-like functions in SOCs can be completed by selecting a class of single variate convex functions.

Example 3.1. Let $\phi(t) = t \ln t - t + 1$ if $t \geq 0$, and $\phi(t) = +\infty$ if $t < 0$. It is easy to verify that ϕ satisfies Conditions C.1–C.3. Also, by Proposition 3.1(d), Condition C.4 also holds. From formula (13), it follows that, for any $y \in \mathcal{K}^n$ and $x \in \text{int}(\mathcal{K}^n)$,

$$\phi^{\text{soc}}(y) = y \circ \ln y - y + e \quad \text{and} \quad (\phi')^{\text{soc}}(x) = \ln x.$$

Consequently, the distance-like function induced by ϕ is given by

$$D_1(x, y) = \text{tr}\,(y \circ \ln y - y \circ \ln x + x - y) \quad \forall x \in \text{int}(\mathcal{K}^n), y \in \mathcal{K}^n.$$

This function is precisely the natural extension of the entropy-like distance $d_\varphi(\cdot, \cdot)$, with $\varphi(t) = -\ln t + t - 1$ to the SOCs. In addition, comparing $D_1(x, y)$ with the distance-like function $H(x, y)$ in Example 3.1 of [26], we note that $D_1(x, y) = H(y, x)$, but the proximal-like algorithms corresponding to them are completely different.

Example 3.2. Let $\phi(t) = t \ln t + (1 + t) \ln(1 + t) - (1 + t) \ln 2$ if $t \geq 0$, and $\phi(t) = +\infty$ if $t < 0$. By computing, we can show that ϕ satisfies Conditions C.1–C.3. Furthermore, from Proposition 3.1(d) and Corollary 3.1(a), we learn that ϕ also satisfies Condition C.4. This means that $\phi \in \Phi$. For any $y \in \mathcal{K}^n$ and $x \in \text{int}(\mathcal{K}^n)$, we can compute that

$$\phi^{\text{soc}}(y) = y \circ \ln y + (e + y) \circ \ln(e + y) - \ln 2(e + y),$$
$$(\phi')^{\text{soc}}(x) = (2 - \ln 2)e + \ln x + \ln(e + x).$$

Therefore, the distance-like function generated by such a ϕ is given by

$$D_2(x, y) = \text{tr}\Big[- \ln(e + x) \circ (e + y) + y \circ (\ln y - \ln x) + (e + y) \circ \ln(e + y) - 2(y - x)\Big]$$

for any $x \in \text{int}(\mathcal{K}^n)$ and $y \in \mathcal{K}^n$. It should be pointed out that $D_2(x, y)$ is not the extension of $d_\varphi(\cdot, \cdot)$, with $\varphi(t) = \phi(t)$ given by [18] to the SOCs.

Example 3.3. Take $\phi(t) = t^{\frac{2r+3}{2}} + t^2$, with $0 \leq r < \frac{1}{2}$ if $t \geq 0$, and $\phi(t) = +\infty$ if $t < 0$. It is easy to verify that ϕ satisfies Conditions C.1–C.3. Furthermore, from

Proposition 3.1(a) it follows that ϕ satisfies Condition C.4. Thus, $\phi \in \Phi$. By a simple computation,

$$\phi^{\rm soc}(y) = y^{\frac{2r+3}{2}} + y^2 \quad \forall y \in \mathcal{K}^n \quad \text{and} \quad (\phi')^{\rm soc}(x) = \frac{2r+3}{2} x^{\frac{2r+1}{2}} + 2x \quad \forall x \in {\rm int}(\mathcal{K}^n).$$

Hence, the distance-like function induced by ϕ has the following expression:

$$D_3(x,y) = {\rm tr}\left[\frac{2r+1}{2} x^{\frac{2r+3}{2}} + x^2 - y \circ \left(\frac{2r+3}{2} x^{\frac{2r+1}{2}} + 2x\right) + y^{\frac{2r+3}{2}} + y^2\right].$$

Example 3.4. Let $\phi(t) = t^{a+1} + at\ln t - at$, with $0 < a \le 1$ if $t \ge 0$, and $\phi(t) = +\infty$ if $t < 0$. It is easily shown that ϕ satisfies Conditions C.1–C.3. By Proposition 3.1(a) and Proposition 3.1(d), ϕ' is SOC-concave on $(0, +\infty)$. Hence, $\phi \in \Phi$. For any $y \in \mathcal{K}^n$ and $x \in {\rm int}(\mathcal{K}^n)$,

$$\phi^{\rm soc}(y) = y^{a+1} + ay \circ \ln y - ay \quad \text{and} \quad (\phi')^{\rm soc}(x) = (a+1)x^a + a\ln x.$$

Consequently, the distance-like function induced by ϕ has the following expression:

$$D_4(x,y) = {\rm tr}\Big[ax^{a+1} + ax - y \circ \Big((a+1)x^a + a\ln x\Big) + y^{a+1} + ay \circ \ln y - ay\Big].$$

4. Properties of distance-like functions. In what follows, we study some favorable properties of the function $D(x,y)$. We begin with two technical lemmas that will be used in the subsequent analysis. Among others, the first lemma is a direct consequence of Lemma 2.2 and the definition of Φ.

LEMMA 4.1. *Given a $\phi \in \Phi$, let $\phi^{\rm soc}$ and $(\phi')^{\rm soc}$ be the vector-valued functions given as in* (13). *Then, we have the following results:*

(a) *$\phi^{\rm soc}(x)$ and $(\phi')^{\rm soc}(x)$ are well-defined on $\mathcal{K}^n$ and ${\rm int}(\mathcal{K}^n)$, respectively, and*

$$\lambda_i[\phi^{\rm soc}(x)] = \phi[\lambda_i(x)], \quad \lambda_i[(\phi')^{\rm soc}(x)] = \phi'[\lambda_i(x)], \quad i = 1, 2.$$

(b) *$\phi^{\rm soc}(x)$ and $(\phi')^{\rm soc}(x)$ are continuously differentiable on ${\rm int}(\mathcal{K}^n)$, with the transposed Jacobian at x given as in formulas* (14)–(15).

(c) *${\rm tr}[\phi^{\rm soc}(x)]$ and ${\rm tr}[(\phi')^{\rm soc}(x)]$ are continuously differentiable on ${\rm int}(\mathcal{K}^n)$, and*

$$\nabla {\rm tr}\Big[\phi^{\rm soc}(x)\Big] = 2\nabla\phi^{\rm soc}(x) \cdot e = 2(\phi')^{\rm soc}(x),$$
$$\nabla {\rm tr}\Big[(\phi')^{\rm soc}(x)\Big] = 2\nabla(\phi')^{\rm soc}(x) \cdot e = 2(\phi'')^{\rm soc}(x). \tag{22}$$

(d) *The function ${\rm tr}[\phi^{\rm soc}(x)]$ is strictly convex on ${\rm int}(\mathcal{K}^n)$.*

LEMMA 4.2. *Given a $\phi \in \Phi$ and a fixed point $z \in \mathbb{R}^n$, let $\phi_z : {\rm int}(\mathcal{K}^n) \to \mathbb{R}$ be given by*

$$\phi_z(x) := {\rm tr}\Big[-z \circ (\phi')^{\rm soc}(x)\Big]. \tag{23}$$

Then, the function $\phi_z(x)$ possesses the following properties:

(a) *$\phi_z(x)$ is continuously differentiable on ${\rm int}(\mathcal{K}^n)$, with $\nabla\phi_z(x) = -2\nabla(\phi')^{\rm soc}(x) \cdot z$.*

(b) *$\phi_z(x)$ is convex over ${\rm int}(\mathcal{K}^n)$ when $z \in \mathcal{K}^n$, and furthermore, it is strictly convex over ${\rm int}(\mathcal{K}^n)$ when $z \in {\rm int}(\mathcal{K}^n)$.*

Proof. (a) Since $\phi_z(x) = -2\langle(\phi')^{\rm soc}(x), z\rangle$ for any $x \in {\rm int}(\mathcal{K}^n)$, we have that $\phi_z(x)$ is continuously differentiable on ${\rm int}(\mathcal{K}^n)$ by Lemma 4.1(c). Moreover, applying the chain rule for the inner product of two functions readily yields $\nabla\phi_z(x) = -2\nabla(\phi')^{\rm soc}(x)\cdot z$.

(b) By the continuous differentiability of $\phi_z(x)$, to prove the convexity of ϕ_z on ${\rm int}(\mathcal{K}^n)$, it suffices to prove the following inequality:

$$\phi_z\left(\frac{x+y}{2}\right) \le \frac{1}{2}\Big(\phi_z(x)+\phi_z(y)\Big) \quad \forall x, y \in {\rm int}(\mathcal{K}^n). \tag{24}$$

By Condition C.4, ϕ' is SOC-concave on $(0, +\infty)$. Therefore, we have that

$$-(\phi')^{\rm soc}\left(\frac{x+y}{2}\right) \preceq_{\mathcal{K}^n} -\frac{1}{2}\Big[(\phi')^{\rm soc}(x)+(\phi')^{\rm soc}(y)\Big],$$

i.e.,

$$(\phi')^{\rm soc}\left(\frac{x+y}{2}\right) - \frac{1}{2}(\phi')^{\rm soc}(x) - \frac{1}{2}(\phi')^{\rm soc}(y) \succeq_{\mathcal{K}^n} 0.$$

Using Lemma 2.1(a) and the fact that $z \in \mathcal{K}^n$, we then obtain that

$$\left\langle z,\ (\phi')^{\rm soc}\left(\frac{x+y}{2}\right) - \frac{1}{2}(\phi')^{\rm soc}(x) - \frac{1}{2}(\phi')^{\rm soc}(y)\right\rangle \ge 0, \tag{25}$$

which in turn implies that

$$\left\langle -z,\ (\phi')^{\rm soc}\left(\frac{x+y}{2}\right)\right\rangle \le \frac{1}{2}\Big\langle -z,\ (\phi')^{\rm soc}(x)\Big\rangle + \frac{1}{2}\Big\langle -z,\ (\phi')^{\rm soc}(y)\Big\rangle.$$

The last inequality is exactly the one in (24). Hence, ϕ_z is convex on ${\rm int}(\mathcal{K}^n)$ for $z \in \mathcal{K}^n$.

To prove the second part of the conclusions, we need only to prove that the inequality in (25) holds strictly for any $x, y \in {\rm int}(\mathcal{K}^n)$ and $x \ne y$. By Lemma 2.1(b), this is also equivalent to proving the vector $(\phi')^{\rm soc}\left(\frac{x+y}{2}\right) - \frac{1}{2}(\phi')^{\rm soc}(x) - \frac{1}{2}(\phi')^{\rm soc}(y)$ is nonzero, since

$$(\phi')^{\rm soc}\left(\frac{x+y}{2}\right) - \frac{1}{2}(\phi')^{\rm soc}(x) - \frac{1}{2}(\phi')^{\rm soc}(y) \in \mathcal{K}^n \quad \text{and} \quad z \in {\rm int}(\mathcal{K}^n).$$

From Condition C.4, it follows that ϕ' is concave on $(0, +\infty)$, since the SOC-concavity implies the concavity. This, together with the strict monotonicity of ϕ', implies that ϕ' is strictly concave on $(0, +\infty)$. Using Lemma 2.2(d), we then have that ${\rm tr}[(\phi')^{\rm soc}(x)]$ is strictly concave on ${\rm int}(\mathcal{K}^n)$. This means that, for any $x, y \in {\rm int}(\mathcal{K}^n)$ and $x \ne y$,

$${\rm tr}\left[(\phi')^{\rm soc}\Big(\frac{x+y}{2}\Big)\right] - \frac{1}{2}{\rm tr}\left[(\phi')^{\rm soc}(x)\right] - \frac{1}{2}{\rm tr}\left[(\phi')^{\rm soc}(y)\right] > 0. \tag{26}$$

In addition, we note that the first element of $(\phi')^{\rm soc}\left(\frac{x+y}{2}\right) - \frac{1}{2}(\phi')^{\rm soc}(x) - \frac{1}{2}(\phi')^{\rm soc}(y)$ is

$$\frac{\phi'\Big(\lambda_1\big(\frac{x+y}{2}\big)\Big) + \phi'\Big(\lambda_2\big(\frac{x+y}{2}\big)\Big)}{2} - \frac{\phi'(\lambda_1(x)) + \phi'(\lambda_2(x))}{4} - \frac{\phi'(\lambda_1(y)) + \phi'(\lambda_2(y))}{4},$$

which, by Property 2.1(c), can be rewritten as

$$\frac{1}{2}\mathrm{tr}\left[(\phi')^{\mathrm{soc}}\Big(\frac{x+y}{2}\Big)\right] - \frac{1}{4}\mathrm{tr}\left[(\phi')^{\mathrm{soc}}(x)\right] - \frac{1}{4}\mathrm{tr}\left[(\phi')^{\mathrm{soc}}(y)\right].$$

This together with (26) shows that $(\phi')^{\mathrm{soc}}\left(\frac{x+y}{2}\right) - \frac{1}{2}(\phi')^{\mathrm{soc}}(x) - \frac{1}{2}(\phi')^{\mathrm{soc}}(y)$ is nonzero for any $x, y \in \mathrm{int}(\mathcal{K}^n)$ and $x \neq y$. Consequently, ϕ_z is strictly convex on $\mathrm{int}(\mathcal{K}^n)$. ◻

Now we are in a position to study the properties of the distance-like function $D(x, y)$.

PROPOSITION 4.1. *Given a $\phi \in \Phi$, let $D(x, y)$ be defined as in* (18). *Then,*

(a) *$D(x, y) \geq 0$ for any $x \in \mathrm{int}(\mathcal{K}^n)$ and $y \in \mathcal{K}^n$, and $D(x, y) = 0$ if and only if $x = y$;*

(b) *for any fixed $y \in \mathcal{K}^n$, $D(\cdot, y)$ is continuously differentiable on $\mathrm{int}(\mathcal{K}^n)$, with*

$$\nabla_x D(x, y) = 2\nabla(\phi')^{\mathrm{soc}}(x) \cdot (x - y); \tag{27}$$

(c) *for any fixed $y \in \mathcal{K}^n$, the function $D(\cdot, y)$ is convex over $\mathrm{int}(\mathcal{K}^n)$, and for any fixed $y \in \mathrm{int}(\mathcal{K}^n)$, $D(\cdot, y)$ is strictly convex over $\mathrm{int}(\mathcal{K}^n)$;*

(d) *for any fixed $y \in \mathrm{int}(\mathcal{K}^n)$, the function $D(\cdot, y)$ is essentially smooth;*

(e) *for any fixed $y \in \mathcal{K}^n$, the level sets $L_D(y, \gamma) := \{x \in \mathrm{int}(\mathcal{K}^n) : D(x, y) \leq \gamma\}$ for all $\gamma \geq 0$ are bounded.*

Proof. (a) By Lemma 4.1(c), for any $x \in \mathrm{int}(\mathcal{K}^n)$ and $y \in \mathcal{K}^n$, we can rewrite $D(x, y)$ as

$$D(x, y) = \mathrm{tr}[\phi^{\mathrm{soc}}(y)] - \mathrm{tr}[\phi^{\mathrm{soc}}(x)] - \langle \nabla\mathrm{tr}[\phi^{\mathrm{soc}}(x)],\ y - x\rangle.$$

Notice that $\mathrm{tr}[\phi^{\mathrm{soc}}(x)]$ is strictly convex on $\mathrm{int}(\mathcal{K}^n)$ by Lemma 4.1(d), and hence $D(x, y) \geq 0$ for any $x \in \mathrm{int}(\mathcal{K}^n)$ and $y \in \mathcal{K}^n$, and $D(x, y) = 0$ if and only if $x = y$.

(b) By Lemma 4.1(b) and Lemma 4.1(c), the functions $\mathrm{tr}[\phi^{\mathrm{soc}}(x)]$ and $\langle(\phi')^{\mathrm{soc}}(x), x\rangle$ are continuously differentiable on $\mathrm{int}(\mathcal{K}^n)$. Noting that, for any $x \in \mathrm{int}(\mathcal{K}^n)$ and $y \in \mathcal{K}^n$,

$$D(x, y) = \mathrm{tr}[\phi^{\mathrm{soc}}(y)] - \mathrm{tr}[\phi^{\mathrm{soc}}(x)] - 2\langle(\phi')^{\mathrm{soc}}(x), y - x\rangle;$$

we then have the continuous differentiability of $D(\cdot, y)$ on $\mathrm{int}(\mathcal{K}^n)$. Furthermore,

$$\begin{aligned}\nabla_x D(x, y) &= -\nabla\mathrm{tr}[\phi^{\mathrm{soc}}(x)] - 2\nabla(\phi')^{\mathrm{soc}}(x) \cdot (y - x) + 2(\phi')^{\mathrm{soc}}(x)\\ &= -2(\phi')^{\mathrm{soc}}(x) + 2\nabla(\phi')^{\mathrm{soc}}(x) \cdot (x - y) + 2(\phi')^{\mathrm{soc}}(x)\\ &= 2\nabla(\phi')^{\mathrm{soc}}(x) \cdot (x - y).\end{aligned}$$

(c) By the definition of ϕ_z given as in (23), $D(x, y)$ can be rewritten as

$$D(x, y) = \mathrm{tr}[(\phi')^{\mathrm{soc}}(x) \circ x - \phi^{\mathrm{soc}}(x)] + \phi_y(x) + \mathrm{tr}[\phi^{\mathrm{soc}}(y)].$$

Thus, to prove the (strict) convexity of $D(\cdot, y)$ on $\mathrm{int}(\mathcal{K}^n)$, it suffices to show that

$$\mathrm{tr}[(\phi')^{\mathrm{soc}}(x) \circ x - \phi^{\mathrm{soc}}(x)] + \phi_y(x)$$

is (strictly) convex on $\mathrm{int}(\mathcal{K}^n)$. Let $\psi : (0, +\infty) \to \mathbb{R}$ be the function defined by

$$\psi(t) := \phi'(t)t - \phi(t). \tag{28}$$

Then, the vector-valued function induced by ψ via (13) is $(\phi')^{\rm soc}(x)\circ x-\phi^{\rm soc}(x)$, i.e.,

$$\psi^{\rm soc}(x)=(\phi')^{\rm soc}(x)\circ x-\phi^{\rm soc}(x). \tag{29}$$

From Condition C.3 and Lemma 2.2(d), it follows that ${\rm tr}[(\phi')^{\rm soc}(x)\circ x-\phi^{\rm soc}(x)]$ is convex over ${\rm int}(\mathcal{K}^n)$. In addition, by Lemma 4.2(b), $\phi_y(x)$ is convex on ${\rm int}(\mathcal{K}^n)$ if $y\in\mathcal{K}^n$, and it is strictly convex if $y\in{\rm int}(\mathcal{K}^n)$. Thus, we get the desired results.

(d) From [29, p. 251] and parts (a)–(b), to prove that $D(\cdot,y)$ is essentially smooth for any fixed $y\in{\rm int}(\mathcal{K}^n)$, it suffices to show that $\|\nabla_x D(x^k,y)\|\to+\infty$ for any $\{x^k\}\subset{\rm int}(\mathcal{K}^n)$, with $x^k\to x\in{\rm bd}(\mathcal{K}^n)$. We next prove the conclusion by the following two cases: $x_1>0$ and $x_1=0$. For the sake of notation, let $x^k=(x_1^k,x_2^k)\in\mathbb{R}\times\mathbb{R}^{n-1}$.

Case 1. $x_1>0$. In this case, $\|x_2\|=x_1>0$, since $x\in{\rm bd}(\mathcal{K}^n)$. Noting that $x^k\to x$, we have that $x_2^k\neq 0$ for all sufficiently large k. From the gradient formula (27),

$$\|\nabla_x D(x^k,y)\|=\|2\nabla(\phi')^{\rm soc}(x^k)\cdot(x^k-y)\|\geq\left|2[\nabla(\phi')^{\rm soc}(x^k)\cdot(x^k-y)]_1\right|, \tag{30}$$

where $[\nabla(\phi')^{\rm soc}(x^k)\cdot(x^k-y)]_1$ denotes the first element of the vector $\nabla(\phi')^{\rm soc}(x^k)\cdot(x^k-y)$. By the gradient formula (15), we can compute that

$$\begin{aligned}2[\nabla(\phi')^{\rm soc}(x^k)\cdot(x^k-y)]_1&=[\phi''(\lambda_2(x^k))+\phi''(\lambda_1(x^k))](x_1^k-y_1)\\&\quad+[\phi''(\lambda_2(x^k))-\phi''(\lambda_1(x^k))]\frac{(x_2^k-y_2)^Tx_2^k}{\|x_2^k\|}\\&=\phi''(\lambda_2(x^k))\left(\lambda_2(x^k)-y_1-y_2^Tx_2^k/\|x_2^k\|\right)\\&\quad-\phi''(\lambda_1(x^k))\left(y_1-y_2^Tx_2^k/\|x_2^k\|-\lambda_1(x^k)\right).\end{aligned} \tag{31}$$

Therefore,

$$\begin{aligned}\left|2[\nabla(\phi')^{\rm soc}(x^k)\cdot(x^k-y)]_1\right|&\geq\left|\phi''(\lambda_1(x^k))\left(y_1-y_2^Tx_2^k/\|x_2^k\|-\lambda_1(x^k)\right)\right|\\&\quad-\left|\phi''(\lambda_2(x^k))\left(\lambda_2(x^k)-y_1-y_2^Tx_2^k/\|x_2^k\|\right)\right|\\&\geq\left|\phi''(\lambda_1(x^k))\right|\cdot\left(\left|y_1-y_2^Tx_2^k/\|x_2^k\|\right|-\lambda_1(x^k)\right)\\&\quad-\left|\phi''(\lambda_2(x^k))\right|\cdot\left|\lambda_2(x^k)-y_1-y_2^Tx_2^k/\|x_2^k\|\right|\\&\geq\left|\phi''(\lambda_1(x^k))\right|\cdot\Big(\lambda_1(y)-\lambda_1(x^k)\Big)\\&\quad-\left|\phi''(\lambda_2(x^k))\right|\cdot\left|\lambda_2(x^k)-y_1-y_2^Tx_2^k/\|x_2^k\|\right|.\end{aligned}$$

Noting that $\lambda_1(x^k)\to\lambda_1(x)=0$, $\lambda_2(x^k)\to\lambda_2(x)>0$, and $\frac{y_2^Tx_2^k}{\|x_2^k\|}\to\frac{y_2^Tx_2}{\|x_2\|}$ as $k\to+\infty$, the second term in the right-hand side of the last inequality converges to a finite value, whereas the first term approaches to $+\infty$, since $|\phi''(\lambda_1(x^k))|\to+\infty$ by Condition C.2 and $\lambda_1(y)-\lambda_1(x^k)\to\lambda_1(y)>0$. This implies that as $k\to+\infty$,

$$\left|2[\nabla(\phi')^{\rm soc}(x^k)\cdot(x^k-y)]_1\right|\to+\infty.$$

Combining with the inequality (30) immediately yields $\|\nabla_x D(x^k,y)\|\to+\infty$.

Case 2. $x_1=0$. In this case, we necessarily have that $x=0$, since $x\in\mathcal{K}^n$. Considering that $x^k\to x$, it then follows that $x_2^k=0$ or $x_2^k>0$ for all sufficiently large k. If $x_2^k=0$ for all sufficiently large k, then from (14) we have that

$$\|\nabla_x D(x^k,y)\|=\|2\phi''(x_1^k)(x^k-y)\|\geq 2|\phi''(x_1^k)|\cdot|x_1^k-y_1|.$$

Since $y_1 > 0$ by $y \in \text{int}(\mathcal{K}^n)$ and $x_1^k \to x_1 = 0$, applying Condition C.2 yields that the right-hand side tends to $+\infty$, and consequently $\|\nabla_x D(x^k, y)\| \to +\infty$ when $k \to +\infty$.

Next, we consider the case that $x_2^k > 0$ for all sufficiently large k. In this case, the inequalities (30)–(31) still hold. By Cauchy–Schwarz inequality,

$$\lambda_2(x^k) - y_1 - y_2^T x_2^k/\|x_2^k\| \geq \lambda_2(x^k) - y_1 - \|y_2\| = \lambda_2(x^k) - \lambda_2(y),$$
$$y_1 - y_2^T x_2^k/\|x_2^k\| - \lambda_1(x^k) \geq y_1 - \|y_2\| - \lambda_1(x^k) = \lambda_1(y) - \lambda_1(x^k).$$

Since $\lambda_1(x^k), \lambda_2(x^k) \to 0$ as $k \to +\infty$ and $\lambda_1(y), \lambda_2(y) > 0$ by $y \in \text{int}(\mathcal{K}^n)$, the last two inequalities imply that

$$\lambda_2(x^k) - y_1 - y_2^T x_2^k/\|x_2^k\| \to -\lambda_2(y) < 0,$$
$$y_1 - y_2^T x_2^k/\|x_2^k\| - \lambda_1(x^k) \to \lambda_1(y) > 0.$$

On the other hand, by Condition C.2, when $k \to +\infty$,

$$\phi''(\lambda_2(x^k)) \to +\infty, \quad \phi''(\lambda_1(x^k)) \to +\infty.$$

The two sides show that the right-hand side of (31) approaches to $-\infty$ as $k \to +\infty$, and consequently, $2|[\nabla(\phi')^{\text{soc}}(x^k) \cdot (x^k - y)]_1| \to +\infty$. Thus, from (30), it follows that $\|\nabla_x D(x^k, y)\| \to +\infty$ as $k \to +\infty$.

(e) From the definition of $D(x, y)$, it follows that, for any $x, y \in \text{int}(\mathcal{K}^n)$,

$$\begin{aligned} D(x,y) &= \text{tr}[\phi^{\text{soc}}(y)] - \text{tr}[\phi^{\text{soc}}(x)] - \text{tr}[(\phi')^{\text{soc}}(x) \circ y] + \text{tr}[(\phi')^{\text{soc}}(x) \circ x] \\ &= \sum_{i=1}^{2} \phi(\lambda_i(y)) - \sum_{i=1}^{2} \phi(\lambda_i(x)) - \text{tr}[(\phi')^{\text{soc}}(x) \circ y] + \text{tr}[(\phi')^{\text{soc}}(x) \circ x], \end{aligned} \tag{32}$$

where the second equality is from Lemma 4.1(a) and Property 2.1(c). Since

$$\begin{aligned} (\phi')^{\text{soc}}(x) \circ x &= \left[\phi'(\lambda_1(x))u_x^{(1)} + \phi'(\lambda_2(x))u_x^{(2)}\right] \circ \left[\lambda_1(x)u_x^{(1)} + \lambda_2(x)u_x^{(2)}\right] \\ &= \phi'(\lambda_1(x))\lambda_1(x)u_x^{(1)} + \phi'(\lambda_2(x))\lambda_2(x)u_x^{(2)}, \end{aligned}$$

we have from Lemma 2.2(a) that

$$\text{tr}[(\phi')^{\text{soc}}(x) \circ x] = \sum_{i=1}^{2} \phi'(\lambda_i(x))\lambda_i(x).$$

In addition, by Lemma 2.1(c) and Lemma 4.1(a), we have that

$$\text{tr}[(\phi')^{\text{soc}}(x) \circ y] \leq \sum_{i=1}^{2} \phi'(\lambda_i(x))\lambda_i(y).$$

Combining the last two inequalities with (32) yields that

$$\begin{aligned} D(x,y) &\geq \sum_{i=1}^{2} \left[\phi(\lambda_i(y)) - \phi(\lambda_i(x)) - \phi'(\lambda_i(x))\lambda_i(y) + \phi'(\lambda_i(x))\lambda_i(x)\right] \\ &= \sum_{i=1}^{2} \left[\phi(\lambda_i(y)) - \phi(\lambda_i(x)) - \phi'(\lambda_i(x))(\lambda_i(y) - \lambda_i(x))\right] \\ &= \sum_{i=1}^{2} d_B(\lambda_i(y), \lambda_i(x)), \end{aligned}$$

where $d_B : \mathbb{R}_+ \times \mathbb{R}_{++} \to \mathbb{R}$ is the function defined by

$$d_B(s,t) = \phi(s) - \phi(t) - \phi'(t)(s-t).$$

This implies that, for any fixed $y \in \mathcal{K}^n$ and $\gamma \geq 0$,

$$L_D(y,\gamma) \subseteq \left\{ x \in \mathrm{int}(\mathcal{K}^n) \;\middle|\; \sum_{i=1}^{2} d_B(\lambda_i(y), \lambda_i(x)) \leq \gamma \right\}. \tag{33}$$

Note that, for any fixed $s \geq 0$, the set $\{t > 0 : d_B(s,t) \leq 0\}$ equals to $\{s\}$ or $\emptyset$, and hence it is bounded. Thus, from [29, Corollary 8.7.1] and Condition C.3, it follows that the level sets $\{t > 0 : d_B(s,t) \leq \gamma\}$ for any fixed $s \geq 0$ are bounded. This together with (33) implies that the level sets $L_D(y,\gamma)$ are bounded for all $\gamma \geq 0$. ∎

PROPOSITION 4.2. *Given a* $\phi \in \Phi$, *let* $D(x,y)$ *be defined as in* (18). *Then, for all* $x, y \in \mathrm{int}(\mathcal{K}^n)$ *and* $z \in \mathcal{K}^n$, *we have the following inequality:*

$$\begin{aligned} D(x,z) - D(y,z) &\geq 2\langle \nabla(\phi')^{\mathrm{soc}}(y) \cdot (z-y),\ y-x \rangle \\ &= 2\langle \nabla(\phi')^{\mathrm{soc}}(y) \cdot (y-x),\ z-y \rangle. \end{aligned} \tag{34}$$

Proof. From the definition of $D(x,y)$ and $\phi_z(x)$ and equality (29), it follows that

$$\begin{aligned} D(x,z) - D(y,z) &= \mathrm{tr}[(\phi')^{\mathrm{soc}}(x) \circ x - \phi^{\mathrm{soc}}(x)] + \phi_z(x) \\ &\quad - \mathrm{tr}[(\phi')^{\mathrm{soc}}(y) \circ y - \phi^{\mathrm{soc}}(y)] - \phi_z(y) \\ &= \mathrm{tr}[\psi^{\mathrm{soc}}(x)] - \mathrm{tr}[\psi^{\mathrm{soc}}(y)] + \phi_z(x) - \phi_z(y) \\ &\geq \langle \nabla \mathrm{tr}[\psi^{\mathrm{soc}}(y)],\ x-y \rangle + \langle \nabla \phi_z(y),\ x-y \rangle \\ &= \langle 2(\psi')^{\mathrm{soc}}(y),\ x-y \rangle - \langle 2\nabla(\phi')^{\mathrm{soc}}(y) \cdot z,\ x-y \rangle, \end{aligned} \tag{35}$$

where the inequality is due to the convexity of $\mathrm{tr}[\psi^{\mathrm{soc}}(x)]$ and $\phi_z(x)$, and the last equality follows from Lemma 2.2(c) and Lemma 4.2(a). From the definition of ψ given as in (28), it is easy to compute that

$$\langle (\psi')^{\mathrm{soc}}(y),\ x-y \rangle = \langle (\phi'')^{\mathrm{soc}}(y) \circ y,\ x-y \rangle. \tag{36}$$

In addition, by the gradient formulas in (14)–(15), we can compute that

$$\nabla(\phi')^{\mathrm{soc}}(y) \cdot y = (\phi'')^{\mathrm{soc}}(y) \circ y, \tag{37}$$

which in turn implies that

$$\begin{aligned} &\langle \nabla(\phi')^{\mathrm{soc}}(y) \cdot z,\ x-y \rangle \\ &= \langle \nabla(\phi')^{\mathrm{soc}}(y) \cdot (y+z-y),\ x-y \rangle \\ &= \langle \nabla(\phi')^{\mathrm{soc}}(y) \cdot y,\ x-y \rangle + \langle \nabla(\phi')^{\mathrm{soc}}(y) \cdot (z-y),\ x-y \rangle \\ &= \langle (\phi'')^{\mathrm{soc}}(y) \circ y,\ x-y \rangle + \langle \nabla(\phi')^{\mathrm{soc}}(y) \cdot (z-y),\ x-y \rangle. \end{aligned}$$

This, together with (36) and (35), yields the first inequality in (34), whereas the second inequality follows from the symmetry of the matrix $\nabla(\phi')^{\mathrm{soc}}(y)$. ∎

Propositions 4.1–4.2 indicate that $D(x,y)$ possesses some favorable properties similar to those for d_φ. In the next section, we will employ these properties to establish the convergence for an approximate version of the proximal-like algorithm (8).

5. Approximate proximal-like algorithm. The proximal-like algorithm described as (8) for the CSOCP consists of a sequence of exact minimization. However, in practical computations, it is impossible to obtain the exact solution of these minimization problems. In this section, we consider an approximate version of this algorithm, which allows the inexact solution of the subproblems (8). Throughout this section, we make the following assumptions for the CSOCP:

(A1) $\inf\{f(\zeta) \mid \zeta \in \mathcal{F}\} := f_* > -\infty$ and $\mathrm{dom} f \cap \mathrm{int}(\mathcal{F}) \neq \emptyset$.

(A2) The matrix A is of maximal rank m.

Remark 5.1. Assumption A1 is elementary for the existence of the solution of the CSOCP. Assumption A2 is common in the solution of the SOCPs, which is clearly satisfied when $\mathcal{F} = \{\zeta \in \mathbb{R}^n \mid \zeta \succeq_{\mathcal{K}^n} 0\}$. Moreover, if we consider the linear SOCP

$$\begin{aligned} &\min\ \bar{c}^T x \\ &\text{s.t. } \bar{A}x = \bar{b}, \quad x \in \mathcal{K}^n, \end{aligned} \tag{38}$$

where $\bar{A} \in \mathbb{R}^{m\times n}$, with $m \leq n$, $\bar{b} \in \mathbb{R}^m$ and $\bar{c} \in \mathbb{R}^n$, the assumption that $\bar{A}$ has full row rank m is standard. Consequently, its dual problem, given by

$$\begin{aligned} &\max\ \bar{b}^T y \\ &\text{s.t. } \bar{c} - \bar{A}^T y \succeq_{\mathcal{K}^n} 0 \end{aligned} \tag{39}$$

satisfies assumption A2. This shows that we can solve the linear SOCP by applying the approximate proximal-like algorithm described below to the dual problem (39). In addition, from Lemma 1 in the appendix, we know that the recession cone of $\mathcal{F}$ is given by $0^+\mathcal{F} = \{d \in \mathbb{R}^m \mid Ad \succeq_{\mathcal{K}^n} 0\}$. This implies that assumption A2 is also satisfied when $\mathcal{F}$ is supposed to be bounded, since its recession cone $0^+\mathcal{F}$ now reduces to zero.

For the sake of notation, in what follows, we denote $\mathcal{D} : \mathrm{int}(\mathcal{F}) \times \mathcal{F} \to \mathbb{R}$ by

$$\mathcal{D}(\zeta, \xi) := D(A\zeta + b, A\xi + b). \tag{40}$$

From Proposition 4.1, we readily obtain the following properties of $\mathcal{D}(\zeta, \xi)$.

LEMMA 5.1. *Let $\mathcal{D}(\zeta, \xi)$ be defined by* (40). *Then, under assumption* A2, *we have that*

- (a) *$\mathcal{D}(\zeta, \xi) \geq 0$ for any $\zeta \in \mathrm{int}(\mathcal{F})$ and $\xi \in \mathcal{F}$, and $\mathcal{D}(\zeta, \xi) = 0$ if and only if $\zeta = \xi$;*
- (b) *the function $\mathcal{D}(\cdot, \xi)$ for any fixed $\xi \in \mathcal{F}$ is continuously differentiable on* $\mathrm{int}(\mathcal{F})$, *with*

$$\nabla_\zeta \mathcal{D}(\zeta, \xi) = 2A^T \nabla(\phi')^{\mathrm{soc}}(A\zeta + b)A(\zeta - \xi); \tag{41}$$

- (c) *for any fixed $\xi \in \mathcal{F}$, the function $\mathcal{D}(\cdot, \xi)$ is convex on* $\mathrm{int}(\mathcal{F})$, *and for any fixed $\xi \in \mathrm{int}(\mathcal{F})$, then $\mathcal{D}(\cdot, \xi)$ is strictly convex over* $\mathrm{int}(\mathcal{F})$;
- (d) *for any fixed $\xi \in \mathrm{int}(\mathcal{F})$, the function $\mathcal{D}(\cdot, \xi)$ is essentially smooth;*
- (e) *for any fixed $\xi \in \mathcal{F}$, the level sets $L(\xi, \gamma) = \left\{\zeta \in \mathrm{int}(\mathcal{F}) : \mathcal{D}(\zeta, \xi) \leq \gamma\right\}$ for all $\gamma \geq 0$ are bounded.*

Now we describe an approximate version of the proximal-like algorithm (APM) (8).

The APM. *Given a starting point $\zeta^0 \in \mathrm{int}(\mathcal{F})$ and constants $\epsilon_k \geq 0$ and $\mu_k > 0$, generate the sequence $\{\zeta^k\} \subset \mathrm{int}(\mathcal{F})$ satisfying*

$$g^k \in \partial_{\epsilon_k} f(\zeta^k), \tag{42}$$

$$\mu_k g^k + \nabla_\zeta \mathcal{D}(\zeta^k, \zeta^{k-1}) = 0, \tag{43}$$

where $\partial_\epsilon f$ represents the ϵ-subdifferential of f.

Remark 5.2. The APM can be regarded as an approximate version of the proximal algorithm (8) in the following sense. From the relation in (42) and the convexity of $\mathcal{D}(\cdot, \xi)$ over $\text{int}(\mathcal{F})$ for any fixed $\xi \in \text{int}(\mathcal{F})$, it follows that, for any $u \in \text{int}(\mathcal{F})$,

$$f(u) \geq f(\zeta^k) + \langle u - \zeta^k, g^k \rangle - \epsilon_k$$

and

$$\mu_k^{-1}\mathcal{D}(u, \zeta^{k-1}) \geq \mu_k^{-1}\mathcal{D}(\zeta^k, \zeta^{k-1}) + \mu_k^{-1}\langle \nabla_\zeta \mathcal{D}(\zeta^k, \zeta^{k-1}),\ u - \zeta^k \rangle.$$

Adding the last two inequalities and using (43) yields that

$$f(u) + \mu_k^{-1}\mathcal{D}(u, \zeta^{k-1}) \geq f(\zeta^k) + \mu_k \mathcal{D}(\zeta^k, \zeta^{k-1}) - \epsilon_k.$$

This implies that

$$\zeta^k \in \epsilon_k - \text{argmin}\left\{ f(\zeta) + \mu_k^{-1}\mathcal{D}(\zeta, \zeta^{k-1}) \right\}, \tag{44}$$

where, for a given function F and $\epsilon \geq 0$, the notation

$$\epsilon - \text{argmin } F(\zeta) := \Big\{ \zeta^* : F(\zeta^*) \leq \inf F(\zeta) + \epsilon \Big\}. \tag{45}$$

In the rest of this section, we focus on the convergence of the APM under assumptions A1 and A2. First, we prove that the APM generates a sequence $\{\zeta^k\} \subset \text{int}(\mathcal{F})$, and consequently the APM is well-defined. This is implied by the following lemma.

LEMMA 5.2. *For any $\xi \in \text{int}(\mathcal{F})$ and $\mu > 0$, we have that the following results hold:*

(a) *The function $F(\cdot) := f(\cdot) + \mu^{-1}\mathcal{D}(\cdot, \xi)$ has bounded level sets under assumption* A1*.*

(b) *If, in addition, assumption* A2 *holds, then there has a unique $\widehat{\zeta} \in \text{int}(\mathcal{F})$ such that*

$$\widehat{\zeta} = \underset{\zeta \in \text{int}(\mathcal{F})}{\text{argmin}} \ \left\{ f(\zeta) + \mu^{-1}\mathcal{D}(\zeta, \xi) \right\}, \tag{46}$$

and moreover, the minimum in the right-hand side is attained at $\widehat{\zeta}$ satisfying

$$-2\mu^{-1}A^T\nabla(\phi')^{\text{soc}}(A\widehat{\zeta} + b)A(\widehat{\zeta} - \xi) \in \partial f(\widehat{\zeta}). \tag{47}$$

Proof. (a) Fix $\xi \in \text{int}(\mathcal{F})$ and $\mu > 0$. By assumption A1 and the nonnegativity of $\mathcal{D}(\zeta, \xi)$, to show that $F(\zeta)$ has bounded level sets, it suffices to show that, for all $\nu \geq f_*$, the level sets $L(\nu) := \{\zeta \in \text{int}(\mathcal{F}) \mid F(\zeta) \leq \nu\}$ are bounded. Notice that $L(\nu) \subseteq L(\xi, \mu(\nu - f_*))$ and $L(\xi, \gamma) := \{\zeta \in \text{int}(\mathcal{F}) \mid \mathcal{D}(\zeta, \xi) \leq \gamma\}$ are bounded for all $\gamma \geq 0$ by Lemma 5.1 (e). Therefore, the sets $L(\nu)$ all $\nu \geq f_*$ are bounded.

(b) By Lemma 5.1(b), $F(\zeta)$ is a closed proper strictly convex function. Hence, if the minimum exists, it must be unique. From part (a), the minimizer $\widehat{\zeta}$ exists, and so it is unique. Under assumption A2, using the gradient formula in (41) and the optimality conditions for (46) then yields that

$$0 \in \partial f(\widehat{\zeta}) + 2\mu^{-1}A^T\nabla(\phi')^{\text{soc}}(A\widehat{\zeta} + b)A(\widehat{\zeta} - \xi) + \partial\delta(\widehat{\zeta} \mid \mathcal{F}), \tag{48}$$

where $\delta(u \mid \mathcal{F}) = 0$ if $u \in \mathcal{F}$ and $+\infty$ otherwise. By Lemma 5.1(c) and [29, Theorem 26.1], we have that $\partial_\zeta \mathcal{D}(\zeta, \xi) = \emptyset$ for all $\zeta \in \text{bd}(\mathcal{F})$. Hence, the relation in (48) implies that $\widehat{\zeta} \in \text{int}(\mathcal{F})$. On the other hand, from [29, p. 226], we know that

$$\partial\delta(u \mid \mathcal{F}) = \{v \in \mathbb{R}^n \mid v \preceq_{\mathcal{K}^n} 0,\ \text{tr}(v \circ u) = 0\}.$$

Using Lemma 2.1, we then obtain $\partial\delta(\widehat{\zeta} \mid \mathcal{F}) = \{0\}$. Thus, the proof is complete. ∎

Next, we investigate the properties of the sequence $\{\zeta^k\}$ generated by the APM.

PROPOSITION 5.1. *Let* $\{\mu_k\}$ *be any sequence of positive numbers and* $\sigma_n = \sum_{k=1}^n \mu_k$. *Let* $\{\zeta^k\}$ *be the sequence generated by the APM. Then,*

(a) $\mu_k[f(\zeta^k) - f(\zeta)] \le \mathcal{D}(\zeta^{k-1}, \zeta) - \mathcal{D}(\zeta^k, \zeta) + \mu_k\epsilon_k$ *for all* $\zeta \in \mathcal{F}$.

(b) $\mathcal{D}(\zeta^k, \zeta) \le D(\zeta^{k-1}, \zeta) + \mu_k\epsilon_k$ *for all* $\zeta \in \mathcal{F}$ *subject to* $f(\zeta) \le f(\zeta^k)$.

(c) $\sigma_n(f(\zeta^n) - f(\zeta)) \le \mathcal{D}(\zeta^0, \zeta) - D(\zeta^n, \zeta) + \sum_{k=1}^n \sigma_k\epsilon_k$ *for all* $\zeta \in \mathcal{F}$.

Proof. (a) For any $\zeta \in \mathcal{F}$, using the definition of the ϵ-subdifferential, we have that

$$f(\zeta) \ge f(\zeta^k) + \langle g^k,\ \zeta - \zeta^k\rangle - \epsilon_k, \tag{49}$$

where $g^k \in \partial_{\epsilon_k} f(\zeta^k)$. However, from (43) and (41), it follows that

$$g^k = -2\mu_k^{-1} A^T \nabla(\phi')^{\mathrm{soc}}(A\zeta^k + b)A(\zeta^k - \zeta^{k-1}).$$

Substituting this g^k into (49), we then obtain that

$$\mu_k[f(\zeta^k) - f(\zeta)] \le 2\Big\langle A^T\nabla(\phi')^{\mathrm{soc}}(A\zeta^k + b)A(\zeta^k - \zeta^{k-1}),\ \zeta - \zeta^k\Big\rangle + \mu_k\epsilon_k.$$

On the other hand, applying Proposition 4.2 at the points $x = A\zeta^{k-1} + b$, $y = A\zeta^k + b$, and $z = A\zeta + b$ and using the definition of $\mathcal{D}(\zeta, \xi)$ given by (40) yields that

$$\mathcal{D}(\zeta^{k-1}, \zeta) - \mathcal{D}(\zeta^k, \zeta) = 2\Big\langle A^T\nabla(\phi')^{\mathrm{soc}}(A\zeta^k + b)A(\zeta^k - \zeta^{k-1}),\ \zeta - \zeta^k\Big\rangle.$$

Combining the last two equations, we immediately obtain the result.

(b) The result follows directly from part (a) for any $\zeta \in \mathcal{F}$ such that $f(\zeta^k) \ge f(\zeta)$.

(c) First, from (44), it follows that

$$\zeta^k \in \epsilon_k - \operatorname{argmin}\left\{f(\zeta) + \mu_k^{-1}\mathcal{D}(\zeta, \zeta^{k-1})\right\}.$$

This implies that, for any $\zeta \in \operatorname{int}(\mathcal{F})$,

$$f(\zeta) + \mu_k^{-1}\mathcal{D}(\zeta, \zeta^{k-1}) \ge f(\zeta^k) + \mu_k^{-1}\mathcal{D}(\zeta^k, \zeta^{k-1}) - \epsilon_k.$$

Setting $\zeta = \zeta^{k-1}$ in this inequality and using Lemma 5.1(d) then yields that

$$f(\zeta^{k-1}) - f(\zeta^k) \ge \mu_k^{-1}\mathcal{D}(\zeta^k, \zeta^{k-1}) - \epsilon_k \ge -\epsilon_k.$$

Multiplying the above inequality by σ_{k-1} and summing over $k = 1, 2, \ldots, n$, we get

$$\sum_{k=1}^n \left[\sigma_{k-1} f(\zeta^{k-1}) - (\sigma_k - \mu_k) f(\zeta^k)\right] \ge -\sum_{k=1}^n \sigma_{k-1}\epsilon_k,$$

which, by noting that $\sigma_k = \mu_k + \sigma_{k-1}$ (with $\sigma_0 \equiv 0$), can be reduced to

$$\sigma_n f(\zeta^n) - \sum_{k=1}^n \mu_k f(\zeta^k) \le \sum_{k=1}^n \sigma_{k-1}\epsilon_k.$$

On the other hand, using part (a) and summing over $k = 1, 2, \ldots, n$, we have that

$$-\sigma_n f(\zeta) + \sum_{k=1}^n \mu_k f(\zeta^k) \le \mathcal{D}(\zeta^0, \zeta) - D(\zeta^n, \zeta) + \sum_{k=1}^n \mu_k\epsilon_k \quad \forall \zeta \in \mathcal{F}.$$

Adding the last two inequalities yields

$$\sigma_n(f(\zeta^n) - f(\zeta)) \le \mathcal{D}(\zeta^0, \zeta) - D(\zeta^n, \zeta) + \sum_{k=1}^{n}(\mu_k + \sigma_{k-1})\epsilon_k,$$

which proves (c) because $\mu_k + \sigma_{k-1} = \sigma_k$. ◻

We are now in a position to prove our main convergence result for the APM. For convenience, we denote the optimal set of the CSOCP by $\mathcal{X} := \{\zeta \mid f(\zeta) = f_*\}$.

PROPOSITION 5.2. *Let $\{\zeta^k\}$ be the sequence generated by the APM and $\sigma_n = \sum_{k=1}^n \mu_k$. Then, under assumptions* A1 *and* A2*, the following results hold.*

(a) *If $\sigma_n \to +\infty$ and $\mu_k^{-1}\sigma_k\epsilon_k \to 0$, then $\lim_{n\to+\infty} f(\zeta^n) \to f_*$.*

(b) *If the optimal set $\mathcal{X} \ne \emptyset$, $\sigma_n \to \infty$ and $\sum_{k=1}^\infty \mu_k\epsilon_k < \infty$, then the sequence ζ^k is bounded, and every accumulation point is a solution of the CSOCP.*

Proof. (a) From Proposition 5.1(c) and the nonnegativity of $\mathcal{D}(\zeta^n, \zeta)$, it follows that

$$f(\zeta^n) - f(\zeta) \le \sigma_n^{-1}\mathcal{D}(\zeta^0, \zeta) + \sigma_n^{-1}\sum_{k=1}^{n}\sigma_k\epsilon_k \quad \forall \zeta \in \mathcal{F}.$$

Taking the limit $\sigma_n \to +\infty$ to the two sides of the last inequality, we immediately have that the first term in the right-hand side goes to zero. In addition, applying Lemma 2 in the appendix with $a_{nk} := \sigma_n^{-1}\mu_k$ if $k \le n$ and $a_{nk} := 0$ otherwise and $u_k := \mu_k^{-1}\sigma_k\epsilon_k$, we obtain the second term in the right-hand side:

$$\sigma_n^{-1}\sum_{k=1}^{n}\sigma_k\epsilon_k = \sum_k a_{nk}u_k \to 0$$

because $\sigma_n \to +\infty$ and $\mu_k^{-1}\sigma_k\epsilon_k \to 0$. Therefore, we have that the

$$\lim_{n\to+\infty} f(\zeta^n) \le f_*.$$

This, together with the fact that $f(\zeta^n) \ge f_*$, implies the desired result.

(b) Suppose that $\zeta^* \in \mathcal{X}$. For any k, we have that $f(\zeta^k) \ge f(\zeta^*)$. From Proposition 5.1(b), it then follows that

$$\mathcal{D}(\zeta^k, \zeta^*) \le \mathcal{D}(\zeta^{k-1}, \zeta^*) + \mu_k\epsilon_k.$$

Since $\sum_{k=1}^\infty \mu_k\epsilon_k < +\infty$, using Lemma 3 in the appendix with $v_k := \mathcal{D}(\zeta^k, \zeta^*) \ge 0$ and $\beta_k := \mu_k\epsilon_k \ge 0$ yields that the sequence $\{\mathcal{D}(\zeta^k, \zeta^*)\}$ converges. Thus, by Proposition 5.1(e), the sequence $\{\zeta^k\}$ is bounded and consequently has an accumulation point. Without any loss of generality, let $\widehat{\zeta} \in \mathcal{F}$ be an accumulation point of $\{\zeta^k\}$. Then $\{\zeta^{k_j}\} \to \widehat{\zeta}$ for some $k_j \to +\infty$. Since f is lower semicontinuous, we get $f(\widehat{\zeta}) = \liminf_{k_j\to\infty} f(\zeta^{k_j})$. On the other hand, $f(\zeta^{k_j}) \to f_*$ by part (a). The two sides imply that $f(\widehat{\zeta}) = f_*$. Therefore, $\widehat{\zeta}$ is a solution of the CSOCP. The proof is thus complete. ◻

6. Numerical experiments. In this section, we present some preliminary numerical results for CSOCPs with a specific version of the concept APM, described as Algorithm 6.1 below, and compare the numerical performance of the algorithm with that of the merit function approach [8]. The purpose of our numerical experiments is to verify the theoretical results obtained in the last section and to illustrate the effectiveness of the proximal-like method proposed.

ALGORITHM 6.1.

Given a sufficiently small $\tau > 0$, a sufficiently large M_0, and constants $\rho > 1$ and $\mu_1 > 0$. Choose a starting point $\zeta^0 \in \text{int}(\mathcal{F})$ and set $k := 1$.

For $k = 1, 2, \cdots$ **until** $\mu_k \geq M_0$ **do**

1. Use an unconstrained minimization method to solve approximately the problem

$$\min_{\zeta \in \mathbb{R}^m} \ F_k(\zeta) := f(\zeta) + \mu_k^{-1} \mathcal{D}(\zeta, \zeta^{k-1}) \tag{50}$$

and obtain a ζ^k such that $\|\nabla f(\zeta^k) + \mu_k^{-1} \nabla_\zeta \mathcal{D}(\zeta^k, \zeta^{k-1})\| \leq \tau$.

2. Set $\mu_{k+1} = \rho \mu_k$ and $k := k + 1$, and then go back to Step 1.

End

Algorithm 6.1 is in fact a special APM with $\epsilon_k = \tau \|\zeta^k - \widehat{\zeta}^k\|$ and $\mu_k = \rho^{k-1}\mu_1$, where $\widehat{\zeta}^k$ is the solution of the subproblem (44), since using the strict convexity of $F_k(\zeta)$, we have that

$$F_k(\widehat{\zeta}^k) \geq F_k(\zeta^k) + \langle \nabla F_k(\zeta^k), \widehat{\zeta}^k - \zeta^k \rangle \geq F_k(\zeta^k) - \tau \|\zeta^k - \widehat{\zeta}^k\|,$$

which implies that

$$\zeta^k \in \epsilon_k - \text{argmin} F_k(\zeta), \ \text{with } \epsilon_k = \tau \|\zeta^k - \widehat{\zeta}^k\|.$$

Furthermore, such ϵ_k and μ_k at least satisfy the assumptions of Proposition 5.2(a), since

$$\sigma_n = \sum_{k=1}^{n} \mu_k \to +\infty, \quad \mu_k^{-1}\sigma_k \to \frac{\rho}{\rho - 1}, \quad \text{and} \quad \epsilon_k \to 0.$$

In our experiments, we employed the entropy-like distance functions from Example 3.1 and Example 3.2, respectively, for Algorithm 6.1. For convenience, let

$$\mathcal{D}_1(\zeta, \xi) := D_1(A\zeta + b, A\xi + b) \ \text{ and } \ \mathcal{D}_2(\zeta, \xi) := D_2(A\zeta + b, A\xi + b).$$

All numerical experiments were done on a personal computer with 2.8GHz CPU and 512MB memory. The computer codes were all written in Matlab 6.5. We chose a limited-memory BFGS method with 5 limited-memory vector-updates [4] to solve the minimization subproblem (50). In addition, we adopted a nonmonotone line search described as in [16] to seek a suitable steplength, i.e., we computed the smallest nonnegative integer l such that

$$F_k(\zeta^k + \beta^l d^k) \leq \mathcal{W}_k + \sigma \beta^l \nabla F_k(\zeta^k)^T d^k,$$

where d^k denotes the search direction at the kth iterate, and $\mathcal{W}_k = \max_{j=k-m_k,\ldots,k} F_k(\zeta^j)$ and where, for a given nonnegative integer $\hat{m}$ and s,

$$m_k = \begin{cases} 0 & \text{if } k \leq s, \\ \min\{m_{k-1} + 1, \hat{m}\} & \text{otherwise.} \end{cases}$$

Unless otherwise stated, we chose $\beta = 0.5$, $\sigma = 10^{-4}$, $\hat{m} = 5$, and $s = 5$ for the nonmontone line search, and the following parameters for Algorithm 6.1:

$$\tau = 10^{-5}, \quad M_0 = 1000, \quad \rho = 10, \quad \text{and} \quad \mu_1 = 1.$$

We applied Algorithm 6.1 for the following quadratic convex SOC program:

$$\begin{aligned} &\min \ \frac{1}{2}\zeta^T M\zeta + q^T\zeta \\ &\text{s.t. } \zeta \succeq_{\mathcal{K}^n} 0, \end{aligned} \tag{51}$$

where $M \in \mathbb{R}^{n\times n}$ is a symmetric positive semidefinite matrix and $q \in \mathbb{R}^n$ is a vector. In the experiment, the matrix M and the vector q were generated as follows: elements of q were chosen randomly from the interval $[-1, 1]$, and M was obtained by setting $M = DD^T$, where D was a sparse matrix with approximately $\mathit{density} \cdot n \cdot n$ nonzero entries, which were chosen from a normal distribution with mean -1 and variance 4. In this procedure, the number of nonzero entries of D is determined so that the nonzero density of M can be approximately estimated. To construct different types of $\mathcal{K}$, we chose n_i and N such that $n_1 + \cdots + n_N = 1000$ and $n_1 = \cdots = n_N = 100$. For each type of $\mathcal{K}$, we have solved 10 test problems with the matrix M of nonzero density 0.5%, 1% and 10%, respectively, and started Algorithm 6.1 from the initial point $\zeta^0 = (\bar{\zeta}^{n_i}, \ldots, \bar{\zeta}^{n_N})$, where $\bar{\zeta}^{n_i} = (2, \omega_i/\|\omega_i\|)$ for $i = 1, 2, \ldots, N$, with $\omega_i \in \mathbb{R}^{n_i-1}$ generated randomly by Matlab's **randn.m**. We also employed the merit function approach [8] to solve these test problems. In other words, we chose the same limited-memory BFGS method as used by Algorithm 6.1 to solve the unconstrained minimization reformulation for the KKT conditions of (51):

$$\min_{\zeta\in\mathbb{R}^n} \Psi_{\text{FB}}(\zeta) := \frac{1}{2}\left\|(\zeta^2 + (M\zeta+q)^2)^{1/2} - \zeta - (M\zeta+q)\right\|^2. \tag{52}$$

For the merit function approach, we used the same starting point ζ^0 as Algorithm 6.1 and terminated the iterates once $\sqrt{2\Psi_{\text{FB}}(\zeta)} \le 10^{-4}$.

The numerical results were listed in Tables 1–3 (see the appendix), where **Rcond** denotes the condition number of the matrix M computed by Matlab's **rcond.m**, **Gap** means the absolute dual gap, i.e., the value of the function $|\zeta^T(M\zeta + q)|$ at the final iteration, **NF** represents the number of function evaluations for $F_k(\zeta)$ or $\Psi_{\text{FB}}(\zeta)$ to solve each problem, which for Algorithm 6.1 is the total sum of the function evaluations used for every subproblem, and **Cpu** represents the CPU time in seconds for solving each problem.

From Tables 1–3, we see that Algorithm 6.1 with $D_1(x, y)$ and $D_2(x, y)$ can solve almost all of the test problems within 10^5 function evaluations, except three test problems in Table 1 for which the merit function approach cannot yield the desired result within 5×10^4 function evaluations, too. Algorithm 6.1 requires more function evaluations than the merit function approach, especially for the problems with the matrix M of nonzero density 0.5% and 1%. This is reasonable since the APM is only a primal algorithm, whereas the merit function approach is a primal-dual one. When comparing Table 1 with Tables 2–3, we find that the condition number of M has a great influence on the numerical performance of Algorithm 6.1 and the merit function approach; for example, the two methods have the worst robustness when the condition number of M equals 0. In addition, from Tables 1–3, it seems that the number of function evaluations of Algorithm 6.1 is not influenced by the nonzero density of

TABLE 1
Numerical results for the matrix M with 0.5% nonzero density.

		$\mathcal{D}_1(\zeta,\xi)$			$\mathcal{D}_2(\zeta,\xi)$			Merit function approach		
N0.	**Rcond**	**Gap**	**Nf**	**Time**	**Gap**	**Nf**	**Time**	**Gap**	**Nf**	**Time**
1	0	2.32e–3	28524	84.8	7.76e–3	37480	127.1	1.72e–4	6034	47.2
2	0	–	$> 10^5$	–	–	$> 10^5$	–	–	>50000	–
3	0	–	$> 10^5$	–	–	$> 10^5$	–	–	>50000	–
4	0	–	$> 10^5$	–	–	$> 10^5$	–	–	>50000	–
5	0	1.62e–3	17436	50.9	4.34e–3	25882	100.9	8.91e–4	957	7.56
6	0	1.36e–3	20516	62.4	3.56e–3	25125	91.2	6.35e–4	883	5.76
7	0	4.79e–3	53555	168.0	1.58e–2	53460	194.3	8.24e–4	1506	12.9
8	0	4.95e–3	79164	236.8	1.64e–2	99687	378.2	1.97e–5	4343	36.0
9	0	1.93e–3	29433	91.2	4.74e–3	31655	112.8	3.98e–4	795	6.43
10	0	6.75e–4	57670	168.0	–	$> 10^5$	–	7.39e–4	1492	11.4

TABLE 2
Numerical results for the matrix M with 1% nonzero density.

		$\mathcal{D}_1(\zeta,\xi)$			$\mathcal{D}_2(\zeta,\xi)$			Merit function approach		
N0.	**Rcond**	**Gap**	**Nf**	**Cpu**	**Gap**	**Nf**	**Cpu**	**Gap**	**Nf**	**Cpu**
1	0	2.65e–3	24966	117.4	5.48e–3	26395	142.4	1.19e–3	1695	23.0
2	1.64e–8	3.44e–3	32540	161.5	6.35e–3	33246	200.4	1.14e–3	1666	23.1
3	8.88e–11	1.94e–3	22785	123.7	3.45e–3	24865	137.9	3.42e–4	1421	18.8
4	2.45e–9	6.89e–3	60638	325.1	1.34e–2	66483	392.9	4.78e–4	2360	35.3
5	6.19e–10	2.97e–3	29612	157.3	5.62e–3	37027	230.3	8.53e–4	1594	24.4
6	5.22e–11	5.05e–3	22620	116.5	1.05e–2	26916	166.1	8.09e–4	1259	20.8
7	7.50e–11	2.09e–3	12419	63.6	3.75e–3	19326	121.7	2.70e–4	1728	21.2
8	4.97e–9	2.63e–3	20661	106.4	5.27e–3	30375	187.9	6.14e–4	1497	19.8
9	5.16e–11	4.32e–3	29157	153.9	7.85e–3	42502	266.1	4.04e–4	1734	24.7
10	1.96e–9	2.48e–3	23804	119.5	4.81e–3	34085	201.2	1.26e–3	1550	22.4

M, but the merit function approach clearly requires more function evaluations as the nonzero density of M increases. Moreover, the merit function approach needs more CPU time at each iteration than Algorithm 6.1 due to an extra multiplication of the matrix M and the vector $\nabla_y \psi_{\rm FB}(\zeta, M\zeta + q)$ involved in the computation of $\nabla \Psi_{\rm FB}(\zeta)$. This accounts for the fact that Algorithm 6.1 is superior to the merit

Table 3
Numerical results for the matrix M with 10% nonzero density.

		$\mathcal{D}_1(\zeta,\xi)$			$\mathcal{D}_2(\zeta,\xi)$			Merit function approach		
N0.	**Rcond**	**Gap**	**Nf**	**Cpu**	**Gap**	**Nf**	**Cpu**	**Gap**	**Nf**	**Cpu**
1	4.99e–9	1.12e–3	21486	671.4	1.37e–3	23255	755.6	5.81e–4	15208	1486.4
2	1.05e–8	1.03e–3	27822	896.8	1.24e–3	36318	1225.4	1.19e–3	21823	2054.5
3	2.63e–9	1.08e–3	32345	1031.8	1.33e–3	18571	603.6	1.57e–3	15770	1495.0
4	9.11e–9	1.03e–3	30856	714.6	1.20e–3	38199	890.0	7.05e-4	14742	1372.5
5	4.68e–10	1.11e–3	26957	869.5	1.34e–3	49756	1713.8	1.52e–3	16949	1691.5
6	1.13e–8	1.08e–3	33621	799.1	1.26e–3	38470	904.0	1.10e–3	17071	1591.9
7	1.20e–9	8.33e–4	19452	623.7	9.76e–4	22501	728.7	1.17e–3	23102	2140.7
8	1.21e–8	1.04e–3	24345	763.0	1.21e-3	33600	1185.9	1.46e–3	16011	1492.9
9	1.26e–8	1.04e-3	35613	821.3	1.22e-3	38723	923.2	1.15e–3	27110	2528.3
10	1.16e–8	8.43e-4	13580	410.8	1.00e-3	33869	1204.9	1.61e–3	17883	1655.3

function approach by the CPU time for the problems with the matrix M of nonzero density 10%.

We also applied Algorithm 6.1 for a nonlinear convex SOCP taken from [17].

Example 6.1. Consider the following nonlinear convex SOCP:

$$\begin{aligned} &\min\ \exp(\zeta_1-\zeta_3)+3(2\zeta_1-\zeta_2)^4+\sqrt{1+(3\zeta_2+5\zeta_3)^2} \\ &\text{s.t.}\ \begin{pmatrix} 4 & 6 & 3 \\ -1 & 7 & -5 \end{pmatrix}\zeta+\begin{pmatrix} -1 \\ 2 \end{pmatrix}\in\mathcal{K}^2,\quad \zeta\in\mathcal{K}^3. \end{aligned} \tag{53}$$

In order to obtain an initial interior point $\zeta^0\in\text{int}(\mathcal{F})$ for Algorithm 6.1, we constructed the following conic optimization problem:

$$\begin{aligned} &\min\ w \\ &A\zeta+b+w\hat{e}\succeq_{\mathcal{K}}0, \\ &-w+w^*\geq 0, \\ &\mathcal{K}=\mathcal{K}^3\times\mathcal{K}^2, \end{aligned} \tag{54}$$

where $w^*\in\mathbb{R}$ is a constant and $\hat{e}=\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}, A=\begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, b=\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$, with

$$A_1=\begin{pmatrix} 4 & 6 & 3 \\ -1 & 7 & 5 \end{pmatrix},\ A_2=I,\ b_1=(-1,2)^T,\ b_2=(0,0,0)^T, e_1=(1,0)^T,\ e_2=(1,0,0)^T.$$

It is easy to see that $\zeta=0,\ w=w_0$ belongs to $\text{int}(\mathcal{F})$ only if $w_0>-\lambda_1(b_i),\ i=1,2$ and $w^*>w_0$, and furthermore, when solving (54) with Algorithm 6.1 from

TABLE 4
Numerical results for Example 6.1.

	$\mathcal{D}_1(\zeta,\xi)$			$\mathcal{D}_2(\zeta,\xi)$		
ζ^0	**fopt**	**Nf**	**Time**	**fopt**	**Nf**	**Time**
$(1.8860, -0.1890, -0.4081)^T$	2.597580	89804	79.65	2.597584	65165	62.46
$(4.3425, 0.0875, -0.2332)^T$	2.597591	40402	32.42	2.597610	35820	31.79
$(4.6972, -0.4294, -1.3931)^T$	2.597587	50889	44.36	2.597601	40404	39.59
$(12.3337, -2.6206, -6.2167)^T$	2.597585	67835	50.06	2.597599	64551	66.98
$(3.7282, 0.2875, 0.2737)^T$	2.597591	30763	26.58	2.597611	26485	24.48

$\zeta = 0,\ w = w_0$, if some iterate (ζ^k, w^k) satisfying $w^k < 0$, then the corresponding ζ^k can be used as the starting point to solve (53). This way can also be used to find the starting interior point ζ^0 when applying Algorithm 6.1 for other problems with the form of (1). We have solved the test problem with Algorithm 6.1 from several starting points. The parameters for Algorithm 6.1 were the same as above except $M_0 = 10000$ and $\tau = 10^{-6}$. The numerical results were listed in Table 4, where **fopt** denotes the objective value at the final iteration. We see that the choice of ζ^0 has an influence on the numerical behavior of Algorithm 6.1.

From Tables 1–4, we may draw the following conclusions: the approximate proximal-like algorithm using $D_1(x, y)$ has the better numerical behavior than the one using $D_2(x, y)$ whether by the accuracy of solution or the number of function evaluations required, and the proximal-like algorithm with an appropriate distance measure is superior to the merit function approach by the CPU time for those dense problems.

7. Conclusions. In this paper, we extended the entropy-like proximal algorithm proposed by Eggermont [12] for convex programming subject to nonnegative constraints and proposed a class of interior proximal-like algorithms for solving the CSOCPs. These algorithms are based on a distance-like function generated by a closed proper convex function ϕ satisfying dom$\phi = [0, +\infty)$ and Conditions (C.1)–(C.4). The given examples illustrated that the conditions required by ϕ are not very stringent. For the proposed proximal-like algorithm, we particularly considered an approximate version which allows inexact minimization steps, and we established the convergence properties under some mild assumptions. Numerical results were also reported for the algorithm with the entropy-like distance functions from Examples 3.1 and 3.2, and we made comparisons with those yielded by the merit function approach [8], which verify the effectiveness of the proposed method.

In our future research works, we will analyze the convergence rate of the proposed algorithms and investigate some practical versions of the algorithms. In addition, we will consider the extension of the class of interior proximal-like algorithms to general convex symmetric cone programming problems. It should be pointed out that the extension is not direct. The main difficulty is how to extend the characterizations of SOC-convexity [7, 9] to the setting of symmetric cones.

Appendix A.

LEMMA 1. *Let $\mathcal{F}$ be the set defined as in* (8). *Then its recession cone $0^+\mathcal{F}$ is given by*

$$0^+\mathcal{F} = \Big\{ d \in \mathbb{R}^m \mid Ad \succeq_{\mathcal{K}^n} 0 \Big\}. \tag{55}$$

Proof. Assume that $d \in \mathbb{R}^m$ such that $Ad \succeq_{\mathcal{K}^n} 0$. Then, for any $\lambda > 0$, $\lambda Ad \succeq_{\mathcal{K}^n} 0$. Considering that $\mathcal{K}^n$ is closed under the "+" operation, we have that, for any $\zeta \in \mathcal{F}$,

$$A(\zeta + \lambda d) + b = (A\zeta + b) + \lambda(Ad) \succeq_{\mathcal{K}^n} 0. \tag{56}$$

By [29, p. 61], this shows that every element in the set of the right-hand side of (55) is a recession direction of $\mathcal{F}$. Consequently, $\{d \in \mathbb{R}^m \mid Ad \succeq_{\mathcal{K}^n} 0\} \subseteq 0^+\mathcal{F}$.

Now take any $d \in 0^+\mathcal{F}$ and $\zeta \in \mathcal{F}$. Then, for any $\lambda > 0$, equation (56) holds. By Property 2.1(d), we then have $\lambda_1\Big[(A\zeta + b) + \lambda Ad\Big] \geq 0$ for any $\lambda > 0$. This implies that $\lambda_1(Ad) \geq 0$, since otherwise letting $\lambda \to +\infty$ and using the fact that

$$\begin{aligned} \lambda_1\Big[(A\zeta + b) + \lambda Ad\Big] &= (A\zeta + b)_1 + \lambda(Ad)_1 - \|(A\zeta + b)_2 + \lambda(Ad)_2\| \\ &\leq (A\zeta + b)_1 + \lambda(Ad)_1 - \Big(\lambda\|(Ad)_2\| - \|(A\zeta + b)_2\|\Big) \\ &= \lambda\lambda_1(Ad) + \lambda_2(A\zeta + b), \end{aligned}$$

we obtain that $\lambda_1[(A\zeta + b) + \lambda Ad] \to -\infty$. Thus, we prove that $Ad \succeq_{\mathcal{K}^n} 0$, and consequently $0^+\mathcal{F} \subseteq \{d \in \mathbb{R}^m \mid Ad \succeq_{\mathcal{K}^n} 0\}$. Combining with the above discussions then yields the result. $\square$

LEMMA 2 (see [20, Theorem 2]). *Let $\{a_{nk}\}$ be a sequence of real numbers satisfying*

(i) $a_{nk} \geq 0 \ \forall n = 1, 2, \ldots,\ k = 1, 2, \ldots$.

(ii) $\sum_{k=1}^{\infty} a_{nk} = 1 \ \forall n = 1, 2, \ldots,$ *and* $\lim_{n\to+\infty} \sum_{k=1}^{n} a_{nk} u_k = u \ \forall k = 1, 2, \ldots$.

If $\{u_k\}$ is a sequence such that $\lim_{k\to+\infty} u_k = u$, *then* $\lim_{k\to+\infty} a_{nk} u_k = u$.

LEMMA 3 (see [28, Chapter 2]). *Let $\{\upsilon_k\}$ and $\{\beta_k\}$ be nonnegative sequences of real numbers satisfying* (i) $\upsilon_{k+1} \leq \upsilon_k + \beta_k$, (ii) $\sum_{k=1}^{\infty} \beta_k < +\infty$. *Then the sequence $\{\upsilon_k\}$ is convergent.*

Acknowledgments. The authors would like to thank two anonymous referees for their helpful comments and suggestions on the revision of this paper. In particular, the influence of the condition number of M is observed by one referee.

REFERENCES

[1] F. ALIZADEH AND D. GOLDFARB, *Second-order cone programming*, Math. Program., 95 (2003), pp. 3–51.

[2] A. AUSLENDER AND M. HADDOU, *An interior-proximal method for convex linearly constrained problems and its extension to variational inequalities*, Math. Program., 71 (1995), pp. 77–100.

[3] J. BRINKHUIS, Z.-Q. LUO, AND S.-Z. ZHANG, *Matrix Convex Functions with Applications to Weighted Centers for Semidefinite Programming*, preprint; The Chinese University of Hong Kong, Hong Kong, 2005.

[4] R. H. BYRD, P. LU, J. NOCEDAL, AND C. ZHU, *A limited memory algorithm for bound constrained optimization*, SIAM J. Sci. Comput., 16 (1995), pp. 1190–1208.

[5] Y. CENSOR AND S. A. ZENIOS, *The proximal minimization algorithm with D-functions*, J. Optim. Theory Appl., 73 (1992), pp. 451–464.

[6] J.-S. Chen, X. Chen, and P. Tseng, *Analysis of nonsmooth vector-valued functions associated with second-order cones*, Math. Program., 101 (2004), pp. 95–117.
[7] J.-S. Chen, *The convex and monotone functions associated with second-order cone*, Optimization, 55 (2006), pp. 363–385.
[8] J.-S. Chen and P. Tseng, *An unconstrained smooth minimization reformulation of the second-order cone complementarity problem*, Math. Program., 104 (2005), pp. 293–327.
[9] J.-S. Chen, X. Chen, S.-H. Pan, and J.-W. Zhang, *The Characterizations of SOC-monotone and SOC-convex Function*, manuscript, 2007.
[10] X.-D. Chen, D. Sun, and J. Sun, *Complementarity functions and numerical experiments for second-order cone complementarity problems*, Comput. Optim. Appl., 25 (2003), pp. 39–56.
[11] M. Doljansky and M. Teboulle, *An interior proximal algorithm and the exponential multiplier method for semidefinite programming*, SIAM J. Optim., 9 (1998), pp. 1–13.
[12] P. P. B. Eggermont, *Multiplicative iterative algorithms for convex programming*, Linear Algebra Appl., 130 (1990), pp. 25–42.
[13] J. Eckstein, *Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming*, Math. Oper. Res., 18 (1993), pp. 202–226.
[14] J. Faraut and A. Korányi, *Analysis on Symmetrc Cones*, Oxford Mathematical Monographs, Oxford University Press, N. Y., 1994.
[15] M. Fukushima, Z.-Q. Luo, and P. Tseng, *Smoothing functions for second-order cone complementarity problems*, SIAM J. Optim., 12 (2002), pp. 436–460.
[16] L. Grippo, F. Lampariello, and S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM J. Numer. Anal., 23 (1986), pp. 707–716.
[17] S. Hayashi, N. Yamashita, and M. Fukushima, *A combined smoothing and regularization method for monotone second-order cone complementarity problems*, SIAM J. Optim., 15 (2005), pp. 593–615.
[18] A. N. Iusem, B. F. Svaiter, and M. Teboulle, *Entropy-like proximal methods in convex programming*, Math. Oper. Res., 19 (1994), pp. 790–814.
[19] A. N. Iusem and M. Teboulle, *Convergence rate analysis of nonquadratic proximal and augmented Lagrangian methods for convex and linear programming*, Math. Oper. Res., 20 (1995), pp. 657–677.
[20] K. Knopp, *Infinite Sequences and Series*, Dover Publications, New York, 1956.
[21] Y.-J. Kuo and H. D. Mittelmann, *Interior point methods for second-order cone programming and OR applications*, Comput. Optim. Appl., 28 (2004), pp. 255–285.
[22] C. Kanzow, I. Ferenczi, and M. Fukushima, *Semismooth methods for linear and nonlinear second-order cone programs*, Technical report, Department of Applied Mathematics and Physics, Kyoto University, Kyoto, Japan, 2006.
[23] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, *Application of second-order cone programming*, Linear Algebra Appl., 284 (1998), pp. 193–228.
[24] B. Martinet, *Perturbation des methodes d'Optimisation*, Appl. RAIRO Numer. Anal., 12 (1978), pp. 153–171.
[25] J. J. Moreau, *Promimité et Dualité dans un Espace Hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.
[26] S.-H. Pan and J.-S. Chen, *A proximal-like algorithnm using quasi D-function for convex second-order cone programming*, J. Optim. Theory Appl., 138 (2008), pp. 95–113.
[27] J. M. Peng, C. Roos, and T. Terlaky, *Self-regular functions and new search directions for linear and semi-definite optimization*, Math. Program., 93 (2002), pp. 129–171.
[28] B. T. Polyak, *Introduction to Optimization*, Optimization Software, Inc., New York, 1987.
[29] R. T. Rockafella, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[30] R. T. Rockafella, *Augmented Lagrangians and applications of proximial point algorithm in convex programming*, Math. Oper. Res., 1 (1976), pp. 97–116.
[31] R. T. Rockafella, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.
[32] M. Teboulle, *Entropic proximal mappings with applications to nonlinear programming*, Math. Oper. Res., 17 (1992), pp. 670–690.
[33] A. Yoshise, *Interior point trajectories and a homogeneous model for nonlinear complementarity problems over symmetric cones*, SIAM J. Optim., 17 (2006), pp. 1129–1153.