J Optim Theory Appl (2008) 138: 95–113
DOI 10.1007/s10957-008-9380-8

Proximal-Like Algorithm Using the Quasi D-Function for Convex Second-Order Cone Programming

S.H. Pan · J.S. Chen

Published online: 12 April 2008

Abstract In this paper, we present a measure of distance in a second-order cone based on a class of continuously differentiable strictly convex functions on $\mathbb{R}_{++}$. Since the distance function has some favorable properties similar to those of the D-function (Censor and Zenios in J. Optim. Theory Appl. 73:451–464 1992), we refer to it as a quasi D-function. Then, a proximal-like algorithm using the quasi D-function is proposed and applied to the second-cone programming problem, which is to minimize a closed proper convex function with general second-order cone constraints. Like the proximal point algorithm using the D-function (Censor and Zenios in J. Optim. Theory Appl. 73:451–464 1992; Chen and Teboulle in SIAM J. Optim. 3:538–543 1993), under some mild assumptions we establish the global convergence of the algorithm expressed in terms of function values; we show that the sequence generated by the proposed algorithm is bounded and that every accumulation point is a solution to the considered problem.

Keywords Bregman functions · Quasi D-functions · Proximal-like methods · Convex second-order cone programming

Communicated by M. Fukushima.

Research of Shaohua Pan was partially supported by the Doctoral Starting-up Foundation (B13B6050640) of GuangDong Province.
Jein-Shan Chen is a member of the Mathematics Division, National Center for Theoretical Sciences, Taipei Office. The author's work was partially supported by National Science Council of Taiwan.

S.H. Pan
School of Mathematical Sciences, South China University of Technology, Guangzhou 510640, China
e-mail: shhpan@scut.edu.cn

J.S. Chen (✉)
Department of Mathematics, National Taiwan Normal University, Taipei 11677, Taiwan
e-mail: jchen@math.ntnu.edu.tw

1 Introduction

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

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

where A is an $n \times m$ matrix with $n \geq m$, b is a vector in $\mathbb{R}^n$, $f : \mathbb{R}^m \to (-\infty, +\infty]$ is a closed proper convex function, $\mathcal{K}^n$ is a second-order cone (SOC for short) in $\mathbb{R}^n$ given by

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

and $x \succeq_{\mathcal{K}^n} 0$ means that $x \in \mathcal{K}^n$. Note that a function is closed if and only if it is lower semi-continuous (l.s.c. for short) and a function is proper if $f(\zeta) < +\infty$ for at least one $\zeta \in \mathbb{R}^m$ and $f(\zeta) > -\infty$ for all $\zeta \in \mathbb{R}^m$. The CSOCP, as an extension of the standard second-order cone programming (SOCP) (see Sect. 4), has applications in a broad range of fields from engineering, control and finance to robust optimization and combinatorial optimization; see [3–7, and references therein].

Recently, the SOCP has received much attention in optimization, particularly in the context of solutions methods. In this paper, we focus on the solution of the more general CSOCP. Note that the CSOCP is a special class of convex programs, and therefore it can be solved via general convex programming methods. One of these methods is the proximal point algorithm for minimizing a convex function $f(\zeta)$ defined on $\mathbb{R}^m$, which replaces the problem $\min_{\zeta \in \mathbb{R}^m} f(\zeta)$ by a sequence of minimization problems with strictly convex objectives and generates a sequence $\{\zeta^k\}$ by

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

where μ_k is a sequence of positive numbers and $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^m$. The method was due to Martinet [8] who introduced the above proximal minimization problem based on the Moreau proximal approximation [9] of f. The proximal point algorithm was then further developed and studied by Rockafellar [10, 11]. Later, several researchers [1, 2, 12–14] proposed and investigated nonquadratic proximal point algorithm for the convex programming with nonnegative constraints, by replacing the quadratic distance in (2) with other distance-like functions. Among others, Censor and Zenios [1] replaced the method (2) by a method of the form

$$\zeta^k = \operatorname*{argmin}_{\zeta \in \mathbb{R}^m} \{f(\zeta) + (1/\mu_k) D(\zeta, \zeta^k)\}, \tag{3}$$

where $D(\cdot, \cdot)$, called the D-function, is a measure of distance based on a Bregman function.

Recall that, given an open convex set S of $\mathbb{R}^m$, a convex real function g defined on the closure of S, is called a Bregman function [15–17] if it satisfies the properties listed in Definition 1.1 below; the induced D-function is given by

$$D_\varphi(\zeta, \xi) := \varphi(\zeta) - \varphi(\xi) - \langle \nabla\varphi(\xi), \zeta - \xi \rangle, \tag{4}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbb{R}^m$ and $\nabla\varphi$ denotes the gradient of φ.

Definition 1.1 Let $S \subseteq \mathbb{R}^m$ be an open set and let $\bar{S}$ be its closure. Then, $\varphi : \bar{S} \to \mathbb{R}$ is called a Bregman function with zone S if the following properties hold:

(i) φ is continuously differentiable on S.
(ii) φ is strictly convex and continuous on $\bar{S}$.
(iii) For each $\gamma \in \mathbb{R}$, the partial level sets $L_1(\xi, \gamma) := \{\zeta \in \bar{S} \mid D_\varphi(\zeta, \xi) \leq \gamma\}$ and $L_2(\zeta, \gamma) := \{\xi \in S \mid D_\varphi(\zeta, \xi) \leq \gamma\}$ are bounded for any $\xi \in S$ and $\zeta \in \bar{S}$.
(iv) If $\{\xi^k\} \subset S$ converges to ξ^*, then $D_\varphi(\xi^*, \xi^k)$ converges to 0.
(v) If $\{\zeta^k\}$ and $\{\xi^k\}$ are sequences such that $\xi^k \to \xi^* \in \bar{S}$, $\{\zeta^k\}$ is bounded and, if $D_\varphi(\zeta^k, \xi^k) \to 0$, then $\zeta^k \to \xi^*$.

The Bregman proximal minimization (BPM) method described as in (3) was further extended by Kiwiel [18] with generalized Bregman functions, called B-functions. Compared with Bregman functions, these functions are possibly nondifferentiable and infinite on the boundary of their domain. For the detailed definition of B-functions and the convergence of BPM method using B-functions, please refer to [18].

The main purpose of this paper is to extend the BPM method (3) so that it can be used to deal with the CSOCP. Specifically, we define a measure of distance in second-order cone $\mathcal{K}^n$ by a class of continuously differentiable strictly convex functions on $\mathbb{R}_{++}$ which are in fact special B-functions in $\mathbb{R}$ (see Property 3.1). The distance measure, including the entropy-like distance in $\mathcal{K}^n$ as a special case, is shown to have some favorable properties similar to those for a Bregman distance, and hence we here refer it as a quasi Bregman distance or quasi D-function. The specific definition is given in Sect. 3. Then, a proximal-like algorithm using quasi D-function is proposed and applied for solving the CSOCP. Like the proximal-point algorithm (3), we establish, under some mild assumptions, the global convergence of the algorithm expressed in terms of function values, and show that the sequence generated is bounded and each accumulation point is a solution of the CSOCP.

The rest of this paper is organized as follows. In Sect. 2, we review some basic concepts and properties associated with SOC. In Sect. 3, we define a quasi D-function in $\mathcal{K}^n$ and explore the relations among the quasi D-function, the D-function, and the double-regularized distance function [19]. In Sect. 4, we present a proximal-like algorithm using quasi D-function and apply it for solving the CSOCP, and meanwhile, analyze the convergence of the algorithm. Finally, we close this paper in Sect. 5.

Some words about our notation. $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote the nonnegative real number set and the positive real number set, respectively, and I represents an identity matrix of suitable dimension. For a differentiable function ϕ in $\mathbb{R}$, ϕ' represents its derivative. Given a set S, we use $\bar{S}$, $\mathrm{int}(S)$ and $\mathrm{bd}(S)$ to denote the closure, the interior and the boundary of S, respectively. For a closed proper convex function $f : \mathbb{R}^m \to (-\infty, +\infty]$, we denote the domain of f by $\mathrm{dom}(f) := \{\zeta \in \mathbb{R}^m \mid f(\zeta) < \infty\}$ and the subdifferential of f at $\bar{\zeta}$ by $\partial f(\bar{\zeta}) := \{w \in \mathbb{R}^m \mid f(\zeta) \geq f(\bar{\zeta}) + \langle w, \zeta - \bar{\zeta}\rangle, \forall \zeta \in \mathbb{R}^m\}$. If f is differentiable at ζ, we use $\nabla f(\zeta)$ to denote its gradient at ζ. 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 $x \succeq_{\mathcal{K}^n} 0$ if and only if $x \in \mathcal{K}^n$; and $x \succ_{\mathcal{K}^n} 0$ if and only if $x \in \mathrm{int}(\mathcal{K}^n)$.

2 Preliminaries

In this section, we review some basic concepts and properties related to the SOC $\mathcal{K}^n$ that will be used in the subsequent analysis. For any $x = (x_1, x_2)$, $y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, we define their Jordan product as

$$x \circ y := (\langle x, y \rangle, y_1 x_2 + x_1 y_2). \tag{5}$$

We write $x + y$ to mean the usual componentwise addition of vectors and x^2 to mean $x \circ x$. Then $\circ$, $+$ and $e = (1, 0, \ldots, 0)^T \in \mathbb{R}^n$ have the following basic properties [20, 21]: (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 $x, y, z \in \mathbb{R}^n$. Note that the Jordan product is not associative, but 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 define $x^0 = e$. Besides, we should point out that $\mathcal{K}^n$ is not closed under Jordan product.

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

$$\det(x) = x_1^2 - \|x_2\|^2, \qquad \operatorname{tr}(x) = 2x_1. \tag{6}$$

In general, $\det(x \circ y) \neq \det(x)\det(y)$ unless x and y are collinear, i.e., $x = \alpha y$ 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 $y \in \mathbb{R}^n$ 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

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

Therefore, $x \in \operatorname{int}(\mathcal{K}^n)$ if and only if $x^{-1} \in \operatorname{int}(\mathcal{K}^n)$. For any $x \in \mathcal{K}^n$, it is known that there exists a unique vector in $\mathcal{K}^n$ denoted by $x^{1/2}$ such that $(x^{1/2})^2 = x^{1/2} \circ x^{1/2} = x$.

Next, we introduce the definition of spectral factorization. Let $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$; then, x can be decomposed as

$$x = \lambda_1(x) u_x^{(1)} + \lambda_2(x) u_x^{(2)}, \tag{7}$$

where $\lambda_i(x)$ and $u_x^{(i)}$ are the spectral value and the associated spectral vector given by

$$\begin{aligned} &\lambda_i(x) := x_1 + (-1)^i \|x_2\|, \\ &u_x^{(i)} := \begin{cases} (1/2)(1, (-1)^i x_2/\|x_2\|), & \text{if } x_2 \neq 0; \\ (1/2)(1, (-1)^i \bar{w}_2), & \text{if } x_2 = 0, \end{cases} \end{aligned} \tag{8}$$

for $i = 1, 2$ with $\bar{w}_2$ being any vector in $\mathbb{R}^{n-1}$ satisfying $\|\bar{w}_2\| = 1$. If $x_2 \neq 0$, the factorization is unique. In the sequel, for any $x \in \mathbb{R}^n$, we write $\lambda(x) := (\lambda_1(x), \lambda_2(x))$, where $\lambda_1(x), \lambda_2(x)$ are the spectral values of x.

The spectral decomposition along with the Jordan algebra associated with SOC has some basic properties as below, whose proofs can be found in [20, 21].

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 above, we have:*

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

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

(b) *$u_x^{(1)}$ and $u_x^{(2)}$ are idempotent under the 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 norm of x can be represented by $\lambda_1(x),\lambda_2(x)$:*

$$\begin{aligned}&\det(x)=\lambda_1(x)\lambda_2(x),\\&\mathrm{tr}(x)=\lambda_1(x)+\lambda_2(x),\\&\|x\|^2=(\lambda_1^2(x)+\lambda_2^2(x))/2.\end{aligned}$$

(d) *$\lambda_1(x),\lambda_2(x)$ are nonnegative (positive) if and only if $x\in\mathcal{K}^n$ ($x\in\mathrm{int}(\mathcal{K}^n)$).*

Finally, for any $g:\mathbb{R}\to\mathbb{R}$, one can define a corresponding function $g^{\mathrm{soc}}(x)$ in $\mathbb{R}^n$ by applying g to the spectral values of the spectral decomposition of x with respect to $\mathcal{K}^n$, i.e.,

$$g^{\mathrm{soc}}(x)=g(\lambda_1(x))u_x^{(1)}+g(\lambda_2(x))u_x^{(2)},\quad \forall x=(x_1,x_2)\in\mathbb{R}\times\mathbb{R}^{n-1}. \tag{9}$$

If g is defined only on a subset of $\mathbb{R}$, then g^{soc} is defined on the corresponding subset of $\mathbb{R}^n$. The definition in (9) is unambiguous whether $x_2\neq 0$ or $x_2=0$. The following lemma states some relations between the vector-valued function g^{soc} and the scalar function g, whose proof can be found in [6, 21].

Lemma 2.1 *Given a function $g:\mathbb{R}\to\mathbb{R}$, let $g^{\mathrm{soc}}(x)$ be the vector-valued function defined by (9). If g is differentiable (respectively, continuously differentiable), then $g^{\mathrm{soc}}(x)$ is also differentiable (respectively, continuously differentiable), and its Jacobian at $x=(x_1,x_2)\in\mathbb{R}\times\mathbb{R}^{n-1}$ is given by $\nabla g^{\mathrm{soc}}(x)=g'(x_1)I$, if $x_2=0$, and otherwise*

$$\nabla g^{\mathrm{soc}}(x)=\begin{bmatrix} b & cx_2^T/\|x_2\| \\ cx_2/\|x_2\| & aI+(b-a)(x_2x_2^T)/\|x_2\|^2\end{bmatrix}, \tag{10}$$

where

$$\begin{aligned}&a=[g(\lambda_2(x))-g(\lambda_1(x))]/[\lambda_2(x)-\lambda_1(x)],\\&b=[g'(\lambda_2(x))+g'(\lambda_1(x))]/2,\qquad c=[g'(\lambda_2(x))-g'(\lambda_1(x))]/2.\end{aligned} \tag{11}$$

3 Quasi D-Functions in SOC and Their Properties

In this section, we present a class of distance measures on SOC and discuss its relations with the D-function and the double-regularized Bregman distance [19]. For this purpose, we need a class of functions $\phi:\mathbb{R}_+\to\mathbb{R}$ satisfying Property 3.1 below, in which the function $d:\mathbb{R}_+\times\mathbb{R}_{++}\to\mathbb{R}$ is defined by

$$d(s,t)=\phi(s)-\phi(t)-\phi'(t)(s-t),\quad \forall s\in\mathbb{R}_+,\ t\in\mathbb{R}_{++}. \tag{12}$$

Property 3.1

(a) *ϕ is continuously differentiable on $\mathbb{R}_{++}$.*
(b) *ϕ is strictly convex and continuous on $\mathbb{R}_+$.*
(c) *For each $\gamma\in\mathbb{R}$, the level sets $\{s\in\mathbb{R}_+\mid d(s,t)\le\gamma\}$ and $\{t\in\mathbb{R}_{++}\mid d(s,t)\le\gamma\}$ are bounded for any $t\in\mathbb{R}_{++}$ and $s\in\mathbb{R}_+$, respectively.*
(d) *If $\{t^k\}\subset\mathbb{R}_{++}$ is a sequence such that $\lim_{k\to+\infty}t^k=0$, then $\lim_{k\to+\infty}\phi'(t^k)(s-t^k)=-\infty$ for all $s\in\mathbb{R}_{++}$.*

The function ϕ satisfying Property 3.1 (d) is said to be boundary coercive in [22]. If setting $\phi(t)=+\infty$ when $t\notin\mathbb{R}_+$, then ϕ becomes a closed proper strictly convex on $\mathbb{R}$. Furthermore, by Lemma 2.4 of [18] and Property 3.1 (c), it is not difficult to see that $\phi(t)$ and $\sum_{i=1}^n\phi(x_i)$ are a B-function on $\mathbb{R}$ and $\mathbb{R}^n$, respectively. Unless otherwise stated, in the rest of this paper, we always assume that ϕ satisfies Property 3.1.

From the discussions in Sect. 2, clearly, the following vector-valued functions

$$\phi^{\mathrm{soc}}(x)=\phi(\lambda_1(x))u_x^{(1)}+\phi(\lambda_2(x))u_x^{(2)} \tag{13}$$

and

$$(\phi')^{\mathrm{soc}}(x)=\phi'(\lambda_1(x))\,u_x^{(1)}+\phi'(\lambda_2(x))\,u_x^{(2)} \tag{14}$$

are well-defined over $\mathcal{K}^n$ and $\mathrm{int}(\mathcal{K}^n)$, respectively. In view of this, we define

$$H(x,y):=\begin{cases}\mathrm{tr}[\phi^{\mathrm{soc}}(x)-\phi^{\mathrm{soc}}(y)-(\phi')^{\mathrm{soc}}(y)\circ(x-y)], & \forall x\in\mathcal{K}^n,\ y\in\mathrm{int}(\mathcal{K}^n),\\ +\infty, & \text{otherwise}.\end{cases} \tag{15}$$

In what follows, we will show that the function $H:\mathbb{R}^n\times\mathbb{R}^n\to(-\infty,+\infty]$ enjoys some favorable properties similar to those of the D-function. Particularly, we prove that $H(x,y)\ge 0$ for any $x\in\mathcal{K}^n$, $y\in\mathrm{int}(\mathcal{K}^n)$, and moreover, $H(x,y)=0$ if and only if $x=y$. Consequently, it can be regarded as a distance measure on the SOC.

We first start with two technical lemmas that will be used in the subsequent analysis.

Lemma 3.1 *For any $x=(x_1,x_2)$, $y=(y_1,y_2)\in\mathbb{R}\times\mathbb{R}^{n-1}$, we have $\mathrm{tr}(x\circ y)\le\langle\lambda(x),\lambda(y)\rangle$ where $\lambda(x)=(\lambda_1(x),\lambda_2(x))$ and $\lambda(y)=(\lambda_1(y),\lambda_2(y))$, and the inequality holds with equality if and only if $x_2=\alpha y_2$ for some $\alpha>0$.*

Proof From (5)–(6) and Cauchy-Schwartz inequality,

$$\mathrm{tr}(x\circ y)=2\langle x,y\rangle=2x_1y_1+2x_2^Ty_2\le 2x_1y_1+2\|x_2\|\cdot\|y_2\|.$$

On the other hand, from the definition of the spectral values given by (8),

$$\begin{aligned}\langle \lambda(x), \lambda(y)\rangle &= (x_1 - \|x_2\|)(y_1 - \|y_2\|) + (x_1 + \|x_2\|)(y_1 + \|y_2\|)\\ &= 2x_1y_1 + 2\|x_2\| \cdot \|y_2\|.\end{aligned}$$

From the above two sides, we obtain immediately the inequality relation. In addition, we note that the inequality becomes an equality if and only if $x_2^T y_2 = \|x_2\| \cdot \|y_2\|$, which is equivalent to saying that $x_2 = \alpha y_2$ for some $\alpha > 0$. □

Lemma 3.2 *Let* $\phi^{\mathrm{soc}}(x)$ *and* $(\phi')^{\mathrm{soc}}(x)$ *be given as in* (13) *and* (14), *respectively. Then*:

(a) $\phi^{\mathrm{soc}}(x)$ *is continuously differentiable on* $\mathrm{int}(\mathcal{K}^n)$ *with the gradient* $\nabla\phi^{\mathrm{soc}}(x)$ *satisfying* $\nabla\phi^{\mathrm{soc}}(x)e = (\phi')^{\mathrm{soc}}(x)$.
(b) $\mathrm{tr}[\phi^{\mathrm{soc}}(x)] = \sum_{i=1}^{2} \phi[\lambda_i(x)]$ *and* $\mathrm{tr}[(\phi')^{\mathrm{soc}}(x)] = \sum_{i=1}^{2} \phi'[\lambda_i(x)]$.
(c) $\mathrm{tr}[\phi^{\mathrm{soc}}(x)]$ *is continuously differentiable on* $\mathrm{int}(\mathcal{K}^n)$ *with* $\nabla\,\mathrm{tr}[\phi^{\mathrm{soc}}(x)] = 2\nabla\phi^{\mathrm{soc}}(x)e$.
(d) $\mathrm{tr}[\phi^{\mathrm{soc}}(x)]$ *is strictly convex and continuous on* $\mathcal{K}^n$.
(e) *If* $\{y^k\} \subset \mathrm{int}(\mathcal{K}^n)$ *is a sequence such that* $\lim_{k\to+\infty} y^k = \bar{y} \in \mathrm{bd}(\mathcal{K}^n)$, *then*

$$\lim_{k\to+\infty} \langle \nabla\,\mathrm{tr}[\phi^{\mathrm{soc}}(y^k)],\ x - y^k\rangle = -\infty \quad \textit{for all } x \in \mathrm{int}(\mathcal{K}^n).$$

In other words, *the function* $\mathrm{tr}[\phi^{\mathrm{soc}}(x)]$ *is boundary coercive.*

Proof (a) The first part is due to Lemma 2.1, and we next prove the second part. If $x_2 \neq 0$, then by formulas (10)–(11) it is easy to compute that

$$\nabla\phi^{\mathrm{soc}}(x)e = \begin{pmatrix} (1/2)[\phi'(\lambda_2(x)) + \phi'(\lambda_1(x))] \\ (1/2)[\phi'(\lambda_2(x)) - \phi'(\lambda_1(x))](x_2/\|x_2\|) \end{pmatrix}.$$

In addition, using (8) and (14), we can prove that the vector in the right hand side is exactly $(\phi')^{\mathrm{soc}}(x)$. Therefore, $\nabla\phi^{\mathrm{soc}}(x)e = (\phi')^{\mathrm{soc}}(x)$. If $x_2 = 0$, from $\nabla\phi^{\mathrm{soc}}(x) = \phi'(x_1)I$ and formula (8), we readily obtain $\nabla\phi^{\mathrm{soc}}(x)e = (\phi')^{\mathrm{soc}}(x)$.

(b) The result follows directly from Property 2.1 (c) and (13)–(14).

(c) From part (a) and the fact that $\mathrm{tr}[\phi^{\mathrm{soc}}(x)] = \mathrm{tr}[\phi^{\mathrm{soc}}(x) \circ e] = 2\langle\phi^{\mathrm{soc}}(x), e\rangle$, clearly, $\mathrm{tr}[\phi^{\mathrm{soc}}(x)]$ is continuously differentiable on $\mathrm{int}(\mathcal{K}^n)$. Applying the chain rule for inner product of two functions yields immediately that $\nabla\,\mathrm{tr}[\phi^{\mathrm{soc}}(x)] = 2\nabla\phi^{\mathrm{soc}}(x)e$.

(d) It is clear that $\phi^{\mathrm{soc}}(x)$ is continuous on $\mathcal{K}^n$. We next prove that it is strictly convex on $\mathcal{K}^n$. For any $x, y \in \mathcal{K}^n$ with $x \neq y$ and $\alpha, \beta \in (0,1)$ with $\alpha + \beta = 1$, we have that

$$\begin{aligned}\lambda_1(\alpha x + \beta y) &= \alpha x_1 + \beta y_1 - \|\alpha x_2 + \beta y_2\| \geq \alpha\lambda_1(x) + \beta\lambda_1(y),\\ \lambda_2(\alpha x + \beta y) &= \alpha x_1 + \beta y_1 + \|\alpha x_2 + \beta y_2\| \leq \alpha\lambda_2(x) + \beta\lambda_2(y),\end{aligned}$$

implying that

$$\alpha\lambda_1(x) + \beta\lambda_1(y) \leq \lambda_1(\alpha x + \beta y) \leq \lambda_2(\alpha x + \beta y) \leq \alpha\lambda_2(x) + \beta\lambda_2(y).$$

On the other hand,

$$\begin{aligned}\lambda_1(\alpha x+\beta y)+\lambda_2(\alpha x+\beta y) &= 2\alpha x_1+2\beta y_1\\ &= [\alpha\lambda_1(x)+\beta\lambda_1(y)]+[\alpha\lambda_2(x)+\beta\lambda_2(y)].\end{aligned}$$

The last two equations imply that there exists $\rho \in [0,1]$ such that

$$\begin{aligned}\lambda_1(\alpha x+\beta y) &= \rho[\alpha\lambda_1(x)+\beta\lambda_1(y)]+(1-\rho)[\alpha\lambda_2(x)+\beta\lambda_2(y)],\\ \lambda_2(\alpha x+\beta y) &= (1-\rho)[\alpha\lambda_1(x)+\beta\lambda_1(y)]+\rho[\alpha\lambda_2(x)+\beta\lambda_2(y)].\end{aligned}$$

Thus, from Property 2.1, it follows that

$$\begin{aligned}\operatorname{tr}[\phi^{\mathrm{soc}}(\alpha x+\beta y)] &= \phi[\lambda_1(\alpha x+\beta y)]+\phi[\lambda_2(\alpha x+\beta y)]\\ &= \phi[\rho(\alpha\lambda_1(x)+\beta\lambda_1(y))+(1-\rho)(\alpha\lambda_2(x)+\beta\lambda_2(y))]\\ &\quad +\phi[(1-\rho)(\alpha\lambda_1(x)+\beta\lambda_1(y))+\rho(\alpha\lambda_2(x)+\beta\lambda_2(y))]\\ &\le \rho\phi(\alpha\lambda_1(x)+\beta\lambda_1(y))+(1-\rho)\phi(\alpha\lambda_2(x)+\beta\lambda_2(y))\\ &\quad +(1-\rho)\phi(\alpha\lambda_1(x)+\beta\lambda_1(y))+\rho\phi(\alpha\lambda_2(x)+\beta\lambda_2(y))\\ &= \phi(\alpha\lambda_1(x)+\beta\lambda_1(y))+\phi(\alpha\lambda_2(x)+\beta\lambda_2(y))\\ &< \alpha\phi(\lambda_1(x))+\beta\phi(\lambda_1(y))+\alpha\phi(\lambda_2(x))+\beta\phi(\lambda_2(y))\\ &= \alpha\operatorname{tr}[\phi^{\mathrm{soc}}(x)]+\beta\operatorname{tr}[\phi^{\mathrm{soc}}(y)],\end{aligned}$$

where the first equality and the last one follow from part (b), and the two inequalities are due to the strict convexity of ϕ on $\mathbb{R}_{++}$. By the definition of strict convexity, the conclusion holds.

(e) From part (a) and part (c), we can obtain readily the following equality:

$$\nabla\operatorname{tr}[\phi^{\mathrm{soc}}(x)] = 2(\phi')^{\mathrm{soc}}(x), \quad \forall x\in\operatorname{int}(\mathcal{K}^n). \tag{16}$$

Using the relation and Lemma 3.1, we then have that

$$\begin{aligned}\langle\nabla\operatorname{tr}[\phi^{\mathrm{soc}}(y^k)],\ x-y^k\rangle &= 2\langle(\phi')^{\mathrm{soc}}(y^k),\ x-y^k\rangle\\ &= \operatorname{tr}[(\phi')^{\mathrm{soc}}(y^k)\circ(x-y^k)]\\ &= \operatorname{tr}[(\phi')^{\mathrm{soc}}(y^k)\circ x]-\operatorname{tr}[(\phi')^{\mathrm{soc}}(y^k)\circ y^k]\\ &\le \sum_{i=1}^{2}\phi'[\lambda_i(y^k)]\lambda_i(x)-\operatorname{tr}[(\phi')^{\mathrm{soc}}(y^k)\circ y^k].\end{aligned} \tag{17}$$

In addition, by Property 2.1 (a)–(b), for any $y\in\operatorname{int}(\mathcal{K}^n)$, we can compute

$$(\phi')^{\mathrm{soc}}(y)\circ y = \phi'(\lambda_1(y))\lambda_1(y)u_y^{(1)}+\phi'(\lambda_2(y))\lambda_2(y)u_y^{(2)}, \tag{18}$$

which implies that

$$\mathrm{tr}[(\phi')^{\mathrm{soc}}(y^k)\circ y^k]=\sum_{i=1}^{2}\phi'[\lambda_i(y^k)]\lambda_i(y^k). \tag{19}$$

Combining (17) and (19) immediately yields that

$$\langle\nabla\,\mathrm{tr}[\phi^{\mathrm{soc}}(y^k)],\ x-y^k\rangle\le\sum_{i=1}^{2}\phi'[\lambda_i(y^k)][\lambda_i(x)-\lambda_i(y^k)]. \tag{20}$$

Note that $\lambda_2(\bar{y})\ge\lambda_1(\bar{y})=0$ and $\lambda_2(x)\ge\lambda_1(x)>0$ since $\bar{y}\in\mathrm{bd}(\mathcal{K}^n)$ and $x\in\mathrm{int}(\mathcal{K}^n)$. Hence, if $\lambda_2(\bar{y})=0$, then by Property 3.1 (d) and the continuity of $\lambda_i(\cdot)$ for $i=1,2$, we have

$$\lim_{k\to+\infty}\phi'[\lambda_i(y^k)][\lambda_i(x)-\lambda_i(y^k)]=-\infty,\quad i=1,2,$$

which means that

$$\lim_{k\to+\infty}\sum_{i=1}^{2}\phi'[\lambda_i(y^k)][\lambda_i(x)-\lambda_i(y^k)]=-\infty. \tag{21}$$

If $\lambda_2(\bar{y})>0$, then $\lim_{k\to+\infty}\phi'[\lambda_2(y^k)][\lambda_2(x)-\lambda_2(y^k)]$ is finite and

$$\lim_{k\to+\infty}\phi'[\lambda_1(y^k)][\lambda_1(x)-\lambda_1(y^k)]=-\infty;$$

therefore, the result in (21) also holds under such case. Combining (21) with (20), we prove that the conclusion holds. □

Using the relation (16), we have that, for any $x\in\mathcal{K}^n$ and $y\in\mathrm{int}(\mathcal{K}^n)$,

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

As a consequence, the function $H(x,y)$ in (15) can be rewritten as

$$H(x,y)=\begin{cases}\mathrm{tr}[\phi^{\mathrm{soc}}(x)]-\mathrm{tr}[\phi^{\mathrm{soc}}(y)] & \\ \quad-\langle\nabla\,\mathrm{tr}[\phi^{\mathrm{soc}}(y)],x-y\rangle & \forall x\in\mathcal{K}^n,\ y\in\mathrm{int}(\mathcal{K}^n),\\ +\infty & \text{otherwise}.\end{cases} \tag{22}$$

By the representation, we next investigate several important properties of $H(x,y)$.

Proposition 3.1 *Let $H(x,y)$ be the function defined as in (15) or (22). Then,*

(a) *$H(x,y)$ is continuous on $\mathcal{K}^n\times\mathrm{int}(\mathcal{K}^n)$ and, for any $y\in\mathrm{int}(\mathcal{K}^n)$, the function $H(\cdot,y)$ is strictly convex on $\mathcal{K}^n$.*

(b) *For any given $y\in\mathrm{int}(\mathcal{K}^n)$, $H(x,y)$ is continuously differentiable on $\mathrm{int}(\mathcal{K}^n)$ with*

$$\nabla_x H(x,y)=\nabla\,\mathrm{tr}[\phi^{\mathrm{soc}}(x)]-\nabla\,\mathrm{tr}[\phi^{\mathrm{soc}}(y)]=2[(\phi')^{\mathrm{soc}}(x)-(\phi')^{\mathrm{soc}}(y)]. \tag{23}$$

(c) $H(x,y)\ge\sum_{i=1}^{2}d(\lambda_i(x),\lambda_i(y))\ge 0$ *for any* $x\in\mathcal{K}^n$ *and* $y\in\operatorname{int}(\mathcal{K}^n)$, *where* $d(\cdot,\cdot)$ *is defined by* (12). *Moreover,* $H(x,y)=0$ *if and only if* $x=y$.

(d) *For each* $\gamma\in\mathbb{R}$, *the level sets* $L_H(y,\gamma):=\{x\in\mathcal{K}^n \mid H(x,y)\le\gamma\}$ *and* $L_H(x,\gamma):=\{y\in\operatorname{int}(\mathcal{K}^n)\mid H(x,y)\le\gamma\}$ *are bounded for any* $y\in\operatorname{int}(\mathcal{K}^n)$ *and* $x\in\mathcal{K}^n$, *respectively.*

(e) *If* $\{y^k\}\subset\operatorname{int}(\mathcal{K}^n)$ *is a sequence converging to* $y^*\in\operatorname{int}(\mathcal{K}^n)$, *then* $H(y^*,y^k)\to 0$.

(f) *If* $\{x^k\}\subset\operatorname{int}(\mathcal{K}^n)$ *and* $\{y^k\}\subset\operatorname{int}(\mathcal{K}^n)$ *are sequences such that* $\{y^k\}\to y^*\in\operatorname{int}(\mathcal{K}^n)$, $\{x^k\}$ *is bounded, and* $H(x^k,y^k)\to 0$, *then* $x^k\to y^*$.

Proof (a) Note that $\phi^{\mathrm{soc}}(x)$, $(\phi')^{\mathrm{soc}}(y)$, $(\phi')^{\mathrm{soc}}(y)\circ(x-y)$ are continuous for any $x\in\mathcal{K}^n$ and $y\in\operatorname{int}(\mathcal{K}^n)$ and the trace function $\operatorname{tr}(\cdot)$ is also continuous, and hence $H(x,y)$ is continuous on $\mathcal{K}^n\times\operatorname{int}(\mathcal{K}^n)$. From Lemma 3.2 (d), $\operatorname{tr}[\phi^{\mathrm{soc}}(x)]$ is strictly convex over $\mathcal{K}^n$, whereas $-\operatorname{tr}[\phi^{\mathrm{soc}}(y)]-\langle\nabla\operatorname{tr}[\phi^{\mathrm{soc}}(y)],x-y\rangle$ is clearly convex in $\mathcal{K}^n$ for fixed $y\in\operatorname{int}(\mathcal{K}^n)$. This means that $H(\cdot,y)$ is strictly convex for any $y\in\operatorname{int}(\mathcal{K}^n)$.

(b) By Lemma 3.2 (c), the function $H(\cdot,y)$ for any given $y\in\operatorname{int}(\mathcal{K}^n)$ is continuously differentiable on $\operatorname{int}(\mathcal{K}^n)$. The first equality in (23) is obvious and the second is due to (16).

(c) The result follows directly from the following equalities and inequalities:

$$\begin{aligned}
H(x,y)&=\operatorname{tr}[\phi^{\mathrm{soc}}(x)]-\operatorname{tr}[\phi^{\mathrm{soc}}(y)]-\operatorname{tr}[(\phi')^{\mathrm{soc}}(y)\circ(x-y)]\\
&=\operatorname{tr}[\phi^{\mathrm{soc}}(x)]-\operatorname{tr}[\phi^{\mathrm{soc}}(y)]-\operatorname{tr}[(\phi')^{\mathrm{soc}}(y)\circ x]+\operatorname{tr}[(\phi')^{\mathrm{soc}}(y)\circ y]\\
&\ge\operatorname{tr}[\phi^{\mathrm{soc}}(x)]-\operatorname{tr}[\phi^{\mathrm{soc}}(y)]-\sum_{i=1}^{2}\phi'(\lambda_i(y))\lambda_i(x)+\operatorname{tr}[(\phi')^{\mathrm{soc}}(y)\circ y]\\
&=\sum_{i=1}^{2}[\phi(\lambda_i(x))+\phi(\lambda_i(y))-\phi'(\lambda_i(y))\lambda_i(x)+\phi'(\lambda_i(y))\lambda_i(y)]\\
&=\sum_{i=1}^{2}[\phi(\lambda_i(x))-\phi(\lambda_i(y))-\phi'(\lambda_i(y))(\lambda_i(x)-\lambda_i(y))]\\
&=\sum_{i=1}^{2}d(\lambda_i(x),\lambda_i(y))\ge 0,
\end{aligned}$$

where the first equality is due to (15), the second and fourth are obvious, the third follows from Lemma 3.2 (b) and (18), the last one is from (12), and the first inequality follows from Lemma 3.1 and the last one is due to the strict convexity of ϕ on $\mathbb{R}_+$. Note that $\operatorname{tr}[\phi^{\mathrm{soc}}(x)]$ is strictly convex for any $x\in\mathcal{K}^n$ by Lemma 3.2 (d), and so $H(x,y)=0$ if and only if $x=y$ by (22).

(d) From part (c), we have $L_H(y,\gamma)\subseteq\{x\in\mathcal{K}^n\mid \sum_{i=1}^{2}d(\lambda_i(x),\lambda_i(y))\le\gamma\}$. By Property 3.1 (c), the set in the right-hand side is bounded. So, $L_H(y,\gamma)$ is bounded for $y\in\operatorname{int}(\mathcal{K}^n)$. Similarly, $L_H(x,\gamma)$ is bounded for $x\in\mathcal{K}^n$.

From part (a)–(d), we obtain immediately the results in (e) and (f). □

Remark 3.1

(i) From (22), it is not difficult to see that $H(x,y)$ is exactly a distance measure induced by $\mathrm{tr}[\phi^{\mathrm{soc}}(x)]$ via formula (4). Therefore, if $n=1$ and ϕ is a Bregman function with zone $\mathbb{R}_{++}$, i.e., ϕ also satisfies the property: (e) if $\{s^k\}\subseteq\mathbb{R}_+$ and $\{t^k\}\subset\mathbb{R}_{++}$ are sequences such that $t^k\to t^*$, $\{s^k\}$ is bounded, and $d(s^k,t^k)\to 0$, then $s^k\to t^*$; then $H(x,y)$ reduces to the Bregman distance function $d(x,y)$ in (12).

(ii) When $n>1$, $H(x,y)$ is generally not a Bregman distance even if ϕ is a Bregman function with zone $\mathbb{R}_{++}$, since Proposition 3.1 (e) and (f) do not hold for $\{y^k\}\subseteq \mathrm{bd}(\mathcal{K}^n)$ and $y^*\in\mathrm{bd}(\mathcal{K}^n)$. By the proof of Proposition 3.1 (c), the main reason is that, to guarantee that

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

for any $x\in\mathcal{K}^n$ and $y\in\mathrm{int}(\mathcal{K}^n)$, the relation $[(\phi')^{\mathrm{soc}}(y)]_2=\alpha x_2$ with some $\alpha>0$ is required, where $[(\phi')^{\mathrm{soc}}(y)]_2$ is a vector composed of the last $n-1$ elements of $(\phi')^{\mathrm{soc}}(y)$. It is very stringent for ϕ to satisfy such relation. By this, $\mathrm{tr}[\phi^{\mathrm{soc}}(x)]$ is not a B-function [18] on $\mathbb{R}^n$ either, even if ϕ itself is a B-function.

(iii) We observe that $H(x,y)$ is inseparable, whereas the double-regularized distance function proposed by [19] belongs to the separable class of functions. In view of this, $H(x,y)$ cannot become a double-regularized distance function in $\mathcal{K}^n\times\mathrm{int}(\mathcal{K}^n)$, even when ϕ is such that $\tilde{d}(s,t)=d(s,t)/\phi''(t)+\frac{\mu}{2}(s-t)^2$ is a double regularized component (see [19]).

By Proposition 3.1 and Remark 3.1, we call $H(x,y)$ a quasi D-function in this paper. In the following, we present several specific examples of quasi D-functions.

Example 3.1 Let $\phi(t)=t\log t-t$ (with the convention $0\log 0=0$). It is easy to verify that ϕ satisfies Property 3.1. By Proposition 3.2 (b) of [21] and (13)–(14), we can compute $\phi^{\mathrm{soc}}(x)=x\circ\log x-x$ and $(\phi')^{\mathrm{soc}}(y)=\log y$ for any $x\in\mathcal{K}^n$ and $y\in\mathrm{int}(\mathcal{K}^n)$. Therefore,

$$H(x,y)=\mathrm{tr}(x\circ\log x-x\circ\log y+y-x),\quad \forall x\in\mathcal{K}^n,\ y\in\mathrm{int}(\mathcal{K}^n).$$

Example 3.2 Let $\phi(t)=t^2-\sqrt{t}$. It is not hard to verify that ϕ satisfies Property 3.1. Notice that, for any $x\in\mathcal{K}^n$, $x^2=x\circ x=\lambda_1^2(x)u_x^{(1)}+\lambda_2^2(x)u_x^{(2)}$ and $\sqrt{x}=\sqrt{\lambda_1(x)}u_x^{(1)}+\sqrt{\lambda_2(x)}u_x^{(2)}$, and a direct computation then yields $\phi^{\mathrm{soc}}(x)=x\circ x-\sqrt{x}$ and $(\phi')^{\mathrm{soc}}(y)=2y-(1/2)[\mathrm{tr}(\sqrt{y})e-\sqrt{y}]/\sqrt{\det(y)}$. This implies that, for any $x\in\mathcal{K}^n$, $y\in\mathrm{int}(\mathcal{K}^n)$,

$$H(x,y)=\mathrm{tr}\left[(x-y)^2-(\sqrt{x}-\sqrt{y})+\frac{(\mathrm{tr}(\sqrt{y})e-\sqrt{y})\circ(x-y)}{2\sqrt{\det(y)}}\right].$$

Example 3.3 Take $\phi(t)=t\log t-(1+t)\log(1+t)+(1+t)\log 2$ (with $0\log 0=0$). It is easily shown that ϕ satisfies Property 3.1. Using Property 2.1 (a)–(b), we can compute

$$\phi^{\mathrm{soc}}(x) = x \circ \log x - (e+x) \circ \log(e+x) + (e+x)\log 2, \quad x \in \mathcal{K}^n$$

and

$$(\phi')^{\mathrm{soc}}(y) = \log y - \log(e+y) + e \log 2, \quad y \in \mathrm{int}(\mathcal{K}^n).$$

Consequently, for any $x \in \mathcal{K}^n$ and $y \in \mathrm{int}(\mathcal{K}^n)$,

$$H(x,y) = \mathrm{tr}[x \circ (\log x - \log y) - (e+x) \circ (\log(e+x) - \log(e+y))].$$

In addition, from [14, 22], it follows that $\sum_{i=1}^m \phi(\zeta_i)$ generated by ϕ in the above examples is a Bregman function with zone $S = \mathbb{R}_+^m$, and consequently $\sum_{i=1}^m d(\zeta_i, \xi_i)$ defined as in (12) is a D-function induced by $\sum_{i=1}^m \phi(\zeta_i)$.

To close this section, we present another important property of $H(x,y)$.

Proposition 3.2 *Let $H(x,y)$ be defined as in (15) or (22). Then, for all $x, y \in \mathrm{int}(\mathcal{K}^n)$ and $z \in \mathcal{K}^n$, the following three-points identity holds*:

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

Proof Using the definition of H given as in (22), we have that

$$\begin{aligned} \langle \nabla \mathrm{tr}[\phi^{\mathrm{soc}}(x)],\ z - x\rangle &= \mathrm{tr}[\phi^{\mathrm{soc}}(z)] - \mathrm{tr}[\phi^{\mathrm{soc}}(x)] - H(z,x), \\ \langle \nabla \mathrm{tr}[\phi^{\mathrm{soc}}(y)],\ x - y\rangle &= \mathrm{tr}[\phi^{\mathrm{soc}}(x)] - \mathrm{tr}[\phi^{\mathrm{soc}}(y)] - H(x,y), \\ \langle \nabla \mathrm{tr}[\phi^{\mathrm{soc}}(y)],\ z - y\rangle &= \mathrm{tr}[\phi^{\mathrm{soc}}(z)] - \mathrm{tr}[\phi^{\mathrm{soc}}(y)] - H(z,y). \end{aligned}$$

Subtracting the first two equations from the last one gives the first equality. By (16),

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

This, together with the fact that $\mathrm{tr}(x \circ y) = \langle x, y\rangle$, leads to the second equality. □

4 Proximal-Like Algorithm for the CSOCP

In this section, we propose a proximal-like algorithm for solving the CSOCP based on the quasi D-function $H(x,y)$. For the sake of notation, we denote $\mathcal{F}$ by the feasible set

$$\mathcal{F} := \{\zeta \in \mathbb{R}^m \mid A\zeta + b \succeq_{\mathcal{K}^n} 0\}. \tag{24}$$

It is easy to verify that $\mathcal{F}$ is convex and its interior $\mathrm{int}(\mathcal{F})$ is given by

$$\mathrm{int}(\mathcal{F}) = \{\zeta \in \mathbb{R}^m \mid A\zeta + b \succ_{\mathcal{K}^n} 0\}. \tag{25}$$

Let $\psi : \mathbb{R}^m \to (-\infty, +\infty]$ be the function defined by

$$\psi(\zeta) := \begin{cases} \operatorname{tr}[\phi^{\mathrm{soc}}(A\zeta + b)], & \text{if } \zeta \in \mathcal{F}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{26}$$

By Lemma 3.2, it is easily shown that the following conclusions hold for $\psi(\zeta)$.

Lemma 4.1 *Let $\psi(\zeta)$ be given as in (26). If the matrix A has full rank m, then*:

(a) *$\psi(\zeta)$ is continuously differentiable on $\operatorname{int}(\mathcal{F})$ with $\nabla\psi(\zeta) = 2A^T(\phi')^{\mathrm{soc}} \times (A\zeta + b)$.*
(b) *$\psi(\zeta)$ is strictly convex and continuous on $\mathcal{F}$.*
(c) *$\psi(\zeta)$ is boundary coercive, i.e., if $\{\xi^k\} \subset \operatorname{int}(\mathcal{F})$ is such that $\lim_{k\to+\infty} \xi^k = \xi \in \operatorname{bd}(\mathcal{F})$, then for all $\zeta \in \operatorname{int}(\mathcal{F})$, there holds that $\lim_{k\to+\infty} \nabla\psi(\xi^k)^T \times (\zeta - \xi^k) = -\infty$.*

Let $\mathcal{D}(\zeta, \xi)$ be the function induced by the above $\psi(\zeta)$ via formula (4), i.e.,

$$\mathcal{D}(\zeta, \xi) := \psi(\zeta) - \psi(\xi) - \langle \nabla\psi(\xi), \zeta - \xi \rangle. \tag{27}$$

Then, from (26) and (22), it is not difficult to see that

$$\mathcal{D}(\zeta, \xi) = H(A\zeta + b, A\xi + b). \tag{28}$$

So, by Proposition 3.1 and Lemma 4.1, we can prove the following conclusions.

Lemma 4.2 *Let $\mathcal{D}(\zeta, \xi)$ be given by (27) or (28). If the matrix A has full rank m, then*:

(a) *$\mathcal{D}(\zeta, \xi)$ is continuous on $\mathcal{F} \times \operatorname{int}(\mathcal{F})$ and, for any given $\xi \in \operatorname{int}(\mathcal{F})$, the function $\mathcal{D}(\cdot, \xi)$ is strictly convex on $\mathcal{F}$.*
(b) *For any fixed $\xi \in \operatorname{int}(\mathcal{F})$, $\mathcal{D}(\cdot, \xi)$ is continuously differentiable on $\operatorname{int}(\mathcal{F})$ with*

$$\nabla_\zeta \mathcal{D}(\zeta, \xi) = \nabla\psi(\zeta) - \nabla\psi(\xi) = 2A^T[(\phi')^{\mathrm{soc}}(A\zeta + b) - (\phi')^{\mathrm{soc}}(A\xi + b)].$$

(c) *$\mathcal{D}(\zeta, \xi) \ge \sum_{i=1}^2 d(\lambda_i(A\zeta + b), \lambda_i(A\xi + b)) \ge 0$ for any $\zeta \in \mathcal{F}$ and $\xi \in \operatorname{int}(\mathcal{F})$, where $d(\cdot, \cdot)$ is defined by (12). Moreover, $\mathcal{D}(\zeta, \xi) = 0$ if and only if $\zeta = \xi$.*
(d) *For each $\gamma \in \mathbb{R}$, the partial level sets of $L_{\mathcal{D}}(\xi, \gamma) = \{\zeta \in \mathcal{F} \mid \mathcal{D}(\zeta, \xi) \le \gamma\}$ and $L_{\mathcal{D}}(\zeta, \gamma) = \{\xi \in \operatorname{int}(\mathcal{F}) : \mathcal{D}(\zeta, \xi) \le \gamma\}$ are bounded for any $\xi \in \operatorname{int}(\mathcal{F})$ and $\zeta \in \mathcal{F}$, respectively.*

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

$$\zeta^0 \in \operatorname{int}(\mathcal{F}), \tag{29}$$

$$\zeta^k = \operatorname*{argmin}_{\zeta \in \mathcal{F}} \{ f(\zeta) + (1/\mu_k)\mathcal{D}(\zeta, \zeta^{k-1}) \}, \quad k \ge 1, \tag{30}$$

where $\{\mu_k\}_{k\ge1}$ is a sequence of positive numbers.

To establish the convergence of the algorithm, we make the following assumptions:

(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 4.1 Assumption (A1) is elementary for the solution of the CSOCP. Assumption (A2) is common in the solution of SOCPs and it is obviously satisfied when $\mathcal{F} = \mathcal{K}^n$. Moreover, if we consider the standard SOCP

$$\begin{array}{ll} \min & c^T x, \\ \text{s.t.} & Ax = b, \quad x \in \mathcal{K}^n, \end{array} \tag{31}$$

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

$$\begin{array}{ll} \max & b^T y, \\ \text{s.t.} & c - A^T y \succeq_{\mathcal{K}^n} 0, \end{array} \tag{32}$$

satisfies Assumption (A2). This shows that we can solve the SOCP by applying the proximal-like algorithm in (29)–(30) to the dual problem (32).

In what follows, we are ready to prove the convergence of the proximal-like algorithm in (29)–(30) under Assumptions (A1) and (A2). We first show that the algorithm is well-defined.

Proposition 4.1 *Suppose that Assumptions* (A1)–(A2) *hold. Then, the algorithm described as in* (29)–(30) *generates a sequence* $\{\zeta^k\} \subset \mathrm{int}(\mathcal{F})$ *such that*

$$-2\mu_k^{-1} A^T [(\phi')^{\mathrm{soc}}(A\zeta^k + b) - (\phi')^{\mathrm{soc}}(A\zeta^{k-1} + b)] \in \partial f(\zeta^k). \tag{33}$$

Proof The proof proceeds by induction. For $k = 0$, clearly, $\zeta^0 \in \mathrm{int}(\mathcal{F})$. Assume that $\zeta^{k-1} \in \mathrm{int}(\mathcal{F})$. Let $f_k(\zeta) := f(\zeta) + \mu_k^{-1}\mathcal{D}(\zeta, \zeta^{k-1})$. Then Assumption (A1) and Lemma 4.2 (d) imply that f_k has bounded level sets in $\mathcal{F}$. By the lower semicontinuity of f and Lemma 4.2 (a), the minimization problem $\min_{\zeta\in\mathcal{F}} f_k(\zeta)$, i.e. the subproblem (30), has solutions. Moreover, the solution ζ^k is unique due to the convexity of f and the strict convexity of $\mathcal{D}(\cdot, \xi)$. In the following, we prove that $\zeta^k \in \mathrm{int}(\mathcal{F})$.

By Theorem 23.8 of [23] and the optimal condition for (30), ζ^k is the only $\zeta \in \mathbb{R}^n$ such that

$$2\mu_k^{-1} A^T (\phi')^{\mathrm{soc}}(A\zeta^{k-1} + b) \in \partial(f(\zeta) + \mu_k^{-1}\psi(\zeta) + \delta(\zeta \mid \mathcal{F})), \tag{34}$$

where $\delta(\zeta \mid \mathcal{F}) = 0$ if $\zeta \in \mathcal{F}$ and $+\infty$ otherwise. We will show that

$$\partial(f(\zeta) + \mu_k^{-1}\psi(\zeta) + \delta(\zeta \mid \mathcal{F})) = \emptyset, \quad \text{for all } \zeta \in \mathrm{bd}(\mathcal{F}), \tag{35}$$

which by (34) implies that $\zeta^k \in \mathrm{int}(\mathcal{F})$. Take $\zeta \in \mathrm{bd}(\mathcal{F})$ and assume that there exists $w \in \partial(f(\zeta) + \mu_k^{-1}\psi(\zeta) + \delta(\zeta|\mathcal{F}))$. Take $\widehat{\zeta} \in \mathrm{dom}(f) \cap \mathrm{int}(\mathcal{F})$ and let

$$\zeta^l = (1 - \epsilon_l)\zeta + \epsilon_l \widehat{\zeta} \tag{36}$$

with $\lim_{l\to+\infty}\epsilon_l=0$. From the convexity of $\mathrm{int}(\mathcal{F})$ and $\mathrm{dom}(f)$, it then follows that $\zeta^l\in\mathrm{dom}(f)\cap\mathrm{int}(\mathcal{F})$ and, moreover, $\lim_{l\to+\infty}\zeta^l=\zeta$. Consequently,

$$\begin{aligned}\epsilon_l w^T(\widehat{\zeta}-\zeta)&=w^T(\zeta^l-\zeta)\\&\le f(\zeta^l)-f(\zeta)+\mu_k^{-1}[\psi(\zeta^l)-\psi(\zeta)]\\&\le f(\zeta^l)-f(\zeta)+\mu_k^{-1}\langle 2A^T(\phi')^{\mathrm{soc}}(A\zeta^l+b),\zeta^l-\zeta\rangle\\&\le \epsilon_l(f(\widehat{\zeta})-f(\zeta))+\mu_k^{-1}\frac{\epsilon_l}{1-\epsilon_l}\,\mathrm{tr}[(\phi')^{\mathrm{soc}}(A\zeta^l+b)\circ(A\widehat{\zeta}-A\zeta^l)],\end{aligned}$$

where the first equality is due to (36), the first inequality follows from the definition of subdifferential and the convexity of $f(\zeta)+\mu_k^{-1}\psi(\zeta)+\delta(\zeta|\mathcal{F})$ in $\mathcal{F}$, the second one is due to the convexity and differentiability of $\psi(\zeta)$ in $\mathrm{int}(\mathcal{F})$, and the last one is from (36) and the convexity of f. Using Lemma 3.1 and (18), we then have that

$$\begin{aligned}&\mu_k(1-\epsilon_l)[f(\zeta)-f(\widehat{\zeta})+w^T(\widehat{\zeta}-\zeta)]\\&\quad\le \mathrm{tr}[(\phi')^{\mathrm{soc}}(A\zeta^l+b)\circ(A\widehat{\zeta}+b)]-\mathrm{tr}[(\phi')^{\mathrm{soc}}(A\zeta^l+b)\circ(A\zeta^l+b)]\\&\quad\le\sum_{i=1}^{2}[\phi'(\lambda_i(A\zeta^l+b))\lambda_i(A\widehat{\zeta}+b)-\phi'(\lambda_i(A\zeta^l+b))\lambda_i(A\zeta^l+b)]\\&\quad=\sum_{i=1}^{2}\phi'(\lambda_i(A\zeta^l+b))[\lambda_i(A\widehat{\zeta}+b)-\lambda_i(A\zeta^l+b)].\end{aligned}$$

Since $\zeta\in\mathrm{bd}(\mathcal{F})$, i.e., $A\zeta+b\in\mathrm{bd}(\mathcal{K}^n)$, it follows that $\lim_{l\to+\infty}\lambda_1(A\zeta^l+b)=0$. Thus, using Property 3.1 (d) and following the same line as the proof of Lemma 3.2 (d), we can prove that the right-hand side of the last inequality goes to $-\infty$ when l tends to $+\infty$, whereas the left-hand side has a finite limit. This gives a contradiction. Hence, (35) follows, which means that $\zeta^k\in\mathrm{int}(\mathcal{F})$.

Finally, let us prove $\partial\delta(\zeta^k|\,\mathcal{F})=\{0\}$. From p. 226 of [23], it follows that

$$\partial\delta(z|\mathcal{K}^n)=\{\upsilon\in\mathbb{R}^n\mid \upsilon\preceq_{\mathcal{K}^n}0,\ \mathrm{tr}(\upsilon\circ z)=0\}.$$

Using Theorem 23.9 of [23] and the assumption $\mathrm{dom}(f)\cap\mathrm{int}(\mathcal{F})\neq\emptyset$, we have

$$\partial\delta(\zeta|\,\mathcal{F})=\{A^T\upsilon\in\mathbb{R}^n\mid \upsilon\preceq_{\mathcal{K}^n}0,\ \mathrm{tr}(\upsilon\circ(A\zeta+b))=0\}.$$

In addition, from the self-dual property of the symmetric cone $\mathcal{K}^n$, we know that $\mathrm{tr}(x\circ y)=0$ for any $x\succeq_{\mathcal{K}^n}0$ and $y\succ_{\mathcal{K}^n}0$ implies $x=0$. Thus, we obtain $\partial\delta(\zeta^k|\,\mathcal{F})=\{0\}$. This together with (34) and Theorem 23.8 of [23] yields the desired result. □

Proposition 4.1 implies that the second-order cone constrained subproblem in (30) is actually equivalent to an unconstrained one,

$$\zeta^k=\operatorname*{argmin}_{\zeta\in\mathbb{R}^m}\{f(\zeta)+\mu_k^{-1}\mathcal{D}(\zeta,\zeta^{k-1})\},$$

which is obviously simpler than the original CSOCP. This means that the proximal-like algorithm proposed transforms the CSOCP into the solution of a sequence of simpler problems. We next present some properties satisfied by $\{\zeta^k\}$. For convenience, we denote the optimal set of the CSOCP by $\mathcal{F}^* := \{\zeta \in \mathcal{F} \mid f(\zeta) = f_*\}$.

Proposition 4.2 *Let $\{\zeta^k\}$ be the sequence generated by the algorithm described as in (29)–(30), and let $\sigma_N = \sum_{k=1}^N \mu_k$. Then, the following results hold.*

(a) *$\{f(\zeta^k)\}$ is a nonincreasing sequence.*
(b) *$\mu_k(f(\zeta^k) - f(\zeta)) \le \mathcal{D}(\zeta, \zeta^{k-1}) - \mathcal{D}(\zeta, \zeta^k)$ for all $\zeta \in \mathcal{F}$.*
(c) *$\sigma_N(f(\zeta^N) - f(\zeta)) \le \mathcal{D}(\zeta, \zeta^0) - \mathcal{D}(\zeta, \zeta^N)$ for all $\zeta \in \mathcal{F}$.*
(d) *$\mathcal{D}(\zeta, \zeta^k)$ is nonincreasing for any $\zeta \in \mathcal{F}^*$ if the optimal set $\mathcal{F}^* \ne \emptyset$.*
(e) *$\mathcal{D}(\zeta^k, \zeta^{k-1}) \to 0$ if the optimal set $\mathcal{F}^* \ne \emptyset$.*

Proof (a) By the definition of ζ^k given as in (30), we have

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

Since $\mathcal{D}(\zeta^k, \zeta^{k-1}) \ge 0$ and $\mathcal{D}(\zeta^{k-1}, \zeta^{k-1}) = 0$ by Lemma 4.2 (c), it follows that

$$f(\zeta^k) \le f(\zeta^{k-1}), \quad k \ge 1.$$

(b) By Proposition 4.1, $2\mu_k^{-1} A^T[(\phi')^{\mathrm{soc}}(A\zeta^{k-1} + b) - (\phi')^{\mathrm{soc}}(A\zeta^k + b)] \in \partial f(\zeta^k)$. Hence, from the definition of subdifferential, it follows that, for any $\zeta \in \mathcal{F}$,

$$\begin{aligned}
f(\zeta) &\ge f(\zeta^k) + 2\mu_k^{-1}\langle(\phi')^{\mathrm{soc}}(A\zeta^{k-1} + b) - (\phi')^{\mathrm{soc}}(A\zeta^k + b),\ A\zeta - A\zeta^k\rangle \\
&= f(\zeta^k) + \mu_k^{-1}\,\mathrm{tr}[[(\phi')^{\mathrm{soc}}(A\zeta^{k-1} + b) - (\phi')^{\mathrm{soc}}(A\zeta^k + b)] \\
&\quad \circ [(A\zeta + b) - (A\zeta^k + b)]] \\
&= f(\zeta^k) + \mu_k^{-1}[H(A\zeta + b, A\zeta^k + b) + H(A\zeta^k + b, A\zeta^{k-1} + b) \\
&\quad - H(A\zeta + b, A\zeta^{k-1} + b)] \\
&= f(\zeta^k) + \mu_k^{-1}[\mathcal{D}(\zeta, \zeta^k) + \mathcal{D}(\zeta^k, \zeta^{k-1}) - D(\zeta, \zeta^{k-1})],
\end{aligned} \tag{37}$$

where the first equality is due to (6) and the second follows from Proposition 3.2. From this inequality and the nonnegativity of $\mathcal{D}(\zeta^k, \zeta^{k-1})$, we readily obtain the conclusion.

(c) From the result in part (b), we have

$$\mu_k[f(\zeta^{k-1}) - f(\zeta^k)] \ge \mathcal{D}(\zeta^{k-1}, \zeta^k) - \mathcal{D}(\zeta^{k-1}, \zeta^{k-1}) = \mathcal{D}(\zeta^{k-1}, \zeta^k)$$

Multiplying this inequality by σ_{k-1} and noting that $\sigma_k = \sigma_{k-1} + \mu_k$, one has

$$\sigma_{k-1} f(\zeta^{k-1}) - (\sigma_k - \mu_k) f(\zeta^k) \ge \sigma_{k-1}\mu_k^{-1}\mathcal{D}(\zeta^{k-1}, \zeta^k). \tag{38}$$

Summing up the inequalities in (38) for $k = 1, 2, \dots, N$ and using $\sigma_0 = 0$ yields

$$-\sigma_N f(\zeta^N) + \sum_{k=1}^N \mu_k f(x^k) \ge \sum_{k=1}^N \sigma_{k-1}\mu_k^{-1}\mathcal{D}(\zeta^{k-1}, \zeta^k). \tag{39}$$

On the other hand, summing the inequality in part (b) over $k = 1, 2, \dots, N$, we get

$$-\sigma_N f(\zeta) + \sum_{k=1}^{N} \mu_k f(\zeta^k) \leq \mathcal{D}(\zeta, \zeta^0) - \mathcal{D}(\zeta, \zeta^N). \tag{40}$$

Now, subtracting (39) from (40) yields that

$$\sigma_N[f(\zeta^N) - f(\zeta)] \leq \mathcal{D}(\zeta, \zeta^0) - \mathcal{D}(\zeta, \zeta^N) - \sum_{k=1}^{N} \sigma_{k-1}\mu_k^{-1}\mathcal{D}(\zeta^{k-1}, \zeta^k).$$

This together with the nonnegativity of $\mathcal{D}(\zeta^{k-1}, \zeta^k)$, implies the conclusion.

(d) Note that $f(\zeta^k) - f(\zeta) \geq 0$ for all $\zeta \in \mathcal{F}^*$. So, the result follows from part (b) directly.

(e) From part (d), we know that $\mathcal{D}(\zeta, \zeta^k)$ is nonincreasing for any $\zeta \in \mathcal{F}^*$. This, together with $\mathcal{D}(\zeta, \zeta^k) \geq 0$ for any k, implies that $\mathcal{D}(\zeta, \zeta^k)$ is convergent. Thus, we have that

$$\mathcal{D}(\zeta, \zeta^{k-1}) - \mathcal{D}(\zeta, \zeta^k) \to 0. \tag{41}$$

On the other hand, from (37) it follows that

$$0 \leq \mu_k[f(\zeta^k) - f(\zeta)] \leq \mathcal{D}(\zeta, \zeta^{k-1}) - \mathcal{D}(\zeta, \zeta^k) - \mathcal{D}(\zeta^k, \zeta^{k-1}), \quad \forall \zeta \in \mathcal{F}^*,$$

which implies that

$$\mathcal{D}(\zeta^k, \zeta^{k-1}) \leq \mathcal{D}(\zeta, \zeta^{k-1}) - \mathcal{D}(\zeta, \zeta^k), \quad \forall \zeta \in \mathcal{F}^*.$$

This, together with (41) and the nonnegativity of $\mathcal{D}(\zeta^k, \zeta^{k-1})$, yields the result. □

We have proved that the proximal-like algorithm in (29)–(30) is well-defined and satisfies some favorable properties. By this, we next establish the convergence of the algorithm.

Proposition 4.3 *Let* $\{\zeta^k\}$ *be the sequence generated by the algorithm described as in* (29)–(30), *and let* $\sigma_N = \sum_{k=1}^N \mu_k$. *Then, under Assumptions* (A1)–(A2),

(a) *if* $\sigma_N \to \infty$, *then* $\lim_{N\to+\infty} f(\zeta^N) \to f_*$;
(b) *if* $\sigma_N \to \infty$ *and the optimal set* $\mathcal{F}^* \neq \emptyset$, *then the sequence* $\{x^k\}$ *is bounded and every accumulation point is a solution of the CSOCP.*

Proof (a) From the definition of f_*, there exists a $\widehat{\zeta} \in \mathcal{F}$ such that

$$f(\widehat{\zeta}) < f_* + \epsilon, \quad \forall \epsilon > 0.$$

However, from Proposition 4.2 (c) and the nonnegativity of $\mathcal{D}(\zeta, \zeta^N)$, we have that

$$f(\zeta^N) - f(\zeta) \leq \sigma_N^{-1}\mathcal{D}(\zeta, \zeta^0), \quad \forall \zeta \in \mathcal{F}.$$

Let $\zeta = \widehat{\zeta}$ in the above inequality and take the limit with $\sigma_N \to +\infty$; we then obtain

$$\lim_{N\to+\infty} f(\zeta^N) < f_* + \epsilon.$$

Considering that ϵ is arbitrary and $f(\zeta^N) \geq f_*$, we thus have the desired result.

(b) Suppose that $\zeta^* \in \mathcal{F}^*$. Then, from Proposition 4.2 (d), $\mathcal{D}(\zeta^*, \zeta^k) \leq \mathcal{D}(\zeta^*, \zeta^0)$ for any k. This implies that $\{\zeta^k\} \subseteq L_{\mathcal{D}}(\zeta^*, \mathcal{D}(\zeta^*, \zeta^0))$. By Lemma 4.2 (d), the sequence $\{\zeta^k\}$ is then bounded. Let $\bar{\zeta} \in \mathcal{F}$ be an accumulation point of $\{\zeta^k\}$ with subsequence $\{\zeta^{k_j}\} \to \bar{\zeta}$. Then, from part (a), it follows that $f(\zeta^{k_j}) \to f_*$. On the other hand, since f is lower-semicontinuous, we have $f(\bar{\zeta}) = \liminf_{k_j\to+\infty} f(\zeta^{k_j})$. The two sides show that $f(\bar{\zeta}) \leq f(\zeta^*)$. Consequently, $\bar{\zeta}$ is a solution of the CSOCP. $\square$

5 Conclusions

In this paper, we have extended the proximal-like algorithm associated with some D-function for solving the convex programming with nonnegative constraints to the general CSOCP. The extension is based on a measure of distance $H(x, y)$ on the second-order cone, which can be generated by a single-valued function ϕ satisfying Property 3.1. Some examples are also presented, which includes the entropy-like distance. Like the proximal-like algorithm using the D-function, the algorithm has been shown, under mild assumptions, to generate a bounded sequence and its every accumulation point is a solution of the CSOCP. However, at present, we do not know what additional conditions for ϕ will guarantee that the sequence $\{\zeta^k\}$ itself converges to the solution of the considered problem, and we leave it as a future research topic.

References

1. Censor, Y., Zenios, S.A.: The proximal minimization algorithm with D-functions. J. Optim. Theory Appl. **73**, 451–464 (1992)
2. Chen, G., Teboulle, M.: Convergence analysis of a proximal-like minimization algorithm using Bregman functions. SIAM J. Optim. **3**, 538–543 (1993)
3. Alizadeh, F., Goldfarb, D.: Second-order cone programming. Math. Program. **95**, 3–51 (2003)
4. Kuo, Y.J., Mittelmann, H.D.: Interior point methods for second-order cone programming and OR applications. Comput. Optim. Appl. **28**, 255–285 (2004)
5. Lobo, M.S., Vandenberghe, L., Boyd, S., Lebret, H.: Application of second-order cone programming. Linear Algebra Appl. **284**, 193–228 (1998)
6. Chen, J.S., Chen, X., Tseng, P.: Analysis of nonsmooth vector-valued functions associated with second-order cones. Math. Program. **101**, 95–117 (2004)
7. Chen, J.S., Tseng, P.: An unconstrained smooth minimization reformulation of the second-order cone complementarity problem. Math. Program. **104**, 293–327 (2005)
8. Martinet, B.: Perturbation des methodes d'optimisation. RAIRO Anal. Numer. **12**, 153–171 (1978)
9. Moreau, J.J.: Promimité et dualité dans un espace Hilbertien. Bull. Soc. Math. Fr. **93**, 273–299 (1965)
10. Rockafellar, R.T.: Augmented Lagrangians and applications of proximial point algorithm in convex programming. Math. Oper. Res. **1**, 97–116 (1976)
11. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
12. Eggermont, P.P.B.: Multiplicative iterative algorithms for convex programming. Linear Algebra Appl. **130**, 25–42 (1990)

13. Eckstein, J.: Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. Math. Oper. Res. **18**, 202–226 (1993)
14. Teboulle, M.: Entropic proximal mappings with applications to nonlinear programming. Math. Oper. Res. **17**, 670–690 (1992)
15. Bregman, L.: The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. USSR Comput. Math. Math. Phys. **7**, 200–217 (1967)
16. Censor, Y., Lent, A.: An interval row action method for interval convex programming. J. Optim. Theory Appl. **34**, 321–353 (1981)
17. DePierro, A., Iusem, A.: A relaxed version of Bregman's method for convex programming. J. Optim. Theory Appl. **5**, 421–440 (1986)
18. Kiwiel, K.C.: Proximal minimization methods with generalized Bregman functions. SIAM J. Control Optim. **35**, 1142–1168 (1997)
19. Silva, P.J.S., Eckstein, J.: Double-regularization proximal methods, with complementarity applications. Comput. Optim. Appl. **33**, 115–156 (2006)
20. Faraut, U., Korányi, A.: Analysis on Symmetrc Cones. Oxford Mathematical Monographs. Oxford University Press, Oxford (1994)
21. Fukushima, M., Luo, Z.Q., Tseng, P.: Smoothing functions for second-order cone complementarity problems. SIAM J. Optim. **12**, 436–460 (2002)
22. Iusem, A.N.: Some properties of generalized proximal point methods for quadratic and linear programming. J. Optim. Theory Appl. **85**, 593–612 (1995)
23. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)