Comput Optim Appl (2010) 47: 477–499
DOI 10.1007/s10589-008-9227-0

An entropy-like proximal algorithm and the exponential multiplier method for convex symmetric cone programming

Jein-Shan Chen · Shaohua Pan

Received: 28 April 2006 / Revised: 6 December 2008 / Published online: 18 December 2008

Abstract We introduce an entropy-like proximal algorithm for the problem of minimizing a closed proper convex function subject to symmetric cone constraints. The algorithm is based on a distance-like function that is an extension of the Kullback-Leiber relative entropy to the setting of symmetric cones. Like the proximal algorithms for convex programming with nonnegative orthant cone constraints, we show that, under some mild assumptions, the sequence generated by the proposed algorithm is bounded and every accumulation point is a solution of the considered problem. In addition, we also present a dual application of the proposed algorithm to the symmetric cone linear program, leading to a multiplier method which is shown to possess similar properties as the exponential multiplier method (Tseng and Bertsekas in Math. Program. 60:1–19, 1993) holds.

Keywords Symmetric cone optimization · Proximal-like method · Entropy-like distance · Exponential multiplier method

1 Introduction

Symmetric cone programming provides a unified framework for linear programming, second-order cone programming and semidefinite programming, which arise from a

J.-S. Chen is a member of Mathematics Division, National Center for Theoretical Sciences, Taipei Office. The author's work is partially supported by National Science Council of Taiwan.

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

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

wide range of applications in engineering, economics, management science, optimal control, combinatorial optimization, and other fields; see [1, 16, 28] and references therein. Recently, symmetric cone programming, especially symmetric cone linear programming (SCLP), has attracted the attention of some researchers with a focus on the development of interior point methods similar to those for linear programming; see [9, 10, 23]. Although interior point methods were successfully applied for SCLPs, it is worthwhile to explore other solution methods for general convex symmetric cone optimization problems.

Let $\mathbb{A} = (\mathbb{V}, \circ, \langle\cdot,\cdot\rangle)$ be a Euclidean Jordan algebra, where $(\mathbb{V}, \langle\cdot,\cdot\rangle)$ is a finite dimensional inner product space over the real field $\mathbb{R}$ and "$\circ$" denotes the Jordan product which will be defined in the next section. Let $\mathcal{K}$ be the symmetric cone in $\mathbb{V}$. In this paper, we consider the following convex symmetric cone programming (CSCP):

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t.}\ \ x \succeq_{\mathcal{K}} 0, \end{aligned} \tag{1}$$

where $f : \mathbb{V} \to (-\infty, \infty]$ is a closed proper convex function, and $x \succeq_{\mathcal{K}} 0$ means $x \in \mathcal{K}$. In general, for any $x, y \in \mathbb{V}$, we write $x \succeq_{\mathcal{K}} y$ if $x - y \in \mathcal{K}$ and write $x \succ_{\mathcal{K}} y$ if $x - y \in \operatorname{int}(\mathcal{K})$. A function is closed if and only if it is lower semi-continuous, and a function is proper if $f(x) < \infty$ for at least one $x \in \mathbb{V}$ and $f(x) > -\infty$ for all $x \in \mathbb{V}$.

Notice that the CSCP is a special class of convex programs, and hence in principle it can be solved via general convex programming methods. One such method is the proximal point algorithm for minimizing a convex function $f(x)$ on $\mathbb{R}^n$ which generates a sequence $\{x^k\}_{k\in N}$ via the following iterative scheme:

$$x^k = \operatorname*{argmin}_{x\in\mathbb{R}^n} \left\{ f(x) + \frac{1}{\mu_k}\|x - x^{k-1}\|_2^2 \right\}, \tag{2}$$

where $\mu_k > 0$ and $\|\cdot\|_2$ denotes the Euclidean norm in $\mathbb{R}^n$. This method was first introduced by Martinet [17], based on the Moreau proximal approximation of f (see [18]), and further developed and studied by Rockafellar [20, 21]. Later, several generalizations of the proximal point algorithm have been considered where the usual quadratic proximal term in (2) is replaced by a nonquadratic distance-like function; see, for example, [5, 7, 8, 14, 25]. Among others, the algorithms using an entropy-like distance [13, 14, 25, 26] for minimizing a convex function $f(x)$ subject to $x \in \mathbb{R}^n_+$, generate the iterates by

$$\begin{cases} x^0 \in \mathbb{R}^n_{++} \\ x^k = \operatorname{argmin}_{x\in\mathbb{R}^n_+} \left\{ f(x) + \frac{1}{\mu_k} d_\varphi(x, x^{k-1}) \right\}, \end{cases} \tag{3}$$

where $d_\varphi(\cdot,\cdot) : \mathbb{R}^n_{++} \times \mathbb{R}^n_{++} \to \mathbb{R}^n_+$ is the entropy-like distance defined by

$$d_\varphi(x, y) = \sum_{i=1}^n y_i \varphi(x_i/y_i) \tag{4}$$

with φ satisfying certain conditions; see [13, 14, 25, 26]. An important choice of φ is the function $\varphi(t) = t \ln t - t + 1$, for which the corresponding d_φ given by

$$d_\varphi(x, y) = \sum_{i=1}^{n} \Big[x_i \ln x_i - x_i \ln y_i + y_i - x_i \Big] \tag{5}$$

is the popular Kullback-Leibler entropy from statistics and that is the "entropy" terminology stems from. One key feature of entropic proximal methods is that they generate a sequence staying in the interior of $\mathbb{R}^n_+$ automatically, and thus eliminates the combinatorial nature of the problem. One of the main applications of such proximal methods is to the dual of smooth convex programs, yielding twice continuously differentiable nonquadratic augmented Lagrangians and thereby allowing the usage of Newton's methods.

The main purpose of this paper is to propose an interior proximal-like method and the corresponding dual augmented Lagrangian method for the CSCP (1). Specifically, by using the Euclidean Jordan algebraic techniques, we extend the entropy-like proximal algorithm defined by (3)–(4) with $\varphi(t) = t \ln t - t + 1$ to the solution of (1). For the proposed algorithm, we establish a global convergence estimate in terms of the objective value, and moreover present a dual application to the standard SCLP, which leads to an exponential multiplier method shown to possess properties analogous to the method proposed by [3, 27] for convex programming over nonnegative orthant cone $\mathbb{R}^n_+$.

The paper is organized as follows. Section 2 reviews some basic concepts and materials on Euclidean Jordan algebras which are needed in the analysis of the algorithm. In Sect. 3, we introduce a distance-like function H to measure how close between two points in the symmetric cone $\mathcal{K}$ and investigate some related properties. Furthermore, we outline a basic proximal-like algorithm with the measure function H. The convergence analysis of the algorithm is the main content of Sect. 4. In Sect. 5, we consider a dual application of the algorithm to the SCLP and establish the convergence results for the corresponding multiplier method. We close this paper with some remarks in Sect. 6.

2 Preliminaries on Euclidean Jordan algebra

This section recalls some concepts and results on Euclidean Jordan algebras that will be used in the subsequent sections. More detailed expositions of Euclidean Jordan algebras can be found in Koecher's lecture notes [15] and Faraut and Korányi's monograph [11].

Let $\mathbb{V}$ be a finite dimensional inner space endowed with a bilinear mapping $(x, y) \mapsto x \circ y : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$. For a given $x \in \mathbb{V}$, let $\mathcal{L}(x)$ be the linear operator of $\mathbb{V}$ defined by

$$\mathcal{L}(x)y := x \circ y \quad \text{for every } y \in \mathbb{V}.$$

The pair $(\mathbb{V}, \circ)$ is called a *Jordan algebra* if, for all $x, y \in \mathbb{V}$,

(i) $x \circ y = y \circ x$,

(ii) $x \circ (x^2 \circ y) = x^2 \circ (x \circ y)$, where $x^2 := x \circ x$.

In a Jordan algebra $(\mathbb{V}, \circ)$, $x \circ y$ is said to be the *Jordan product* of x and y. Note that a Jordan algebra is not associative, i.e., $x \circ (y \circ z) = (x \circ y) \circ z$ may not hold in general. If for some element $e \in \mathbb{V}$, $x \circ e = e \circ x = x$ for all $x \in \mathbb{V}$, then e is called a *unit element* of the Jordan algebra $(\mathbb{V}, \circ)$. The unit element, if exists, is unique. A Jordan algebra does not necessarily have a unit element. For $x \in \mathbb{V}$, let $\zeta(x)$ be the degree of the minimal polynomial of x, which can be equivalently defined as

$$\zeta(x) := \min \left\{ k : \{e, x, x^2, \cdots, x^k\} \text{ are linearly dependent} \right\}.$$

Then the *rank* of $(\mathbb{V}, \circ)$ is well defined by $r := \max\{\zeta(x) : x \in \mathbb{V}\}$.

A Jordan algebra $(\mathbb{V}, \circ)$, with a unit element $e \in \mathbb{V}$, defined over the real field $\mathbb{R}$ is called a *Euclidean Jordan algebra* or *formally real Jordan algebra*, if there exists a positive definite symmetric bilinear form on $\mathbb{V}$ which is associative; in other words, there exists on $\mathbb{V}$ an inner product denoted by $\langle \cdot, \cdot \rangle_{\mathbb{V}}$ such that for all $x, y, z \in \mathbb{V}$:

(iii) $\langle x \circ y, z \rangle_{\mathbb{V}} = \langle y, x \circ z \rangle_{\mathbb{V}}$.

In a Euclidean Jordan algebra $\mathbb{A} = (\mathbb{V}, \circ, \langle \cdot, \cdot \rangle_{\mathbb{V}})$, we define the set of squares as

$$\mathcal{K} := \left\{ x^2 : x \in \mathbb{V} \right\}.$$

By [11, Theorem III. 2.1], $\mathcal{K}$ is a symmetric cone. This means that $\mathcal{K}$ is a self-dual closed convex cone with nonempty interior and for any two elements $x, y \in \text{int}(\mathcal{K})$, there exists an invertible linear transformation $\mathcal{T} : \mathbb{V} \to \mathbb{V}$ such that $\mathcal{T}(\mathcal{K}) = \mathcal{K}$ and $\mathcal{T}(x) = y$.

Here are two popular examples of Euclidean Jordan algebras. Let $\mathbb{S}^n$ be the space of $n \times n$ real symmetric matrices with the inner product given by $\langle X, Y \rangle_{\mathbb{S}^n} := \text{Tr}(XY)$, where XY is the matrix multiplication of X and Y and $\text{Tr}(XY)$ is the trace of XY. Then, $(\mathbb{S}^n, \circ, \langle \cdot, \cdot \rangle_{\mathbb{S}^n})$ is a Euclidean Jordan algebra with the Jordan product defined by

$$X \circ Y \ := \ (XY + YX)/2, \quad X, Y \in \mathbb{S}^n.$$

In this case, the unit element is the identity matrix I in $\mathbb{S}^n$ and the cone of squares $\mathcal{K}$ is the set of all positive semidefinite matrices in $\mathbb{S}^n$. Let $\mathbb{R}^n$ be the Euclidean space of dimension n with the usual inner product $\langle x, y \rangle_{\mathbb{R}^n} = x^T y$. For any $x = (x_1, x_2)$, $y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, define $x \circ y := (x^T y, x_1 y_2 + y_1 x_2)^T$. Then $(\mathbb{R}^n, \circ, \langle \cdot, \cdot \rangle_{\mathbb{R}^n})$ is a Euclidean Jordan algebra, also called *the quadratic forms algebra*. In this algebra, the unit element $e = (1, 0, \ldots, 0)^T$ and $\mathcal{K} = \{x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1} : x_1 \geq \|x_2\|_2\}$.

Recall that an element $c \in \mathbb{V}$ is said to be *idempotent* if $c^2 = c$. Two idempotents c and q are said to be *orthogonal* if $c \circ q = 0$. One says that $\{c_1, c_2, \ldots, c_k\}$ is a complete system of orthogonal idempotents if

$$c_j^2 = c_j, \quad c_j \circ c_i = 0 \quad \text{if } j \neq i \text{ for all } j, i = 1, 2, \ldots, k, \quad \text{and} \quad \sum_{j=1}^{k} c_j = e.$$

An idempotent is said to be *primitive* if it is nonzero and cannot be written as the sum of two other nonzero idempotents. We call a complete system of orthogonal primitive idempotents a *Jordan frame*. Then we have the following spectral decomposition theorem.

Theorem 2.1 [11, Theorem III. 1.2] *Suppose that* $\mathbb{A} = (\mathbb{V}, \circ, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ *is a Euclidean Jordan algebra and the rank of* $\mathbb{A}$ *is* r. *Then for any* $x \in \mathbb{V}$, *there exists a Jordan frame* $\{c_1, c_2, \ldots, c_r\}$ *and real numbers* $\lambda_1(x), \lambda_2(x), \ldots, \lambda_r(x)$, *arranged in the decreasing order* $\lambda_1(x) \geq \lambda_2(x) \geq \cdots \geq \lambda_r(x)$, *such that* $x = \sum_{j=1}^{r} \lambda_j(x) c_j$. *The numbers* $\lambda_j(x)$ (*counting multiplicities*), *which are uniquely determined by* x, *are called the eigenvalues*, $\sum_{j=1}^{r} \lambda_j(x) c_j$ *the spectral decomposition of* x, *and* $\mathrm{tr}(x) = \sum_{j=1}^{r} \lambda_j(x)$ *the trace of* x.

From [11, Proposition III.1.5], a Jordan algebra $(\mathbb{V}, \circ)$ with a unit element $e \in \mathbb{V}$ is Euclidean if and only if the symmetric bilinear form $\mathrm{tr}(x \circ y)$ is positive definite. Therefore, we may define another inner product on $\mathbb{V}$ by

$$\langle x, y\rangle \ := \ \mathrm{tr}(x \circ y) \quad \forall x, y \in \mathbb{V}.$$

By the associativity of $\mathrm{tr}(\cdot)$ [11, Proposition II. 4.3], the inner product $\langle\cdot,\cdot\rangle$ is associative, i.e., for all $x, y, z \in \mathbb{V}$, there holds that $\langle x, y \circ z\rangle = \langle y, x \circ z\rangle$. Thus, the operator $\mathcal{L}(x)$ for each $x \in \mathbb{V}$ is symmetric with respect to the inner product $\langle\cdot,\cdot\rangle$ in the sense that

$$\langle \mathcal{L}(x)y, z\rangle = \langle y, \mathcal{L}(x)z\rangle \quad \forall y, z \in \mathbb{V}.$$

In the sequel, we let $\|\cdot\|$ be the norm on $\mathbb{V}$ induced by the inner product $\langle\cdot,\cdot\rangle$, i.e.,

$$\|x\| := \sqrt{\langle x, x\rangle} = \left(\sum_{j=1}^{r} \lambda_j^2(x)\right)^{1/2} \quad \forall x \in \mathbb{V},$$

and denote by $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ the smallest and the largest eigenvalue of x, respectively. Then, by Lemma 13 and the proof of Lemma 14 in [23], we can prove the following lemma.

Lemma 2.1 *Let* $x, y \in \mathbb{V}$, *then we can bound the minimum eigenvalue of* $x + y$ *as follows*:

$$\lambda_{\min}(x) + \lambda_{\min}(y) \leq \lambda_{\min}(x + y) \leq \lambda_{\min}(x) + \lambda_{\max}(y).$$

Let $g : \mathbb{R} \to \mathbb{R}$ be a scalar-valued function. Then, it is natural to define a vector-valued function associated with the Euclidean Jordan algebra $(\mathbb{V}, \circ, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ by

$$g^{\mathrm{sc}}(x) := g(\lambda_1(x))c_1 + g(\lambda_2(x))c_2 + \cdots + g(\lambda_r(x))c_r, \tag{6}$$

where $x \in \mathbb{V}$ has the spectral decomposition $x = \sum_{j=1}^{r} \lambda_j(x) c_j$. This function is called the Löwner operator in [24] and was shown to have the following important property.

Lemma 2.2 [24, Theorem 13] *For any $x=\sum_{j=1}^r \lambda_j(x)c_j$, let g^{sc} be defined by* (6). *Then g^{sc} is* (*continuously*) *differentiable at x if and only if g is* (*continuously*) *differentiable at all $\lambda_j(x)$. Furthermore, the derivative of g^{sc} at x, for any $h\in\mathbb{V}$, is given by*

$$(g^{\mathrm{sc}})'(x)h=\sum_{j=1}^r g'(\lambda_j(x))\langle c_j,h\rangle c_j+\sum_{1\le j<l\le r} 4[\lambda_i(x),\lambda_j(x)]_g c_j\circ(c_l\circ h)$$

with

$$[\lambda_i(x),\lambda_j(x)]_g:=\frac{g(\lambda_i(x))-g(\lambda_j(x))}{\lambda_i(x)-\lambda_j(x)}\quad \forall\, i,j=1,2,\ldots,r \text{ and } i\neq j.$$

In fact, the Jacobian $(g^{\mathrm{sc}})'(\cdot)$ is a linear and symmetric operator, and can be written as

$$(g^{\mathrm{sc}})'(x)=\sum_{j=1}^r g'(\lambda_j(x))\mathcal{Q}(c_j)+2\sum_{i,j=1,i\neq j}^r [\lambda_i(x),\lambda_j(x)]_g\mathcal{L}(c_j)\mathcal{L}(c_i),\tag{7}$$

where $\mathcal{Q}(x):=2\mathcal{L}^2(x)-\mathcal{L}(x^2)$ for any $x\in\mathbb{V}$ is called the *quadratic representation* of $\mathbb{V}$.

Finally, we recall the spectral function generated by a symmetric function. Let $\mathbb{P}$ denote the set of all permutations of r-dimensional vectors. A subset of $\mathbb{R}^r$ is said to be *symmetric* if it remains unchanged under every permutation of $\mathbb{P}$. Let S be a symmetric set in $\mathbb{R}^r$. A real-valued function $f:S\to\mathbb{R}$ is said to be *symmetric* if for every permutation $P\in\mathbb{P}$ and each $s\in S$, there holds that $f(Ps)=f(s)$. For any $x\in\mathbb{V}$ with the spectral decomposition $x=\sum_{j=1}^r\lambda_j(x)c_j$, define $K:=\{x\in\mathbb{V}\mid \lambda(x)\in S\}$ with $\lambda(x)=(\lambda_1(x),\ldots,\lambda_r(x))^T$ being the spectral vector of x. Then $F:K\to\mathbb{R}$ defined by

$$F(x):=f(\lambda(x))\tag{8}$$

is called the *spectral function generated by* f. From Theorem 41 of [2], F is (strictly) convex if f is (strictly) convex.

Unless otherwise stated, in the rest of this paper, the notation $\mathbb{A}=(\mathbb{V},\circ,\langle\cdot,\cdot\rangle)$ represents a Euclidean Jordan algebra of rank r and $\dim(\mathbb{V})=n$. For a closed proper convex function $f:\mathbb{V}\to(-\infty,+\infty]$, we denote by $\operatorname{dom} f:=\{x\in\mathbb{V}\mid f(x)<+\infty\}$ the domain of f. The subdifferential of f at $x_0\in\mathbb{V}$ is the convex set

$$\partial f(x_0)=\{\xi\in\mathbb{V}\mid f(x)\ge f(x_0)+\langle\xi,x-x_0\rangle\ \forall\, x\in\mathbb{V}\}.\tag{9}$$

Since $\langle x,y\rangle=\operatorname{tr}(x\circ y)$ for any $x,y\in\mathbb{V}$, the above subdifferential set is equivalent to

$$\partial f(x_0)=\left\{\xi\in\mathbb{V}\mid f(x)\ge f(x_0)+\operatorname{tr}\big(\xi\circ(x-x_0)\big),\ x\in\mathbb{V}\right\}.\tag{10}$$

For a sequence $\{x^k\}_{k\in N}$, the notation N denotes the set of natural numbers.

3 Entropy-like proximal algorithm

To solve the CSCP (1), we suggest the following proximal-like minimization algorithm:

$$\begin{cases} x^0 \succ_{\mathcal{K}} 0 \\ x^k = \operatorname{argmin}_{x \succeq_{\mathcal{K}} 0} \big\{ f(x) + \frac{1}{\mu_k} H(x, x^{k-1}) \big\}, \end{cases} \tag{11}$$

where $\mu_k > 0$ and $H : \mathbb{V} \times \mathbb{V} \to (-\infty, +\infty]$ is defined by

$$H(x, y) := \begin{cases} \mathrm{tr}(x \circ \ln x - x \circ \ln y + y - x) & \forall\, x \in \mathrm{int}(\mathcal{K}),\ y \in \mathrm{int}(\mathcal{K}), \\ +\infty & \text{otherwise.} \end{cases} \tag{12}$$

This algorithm is indeed a proximal-type one except that the classical quadratic term $\|x - x^{k-1}\|_2^2$ is replaced by the distance-like function H to guarantee that $\{x^k\}_{k \in N} \subset \mathrm{int}(\mathcal{K})$, thus leading to an interior proximal-like method (see Proposition 4.1).

By the definition of Löwner operator, clearly, the function $H(x, y)$ is well-defined for all $x, y \in \mathrm{int}(\mathcal{K})$. Moreover, the domain of $x \in \mathrm{int}(\mathcal{K})$ can be extended to $x \in \mathcal{K}$ by adopting the convention $0 \ln 0 \equiv 0$. The function H is a natural extension of the distance-like entropy function in (5), and is used to measure the "distance" between two points in $\mathcal{K}$. In fact, H will become the entropy function d_φ in (5) if the Euclidean Jordan algebra $\mathbb{A}$ is specified as $(\mathbb{R}^n, \circ, \langle \cdot, \cdot \rangle_{\mathbb{R}^n})$ with "$\circ$" denoting the componentwise product of two vectors in $\mathbb{R}^n$ and $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ the usual Euclidean inner product. As shown by Proposition 3.1 below, most of the important properties, but not all, of d_φ also hold for H.

The following two technical lemmas will be used to investigate the favorable properties of the distance measure H. Lemma 3.1 states an extension of Von Neumann inequality to Euclidean Jordan algebras, and Lemma 3.2 gives the properties of $\mathrm{tr}(x \circ \ln x)$.

Lemma 3.1 [2] *For any $x, y \in \mathbb{V}$, we have $\mathrm{tr}(x \circ y) \le \sum_{j=1}^r \lambda_j(x)\lambda_j(y) = \lambda(x)^T \lambda(y)$, where $\lambda(x)$ and $\lambda(y)$ are the spectral vectors of x and y, respectively.*

Lemma 3.2 *For any $x \in \mathcal{K}$, let $\Phi(x) := \mathrm{tr}(x \circ \ln x)$. Then, we have the following results.*

(a) *$\Phi(x)$ is the spectral function generated by the symmetric entropy function*

$$\phi(u) = \sum_{j=1}^r u_j \ln u_j \quad \forall u \in \mathbb{R}_+^r. \tag{13}$$

(b) *$\Phi(x)$ is continuously differentiable on $\mathrm{int}(\mathcal{K})$ with $\nabla \Phi(x) = \ln x + e$.*
(c) *The function $\Phi(x)$ is strictly convex over $\mathcal{K}$.*

Proof (a) Suppose that x has the spectral decomposition $x = \sum_{j=1}^r \lambda_j(x) c_j$. Let

$$g(t) = t \ln t \quad \forall\, t \in \mathbb{R}.$$

From Sect. 2, it follows that the vector-valued function $x \circ \ln x$ is the Löwner function $g^{\rm sc}(x)$, i.e., $g^{\rm sc}(x) = x \circ \ln x$. Clearly, $g^{\rm sc}$ is well-defined for any $x \in \mathcal{K}$ and

$$g^{\rm sc}(x) = x \circ \ln x = \sum_{j=1}^{r} \lambda_j(x) \ln(\lambda_j(x)) c_j.$$

Therefore,

$$\Phi(x) = \mathrm{tr}(x \circ \ln x) = \mathrm{tr}(g^{\rm sc}(x)) = \sum_{j=1}^{r} \lambda_j(x) \ln(\lambda_j(x)) = \phi(\lambda(x))$$

with $\phi : \mathbb{R}^r_+ \to \mathbb{R}$ given by (13). Since the function ϕ is symmetric, $\Phi(x)$ is the spectral function generated by the symmetric function ϕ in view of (8).

(b) From Lemma 2.2, $g^{\rm sc}(x) = x \circ \ln x$ is continuously differentiable on $\mathrm{int}(\mathcal{K})$. Thus, $\Phi(x)$ is also continuously differentiable on $\mathrm{int}(\mathcal{K})$ because Φ is the composition of the trace function (clearly continuously differentiable) and $g^{\rm sc}$. Now, it remains to find its gradient formula. From the fact that $\mathrm{tr}(x \circ y) = \langle x, y \rangle$, we have

$$\Phi(x) = \mathrm{tr}(x \circ \ln x) = \langle x, \ln x \rangle.$$

Applying the chain rule for inner product of two functions, we then obtain

$$\nabla \Phi(x) = \ln x + (\nabla \ln x) x = \ln x + (\ln x)' x. \tag{14}$$

On the other hand, from formula (7) it follows that for any $h \in \mathbb{V}$,

$$(\ln x)' h = \sum_{j=1}^{r} \frac{1}{\lambda_j(x)} \langle c_j, h \rangle c_j + \sum_{1 \le j < l \le r} \frac{\ln(\lambda_j(x)) - \ln(\lambda_l(x))}{\lambda_j(x) - \lambda_l(x)} c_j \circ (c_l \circ h).$$

By this and the spectral decomposition of x, it is easy to compute that

$$\begin{aligned}
(\ln x)' x &= \sum_{j=1}^{r} \frac{1}{\lambda_j(x)} \langle c_j, x \rangle c_j + \sum_{1 \le j < l \le r} \frac{\ln(\lambda_j(x)) - \ln(\lambda_l(x))}{\lambda_j(x) - \lambda_l(x)} c_j \circ (c_l \circ x) \\
&= \sum_{j=1}^{r} \frac{1}{\lambda_j(x)} \lambda_j(x) c_j + \sum_{1 \le j < l \le r} \frac{\ln(\lambda_j(x)) - \ln(\lambda_l(x))}{\lambda_j(x) - \lambda_l(x)} \lambda_l(x) c_j \circ c_l \\
&= e.
\end{aligned}$$

Combining with (14), we readily obtain the desired result.

(c) Note that the function ϕ in (13) is strictly convex over $\mathbb{R}^n_+$. Therefore, the conclusion immediately follows from part (a) and Theorem 41 of [2]. □

Next we study the favorable properties of the distance-like function H. These properties play a crucial role in the convergence analysis of the algorithm defined by (11)–(12).

Proposition 3.1 *Let $H(x, y)$ be defined by (12). Then the following results hold.*

(a) *$H(x, y)$ is continuous on $\mathcal{K} \times \mathrm{int}(\mathcal{K})$ and $H(\cdot, y)$ is strictly convex for any $y \in \mathrm{int}(\mathcal{K})$.*
(b) *For any fixed $y \in \mathrm{int}(\mathcal{K})$, $H(\cdot, y)$ is continuously differentiable on $\mathrm{int}(\mathcal{K})$ with*

$$\nabla_x H(x, y) = \ln x - \ln y.$$

(c) *$H(x, y) \geq 0$ for any $x \in \mathcal{K}$ and $y \in \mathrm{int}(\mathcal{K})$, and $H(x, y) = 0$ if and only if $x = y$.*
(d) *$H(x, y) \geq d(\lambda(x), \lambda(y)) \geq 0$ for any $x \in \mathcal{K}$, $y \in \mathrm{int}(\mathcal{K})$, where $d(\cdot, \cdot)$ is defined by*

$$d(u, v) = \sum_{i=1}^{n} \Big[u_i \ln u_i - u_i \ln v_i + v_i - u_i \Big] \quad \forall\, u \in \mathbb{R}^r_+,\ v \in \mathbb{R}^r_{++}. \tag{15}$$

(e) *For fixed $y \in \mathrm{int}(\mathcal{K})$, the level sets $L_H(x, \gamma) := \{x \in \mathcal{K} \mid H(x, y) \leq \gamma\}$ are bounded for all $\gamma \geq 0$, and for fixed $x \in \mathcal{K}$, the level sets $L_H(y, \gamma) := \{y \in \mathrm{int}(\mathcal{K}) \mid H(x, y) \leq \gamma\}$ are bounded for all $\gamma \geq 0$.*

Proof (a) Since $x \circ \ln x$, $x \circ \ln y$ are continuous in $x \in \mathcal{K}$ and $y \in \mathrm{int}(\mathcal{K})$, and the trace function is also continuous, the function H is continuous over $\mathcal{K} \times \mathrm{int}(\mathcal{K})$. Notice that

$$H(x, y) = \Phi(x) - \mathrm{tr}(x \circ \ln y) + \mathrm{tr}(y) - \mathrm{tr}(x), \tag{16}$$

$\Phi(x)$ is strictly convex over $\mathcal{K}$ by Lemma 3.2(c), and the other terms on the right hand side of (16) are clearly convex for any fixed $y \in \mathrm{int}(\mathcal{K})$. Therefore, $H(\cdot, y)$ is strictly convex for any fixed $y \in \mathrm{int}(\mathcal{K})$.

(b) From the expression of $H(x, y)$ given by (16) and Lemma 3.2(b), obviously, the function $H(\cdot, y)$ is continuously differentiable in $\mathrm{int}(\mathcal{K})$. Moreover,

$$\nabla_x H(x, y) = \nabla_x \Phi(x) - \ln y - e = \ln x - \ln y.$$

(c) From the definition of $\Phi(x)$ and its gradient formula shown as in Lemma 3.2(b),

$$\begin{aligned} &\Phi(x) - \Phi(y) - \langle \Phi'(y),\ x - y\rangle \\ &\quad = \mathrm{tr}(x \circ \ln x) - \mathrm{tr}(y \circ \ln y) - \langle \ln y + e, x - y\rangle \\ &\quad = \mathrm{tr}(x \circ \ln x) - \mathrm{tr}(x \circ \ln y) - \mathrm{tr}(x) + \mathrm{tr}(y) \\ &\quad = H(x, y) \end{aligned} \tag{17}$$

for any $x \in \mathcal{K}$ and $y \in \mathrm{int}(\mathcal{K})$. In addition, the strict convexity of Φ implies that

$$\Phi(x) - \Phi(y) - \langle \Phi'(y),\ x - y\rangle \geq 0$$

and the equality holds if and only if $x = y$. The two sides readily give the desired result.

(d) Using the definition of H and Lemma 3.1, we have for all $x \in \mathcal{K}$ and $y \in \text{int}(\mathcal{K})$,

$$\begin{aligned} H(x, y) &= \text{tr}(x \circ \ln x + y - x) - \text{tr}(x \circ \ln y) \\ &\geq \text{tr}(x \circ \ln x + y - x) - \sum_{j=1}^{r} \lambda_j(x) \ln(\lambda_j(y)) \\ &= \sum_{j=1}^{r} \Big[\lambda_j(x) \ln(\lambda_j(x)) + \lambda_j(y) - \lambda_j(x)\Big] - \sum_{j=1}^{r} \lambda_j(x) \ln(\lambda_j(y)) \\ &= \sum_{j=1}^{r}\Big[\lambda_j(x) \ln(\lambda_j(x)) - \lambda_j(x) \ln(\lambda_j(y)) + \lambda_j(y) - \lambda_j(x)\Big] \\ &= d(\lambda(x), \lambda(y)). \end{aligned}$$

The nonnegativity of $d(\lambda(x), \lambda(y))$ is direct by the definition of $d(\cdot, \cdot)$ in (15).

(e) For fixed $y \in \text{int}(\mathcal{K})$, from part (d) we have $L_H(x, \gamma) \subseteq \{x \in \mathcal{K} \mid d(\lambda(x), \lambda(y)) \leq \gamma\}$ for all $\gamma \geq 0$. Since the sets $\{u \in \mathbb{R}^r_+ \mid d(u, v) \leq \gamma\}$ are bounded for all $\gamma \geq 0$ by [26, Lemma 2.3], we have from the continuity of $\lambda(\cdot)$ that the sets $L_H(x, \gamma)$ are bounded for all $\gamma \geq 0$. Similarly, for fixed $x \in \mathcal{K}^n$, using Lemma 2.1(i) of [26] the sets $L_H(y, \gamma)$ are bounded for all $\gamma \geq 0$. $\square$

Proposition 3.2 *Let $H(x, y)$ be defined by (12). Suppose $\{x^k\}_{k \in N} \subseteq \mathcal{K}$ and $\{y^k\}_{k \in N} \subset \text{int}(\mathcal{K})$ are bounded sequences such that $H(x^k, y^k) \to 0$. Then, as $k \to \infty$, we have*

(a) $\lambda_j(x^k) - \lambda_j(y^k) \to 0$ *for all* $j = 1, 2, \ldots, r$.
(b) $\text{tr}(x^k - y^k) \to 0$.

Proof (a) From Proposition 3.1(d), $H(x^k, y^k) \geq d(\lambda(x^k), \lambda(y^k)) \geq 0$. Hence, $H(x^k, y^k) \to 0$ implies $d(\lambda(x^k), \lambda(y^k)) \to 0$. By the definition of $d(\cdot, \cdot)$ given by (15),

$$d(\lambda(x^k), \lambda(y^k)) = \sum_{j=1}^{r} \lambda_j(y^k) \varphi\left(\lambda_j(x^k)/\lambda_j(y^k)\right)$$

with $\varphi(t) = t \ln t - t + 1$ $(t \geq 0)$. Since $\varphi(t) \geq 0$ for any $t \geq 0$, each term of the above sum is nonnegative, and consequently, $d(\lambda(x^k), \lambda(y^k)) \to 0$ implies

$$\lambda_j(y^k) \varphi\left(\lambda_j(x^k)/\lambda_j(y^k)\right) \to 0, \quad j = 1, 2, \ldots, r.$$

This is equivalent to saying that

$$\lambda_j(x^k) \ln(\lambda_j(x^k)) - \lambda_j(x^k) \ln(\lambda_j(y^k)) + \lambda_j(y^k) - \lambda_j(x^k) \to 0, \quad j = 1, 2, \ldots, r.$$

Since $\{\lambda_j(x^k)\}$ and $\{\lambda_j(y^k)\}$ are bounded, using Lemma A.1 of [6] then yields that

$$\lambda_j(x^k) - \lambda_j(y^k) \to 0 \quad \text{for all } j = 1, 2, \ldots, r.$$

(b) Since $\mathrm{tr}(x^k - y^k) = \sum_{j=1}^r (\lambda_j(x^k) - \lambda_j(y^k))$, the result follows from part (a). □

To close this section, we present two useful relations for the function H, which can be easily verified by using the definition of H and recalling the nonnegativity of H.

Proposition 3.3 *Let $H(x, y)$ be defined by* (12). *For all $x, y \in \mathrm{int}(\mathcal{K})$ and $z \in \mathcal{K}$, we have*

(a) $H(z, x) - H(z, y) = \mathrm{tr}(z \circ \ln y - z \circ \ln x + x - y)$.
(b) $\mathrm{tr}((z - y) \circ (\ln y - \ln x)) = H(z, x) - H(z, y) - H(y, x) \le H(z, x) - H(z, y)$.

4 Convergence analysis of the algorithm

The convergence analysis of the entropy-like proximal algorithm defined by (11)–(12) is similar to that of the proximal point method [12] for convex minimization problems and the proximal-like algorithm [4] using Bregman functions. In this section, we will present a global convergence rate estimate for the algorithm in terms of function values. For this purpose, we need to make the following assumptions for the CSCP (1):

(A) $\inf\{f(x) \mid x \succeq_{\mathcal{K}} 0\} := f_* > -\infty$ and $\mathrm{dom} f \cap \mathrm{int}(\mathcal{K}) \ne \emptyset$.

In addition, we denote by $\mathcal{X}^* := \{x \in \mathcal{K} \mid f(x) = f_*\}$ the solution set of (1).

First, we show that the algorithm in (11)–(12) is well defined, i.e., it generates a sequence $\{x^k\}_{k\in N} \subset \mathrm{int}(\mathcal{K})$. This is a direct consequence of Proposition 4.1 below. For the proof of Proposition 4.1, we need the following technical lemma.

Lemma 4.1 *For any $x \succeq_{\mathcal{K}} 0$, $y \succ_{\mathcal{K}} 0$ and $\mathrm{tr}(x \circ y) = 0$, we always have $x = 0$.*

Proof From the self-duality of $\mathcal{K}$ and [11, Proposition I. 1.4], we have that

$$u \in \mathrm{int}(\mathcal{K}) \iff \langle u, v\rangle > 0, \quad \forall\, 0 \ne v \in \mathcal{K},$$

which together with $\mathrm{tr}(x \circ y) = \langle x, y\rangle$ immediately implies $x = 0$. □

Proposition 4.1 *For any fixed $y \in \mathrm{int}(\mathcal{K})$ and $\mu > 0$, let $F_\mu(\cdot, y) := f(\cdot) + \mu^{-1} H(\cdot, y)$. Then, under assumption* (A), *we have the following results.*

(a) *The function $F_\mu(\cdot, y)$ has bounded level sets.*
(b) *There exists a unique $x(y, \mu) \in \mathrm{int}(\mathcal{K})$ such that*

$$x(y, \mu) := \operatorname*{argmin}_{x \succeq_{\mathcal{K}} 0} F_\mu(x, y)$$

and

$$\mu^{-1}(\ln y - \ln x(y, \mu)) \in \partial f(x(y, \mu)). \tag{18}$$

Proof (a) Since $F_\mu(\cdot, y)$ is convex, to show that $F_\mu(\cdot, y)$ has bounded level sets, it suffices to show that for any $\nu \geq f_* > -\infty$, the level set $L(\nu) := \{\, x \mid F_\mu(x, y) \leq \nu\}$ is bounded. Let $\nu' := (\nu - f_*)\mu$. Clearly, we have $L(\nu) \subset L_H(x, \nu')$. Moreover, by Proposition 3.1 (e), $L_H(x, \nu')$ is bounded. Therefore, $L(\nu)$ is bounded.

(b) From part (a), $F_\mu(\cdot, y)$ has bounded level sets, which in turn implies the existence of the minimum point $x(y, \mu)$. Also, the strict convexity of $F_\mu(x, y)$ by Proposition 3.1(a) guarantees the uniqueness. Under assumption (A), using Proposition 3.1(b) and the optimality condition for the minimization problem $\mathrm{argmin}_{x \succeq_{\mathcal{K}} 0} F_\mu(x, y)$ gives that

$$0 \in \partial f(x(y, \mu)) + \mu^{-1}\Big(\ln x(y, \mu) - \ln y\Big) + \partial\delta(x(y, \mu)|\mathcal{K}), \tag{19}$$

where $\delta(z|\mathcal{K}) = 0$ if $z \succeq_{\mathcal{K}} 0$ and $+\infty$ otherwise. We will show that $\partial\delta(x(y, \mu)|\mathcal{K}) = \{0\}$ and hence the desired result follows. Notice that for any $x^k \in \mathrm{int}(\mathcal{K})$ with $x^k \to x \in \mathrm{bd}(\mathcal{K})$,

$$\|\ln x^k\| = \left(\sum_{j=1}^{r}[\ln(\lambda_j(x^k))]^2\right)^{1/2} \to +\infty,$$

where the second relation is due to the continuity of $\lambda_j(\cdot)$ and $\ln t$. Consequently,

$$\|\nabla_x H(x^k, y)\| = \|\ln x^k - \ln y\| \geq \|\ln x^k\| - \|\ln y\| \to +\infty.$$

This by [19, pp. 251] means that $H(\cdot, y)$ is essentially smooth, and then $\partial_x H(x, y) = \emptyset$ for all $x \in \mathrm{bd}(\mathcal{K})$ by [19, Theorem 26.1]. Together with (19), we must have $x(y, \mu) \in \mathrm{int}(\mathcal{K})$. Furthermore, from [19, pp. 226], it follows that

$$\partial\delta(z|\mathcal{K}) = \{v \in \mathbb{V} \mid v \preceq_{\mathcal{K}} 0,\ \mathrm{tr}(v \circ z) = 0\}.$$

Using Lemma 4.1 then yields $\partial\delta(x(y, \mu)|\mathcal{K}) = \{0\}$. Thus, the proof is completed. □

Next we establish several important properties for the algorithm defined by (11)–(12), from which our convergence result will follow.

Proposition 4.2 *Let $\{x^k\}_{k\in N}$ be the sequence generated by the algorithm defined by (11)–(12), and let $\sigma_m := \sum_{k=1}^{m} \mu_k$. Then, the following results hold.*

(a) *The sequence $\{f(x^k)\}_{k\in N}$ is nonincreasing.*
(b) *$\mu_k(f(x^k) - f(x)) \leq H(x, x^{k-1}) - H(x, x^k)$ for all $x \in \mathcal{K}$.*
(c) *$\sigma_m(f(x^m) - f(x)) \leq H(x, x^0) - H(x, x^m)$ for all $x \in \mathcal{K}$.*
(d) *$\{H(x, x^k)\}_{k\in N}$ is nonincreasing for any $x \in \mathcal{X}^*$, if the optimal set $\mathcal{X}^* \neq \emptyset$.*
(e) *$H(x^k, x^{k-1}) \to 0$ and $\mathrm{tr}(x^k - x^{k-1}) \to 0$, if the optimal set $\mathcal{X}^* \neq \emptyset$.*

Proof (a) From the definition of x^k in (11), it follows that

$$f(x^k) + \mu_k^{-1} H(x^k, x^{k-1}) \leq f(x^{k-1}) + \mu_k^{-1} H(x^{k-1}, x^{k-1}).$$

Since $H(x^k, x^{k-1}) \geq 0$ and $H(x^{k-1}, x^{k-1}) = 0$ (see Proposition 3.1(c)), we obtain

$$f(x^k) \leq f(x^{k-1}) \quad \forall\, k \in N,$$

and therefore the sequence $\{f(x^k)\}_{k\in N}$ is nonincreasing.

(b) From (18), $-\mu_k^{-1}(\ln x^k - \ln x^{k-1}) \in \partial f(x^k)$. Plugging this into the formula of $\partial f(x^k)$ given by (10), we have

$$f(x) \geq f(x^k) - \mu_k^{-1}\mathrm{tr}\left[(x - x^k) \circ (\ln x^k - \ln x^{k-1})\right] \quad \forall\, x \in \mathbb{V}.$$

Consequently, for any $x \in \mathcal{K}$,

$$\begin{aligned}\mu_k \cdot (f(x^k) - f(x)) &\leq \mathrm{tr}\left[(x - x^k) \circ (\ln x^k - \ln x^{k-1})\right]\\ &= H(x, x^{k-1}) - H(x, x^k) - H(x^k, x^{k-1})\\ &\leq H(x, x^{k-1}) - H(x, x^k),\end{aligned} \tag{20}$$

where the equality holds due to Proposition 3.3(b), and the last inequality follows from the nonnegativity of $H(x, y)$.

(c) First, summing up the inequalities in part (b) over $k = 1, 2, 3, \ldots, m$ yields that

$$\sum_{k=1}^{m} \mu_k f(x^k) - \sigma_m f(x) \leq H(x, x^0) - H(x, x^m) \quad \forall\, x \in \mathcal{K}. \tag{21}$$

On the other hand, since $f(x^m) \leq f(x^k)$ by part (a), multiplying this inequality by μ_k and summing up the inequalities over $k = 1, 2, 3, \ldots, m$ then gives that

$$\sum_{k=1}^{m} \mu_k f(x^k) \geq \sum_{k=1}^{m} \mu_k f(x^m) = \sigma_m f(x^m). \tag{22}$$

Now, combining (22) with (22), we readily obtain the desired result.

(d) Since $f(x^k) - f(x) \geq 0$ for all $x \in \mathcal{X}^*$, the result immediately follows from part (b).

(e) From part (d), the sequence $\{H(x, x^k)\}_{k\in N}$ is nonincreasing for any $x \in \mathcal{X}^*$. This together with $H(x, x^k) \geq 0$ implies that $\{H(x, x^k)\}_{k\in N}$ is convergent. Thus,

$$H(x, x^{k-1}) - H(x, x^k) \to 0 \quad \forall\, x \in \mathcal{X}^*. \tag{23}$$

On the other hand, from (20) and part (d), we have

$$0 \leq \mu_k(f(x^k) - f(x)) \leq H(x, x^{k-1}) - H(x, x^k) - H(x^k, x^{k-1}), \quad \forall\, x \in \mathcal{X}^*,$$

which in turn implies

$$H(x^k, x^{k-1}) \leq H(x, x^{k-1}) - H(x, x^k).$$

Using (23) and the nonnegativity of $H(x^k, x^{k-1})$, the first result is then obtained. Since $\{x^k\}_{k\in N} \subset \{y \in \mathrm{int}(\mathcal{K}) \mid H(x, y) \leq H(x, x^0)\}$ with $x \in \mathcal{X}^*$, the sequence

$\{x^k\}_{k\in N}$ is bounded. Thus, the second result follows by the first result and Proposition 3.2(b). □

By now, we have proved that the algorithm in (11)–(12) is well-defined and satisfies some favorable properties. With these properties, we are ready to establish the convergence results of the algorithm which is one of the main purposes of this paper.

Proposition 4.3 *Let $\{x^k\}_{k\in N}$ be the sequence generated by the algorithm in (11)–(12), and let $\sigma_m := \sum_{k=1}^{m} \mu_k$. Then, under assumptions (A), the following results hold.*

(a) $f(x^m) - f(x) \le \sigma_m^{-1} H(x, x^0)$ *for all* $x \in \mathcal{K}$.
(b) *If* $\sigma_m \to +\infty$, *then* $\lim_{m\to+\infty} f(x^m) = f_*$.
(c) *If the optimal set* $\mathcal{X}^* \neq \emptyset$, *then the sequence* $\{x^k\}_{k\in N}$ *is bounded, and if, in addition,* $\sigma_m \to +\infty$, *every accumulation point is a solution of the CSCP* (1).

Proof (a) This result follows by Proposition 4.2(c) and the nonnegativity of $H(x, x^m)$.

(b) Since $\sigma_m \to +\infty$, passing the limit to the inequality in part (a), we have

$$\limsup_{m\to+\infty} f(x^m) \le f(x) \quad \forall\, x \in \mathcal{K},$$

which particularly implies

$$\limsup_{m\to+\infty} f(x^m) \le \inf\{f(x) :\ x \in \mathcal{K}\}.$$

This together with the fact that $f(x^m) \ge f_* = \inf\{f(x) : x \in \mathcal{K}\}$ yields the result.

(c) From Proposition 3.1(d), $H(x, \cdot)$ has bounded level sets for every $x \in \mathcal{K}$. Also, from Proposition 4.2(d), $\{H(x, x^k)\}_{k\in N}$ is nonincreasing for all $x \in \mathcal{X}^*$ if $\mathcal{X}^* \neq \emptyset$. Thus, the sequence $\{x^k\}_{k\in N}$ is bounded and therefore has an accumulation point. Let $\hat{x} \in \mathcal{K}$ be an accumulation point of $\{x^k\}_{k\in N}$. Then $\{x^{k_j}\} \to \hat{x}$ for some $k_j \to +\infty$. Since f is lower semi-continuous, we have $f(\hat{x}) = \liminf_{k_j\to+\infty} f(x^{k_j})$ (see [22, p. 8]). On the other hand, we also have $f(x^{k_j}) \to f_*$, hence $f(\hat{x}) = f_*$ which says that $\hat{x}$ is a solution of (1). □

Proposition 4.3(a) indicates an estimate for global rate of convergence which is similar to the one obtained for the proximal-like algorithms of convex minimization over nonnegative orthant cone. Analogously, as remarked in [6, Remark 4.1], global convergence of $\{x^k\}$ itself to a solution of (1) can be achieved under the assumption that $\{x^k\} \subset \mathrm{int}(\mathcal{K})$ has a limit point in $\mathrm{int}(\mathcal{K})$. Nonetheless, this assumption is rather stringent.

5 Exponential multiplier method for SCLP

In this section, we give a dual application of the entropy-like proximal algorithm in (11)–(12), leading to an exponential multiplier method for the symmetric cone linear program

$$\text{(SCLP)} \quad \min\left\{-b^T y: \ c - \sum_{i=1}^{m} y_i a_i \succeq_{\mathcal{K}} 0, \ y \in \mathbb{R}^m\right\},$$

where $c \in \mathbb{V}$, $a_i \in \mathbb{V}$ and $b \in \mathbb{R}^m$. For the sake of notation, we denote the feasible set of (SCLP) by $\mathcal{F} := \{y \in \mathbb{R}^m : A(y) \succeq_{\mathcal{K}} 0\}$, where $A : \mathbb{R}^m \to \mathbb{V}$ is a linear operator defined by

$$A(y) := c - \sum_{i=1}^{m} y_i a_i \quad \forall y \in \mathbb{R}^m. \tag{24}$$

In addition, we make the following assumptions for the problem (SCLP):

(A1) The optimal solution set of (SCLP) is nonempty and bounded;
(A2) Slater's condition holds, i.e., there exists a $\hat{y} \in \mathbb{R}^m$ such that $A(\hat{y}) \succ_{\mathcal{K}} 0$.

Notice that the Lagrangian function associated with (SCLP) is defined as follows:

$$L(y, x) = \begin{cases} -b^T y - \mathrm{tr}[x \circ A(y)] & \text{if } x \in \mathcal{K}, \\ +\infty & \text{otherwise.} \end{cases}$$

Therefore, the dual problem corresponding to the SCLP is given by

$$\text{(DSCLP)} \quad \max_{x \succeq_{\mathcal{K}} 0} \inf_{y \in \mathbb{R}^m} L(y, x),$$

which can be explicitly written as

$$\text{(DSCLP)} \quad \begin{cases} \max\ -\mathrm{tr}(c \circ x) \\ \text{s.t.} \quad \mathrm{tr}(a_i \circ x) = b_i, \qquad i = 1, 2, \ldots, m, \\ \qquad\quad x \succeq_{\mathcal{K}} 0. \end{cases}$$

From the standard convex analysis arguments in [19], it follows that under assumption (A2) there is no duality gap between (SCLP) and (DSCLP), and furthermore, the solution set of (DSCLP) is nonempty and compact.

To solve the problem (SCLP), we introduce the following multiplier-type algorithm: *given* $x^0 \in \mathrm{int}(\mathcal{K})$, *generate the sequence* $\{y^k\}_{k\in N} \subseteq \mathbb{R}^m$ *and* $\{x^k\}_{k\in N} \subset \mathrm{int}(\mathcal{K})$ *by*

$$y^k = \operatorname*{argmin}_{y \in \mathbb{R}^m} \left\{-b^T y + \mu_k^{-1} \mathrm{tr}\left[\exp\left(-\mu_k A(y) + \ln x^{k-1}\right)\right]\right\}, \tag{25}$$

$$x^k = \exp\left(-\mu_k A(y^k) + \ln x^{k-1}\right), \tag{26}$$

where the parameters μ_k *satisfy* $\mu_k > \bar{\mu} > 0$ *for all* $k \in N$. The algorithm can be viewed as a natural extension of the exponential multiplier method used in convex programs over nonnegative orthant cones (see, e.g., [3, 27]) to symmetric cones. In the rest of this section, we will show that the algorithm defined by (25)–(26) possesses similar properties.

We first present two technical lemmas, where Lemma 5.1 collects some properties of the Löwner operator $\exp(z)$ and Lemma 5.2 characterizes the recession cone of $\mathcal{F}$.

Lemma 5.1

(a) *For any* $z \in \mathbb{V}$, *there always holds that* $\exp(z) \in \operatorname{int}(\mathcal{K})$.
(b) *The function* $\exp(z)$ *is continuously differentiable everywhere with*

$$(\exp(z))'e = \nabla_z(\exp(z))e = \exp(z) \quad \forall\, z \in \mathbb{V}.$$

(b) *The function* $\operatorname{tr}[\exp(\sum_{i=1}^m y_i a_i)]$ *is continuously differentiable everywhere with*

$$\nabla_y \operatorname{tr}\left[\exp\left(\sum_{i=1}^m y_i a_i\right)\right] = A^T\left(\exp\left(\sum_{i=1}^m y_i a_i\right)\right),$$

where $y \in \mathbb{R}^m$ *and* $a_i \in \mathbb{V}$ *for all* $i = 1, 2, \dots, m$, *and* $A^T : \mathbb{V} \to \mathbb{R}^m$ *be a linear transformation defined by* $A^T(x) = (\langle a_1, x\rangle, \dots, \langle a_m, x\rangle)^T$ *for any* $x \in \mathbb{V}$.

Proof (a) The result is clear since for any $z \in \mathbb{V}$ with $z = \sum_{j=1}^r \lambda_j(z)c_j$, all eigenvalues of $\exp(z)$, given by $\exp(\lambda_j(z))$ for $j = 1, 2, \dots, r$, are positive.

(b) By Lemma 2.2, clearly, $\exp(z)$ is continuously differentiable everywhere and

$$\begin{aligned}(\exp(z))'h &= \sum_{j=1}^r \exp(\lambda_j(z))\langle c_j, h\rangle c_j \\ &\quad + \sum_{1\le j<l\le r} 4\frac{\exp(\lambda_j(z)) - \exp(\lambda_l(z))}{\lambda_j(z) - \lambda_l(z)} c_j \circ (c_l \circ h)\end{aligned}$$

for any $h \in \mathbb{V}$. From this formula, we particularly have

$$\begin{aligned}(\exp(z))'e &= \sum_{j=1}^r \exp(\lambda_j(z))\langle c_j, e\rangle c_j \\ &\quad + \sum_{1\le j<l\le r} 4\frac{\exp(\lambda_j(z)) - \exp(\lambda_l(z))}{\lambda_j(z) - \lambda_l(z)} c_j \circ (c_l \circ e) \\ &= \sum_{j=1}^r \exp(\lambda_j(z))c_j = \exp(z).\end{aligned}$$

(c) The first part is direct by the continuous differentiability of trace function and part (b), and the second part follows from the differential chain rule. □

Lemma 5.2 *Let* $\mathcal{F}_\infty$ *denote the recession cone of the feasible set* $\mathcal{F}$. *Then,*

$$\mathcal{F}_\infty = \left\{d \in \mathbb{R}^m : A(d) - c \succeq_{\mathcal{K}} 0\right\}.$$

Proof Assume that $d \in \mathbb{R}^m$ such that $A(d) - c \succeq_{\mathcal{K}} 0$. If $d = 0$, clearly, $d \in \mathcal{F}_\infty$. If $d \neq 0$, we take any $y \in \mathcal{F}$. From the definition of the linear operator A, for any $\tau > 0$,

$$A(y + \tau d) = c - \sum_{i=1}^{m} (y_i + \tau d_i) a_i = A(y) + \tau (A(d) - c) \succeq_{\mathcal{K}} 0.$$

This, by the definition of recession direction [19, pp. 61], shows that $d \in \mathcal{F}_\infty$. Thus, we prove that $\{d \in \mathbb{R}^m : A(d) - c \succeq_{\mathcal{K}} 0\} \subseteq \mathcal{F}_\infty$.

Now take any $d \in \mathcal{F}_\infty$ and $y \in \mathcal{F}$. Then $A(y + \tau d) \succeq_{\mathcal{K}} 0$ for any $\tau > 0$, and therefore, $\lambda_{\min}[A(y + \tau d)] \geq 0$. This must imply $\lambda_{\min}(A(d) - c) \geq 0$. If not, using the fact that

$$\begin{aligned} \lambda_{\min}[A(y + \tau d)] &= \lambda_{\min}[A(y) + \tau (A(d) - c)] \\ &\leq \lambda_{\min}(\tau (A(d) - c)) + \lambda_{\max}(A(y)) \\ &= \tau \lambda_{\min}(A(d) - c) + \lambda_{\max}(A(y)) \end{aligned}$$

where the inequality is due to Lemma 2.1, and letting $\tau \to +\infty$, we then have

$$\lambda_{\min}[A(y + \tau d)] \to -\infty,$$

contradicting the fact that $\lambda_{\min}[A(y + \tau d)] \geq 0$. Consequently, $A(d) - c \succeq_{\mathcal{K}} 0$. Together with the above discussions, we show that $\mathcal{F}_\infty = \{d \in \mathbb{R}^m \mid A(d) - c \succeq_{\mathcal{K}} 0\}$. □

Next we establish the convergence of the algorithm in (25)–(26). We first prove that the sequence generated is well-defined, which is implied by the following lemma.

Lemma 5.3 *For any $y \in \mathbb{R}^m$ and $\mu > 0$, let $F : \mathbb{R}^m \to \mathbb{R}$ be defined by*

$$F(y) := -b^T y + \mu^{-1} \mathrm{tr}\left[\exp(-\mu A(y) + \ln x^{k-1})\right]. \tag{27}$$

Then under assumption (A1) *the minimum set of F is nonempty and bounded.*

Proof We prove that F is coercive by contradiction. Suppose not, i.e., some level set of F is not bounded. Then, there exists a sequence $\{y^k\} \subseteq \mathbb{R}^m$ such that

$$\|y^k\| \to +\infty, \quad \lim_{k \to +\infty} \frac{y^k}{\|y^k\|} = d \neq 0 \quad \text{and} \quad F(y^k) \leq \delta \tag{28}$$

for some $\delta \in \mathbb{R}$. Since $\exp(-\mu A(y) + \ln x^{k-1}) \in \mathrm{int}(\mathcal{K})$ for any $y \in \mathbb{R}^m$ by Lemma 5.1(a),

$$\mathrm{tr}\left[\exp\left(-\mu A(y) + \ln x^{k-1}\right)\right] = \sum_{j=1}^{r} \exp\left[\lambda_j\left(-\mu A(y) + \ln x^{k-1}\right)\right] > 0.$$

Therefore, $F(y^k) \le \delta$ implies that the following two inequalities hold

$$-b^T y^k < \delta, \tag{29}$$

$$\sum_{j=1}^{r} \exp\left[\lambda_j\left(-\mu A(y^k) + \ln x^{k-1}\right)\right] \le \mu(\delta + b^T y^k). \tag{30}$$

Due to the nonnegativity of the exponential function, (30) is equivalent to saying that

$$\exp\left[\lambda_j\left(-\mu A(y^k) + \ln x^{k-1}\right)\right] \le \mu(\delta + b^T y^k) \quad \forall\, j = 1, 2, \ldots, r. \tag{31}$$

Dividing by $\|y^k\|$ on the both sides and using the monotonicity of $\ln t$ $(t > 0)$ then gives

$$\begin{aligned}
&\lambda_j\left(-\mu A(y^k) + \ln x^{k-1}\right) - \ln(\|y^k\|) \\
&\quad \le \ln\left[\mu(\delta + b^T y^k)/\|y^k\|\right] \\
&\quad \le \frac{\mu(\delta + b^T y^k)}{\|y^k\|} - 1 \quad \forall\, j = 1, 2, \ldots, r.
\end{aligned}$$

where the last inequality is due to $\ln t \le t - 1$ $(t > 0)$. Now dividing $\|y^k\|$ on the both sides again and using the homogeneity of the function $\lambda_j(\cdot)$ yields

$$\lambda_j\left(-\frac{\mu A(y^k)}{\|y^k\|} + \frac{\ln x^{k-1}}{\|y^k\|}\right) - \frac{\ln(\|y^k\|)}{\|y^k\|} \le \frac{\mu(\delta + b^T y^k)}{\|y^k\|^2} - \frac{1}{\|y^k\|} \quad \forall\, j = 1, 2, \ldots, r.$$

Passing to the limit $k \to +\infty$ on the both sides and noting that $\|y^k\| \to +\infty$, we get

$$\lambda_j\left(\mu \sum_{i=1}^{m} d_i a_i\right) \le 0 \quad \forall\, j = 1, 2, \ldots, r,$$

which, by the homogeneity of $\lambda_j(\cdot)$ again, implies

$$\mu\lambda_j\left(A(d) - c\right) = \mu\lambda_j\left(-\sum_{i=1}^{m} d_i a_i\right) \ge 0 \quad \forall\, j = 1, 2, \ldots, r.$$

Consequently, $A(d) - c \succeq_{\mathcal{K}} 0$. From Lemma 5.2, $\mathcal{F}_\infty = \{d \in \mathbb{R}^m : A(d) - c \succeq_{\mathcal{K}} 0\}$. This together with $-b^T d \le 0$ shows that there exists a nonzero $d \in \mathbb{R}^m$ such that $d \in \mathcal{F}_\infty$ but $-b^T d \le 0$, obviously contradicting assumption (A1). Thus, we complete the proof. □

To analyze the convergence of the algorithm defined by (25)–(26), we also need the following lemma which states that the sequence $\{x^k\}_{k\in N}$ generated by (25)–(26) is exactly the one given by the algorithm in (11)–(12) when applied to the dual problem (DSCLP).

Lemma 5.4 *The sequence* $\{x^k\}_{k\in N}$ *generated by the multiplier method* (25)–(26) *can be obtained via the following iterate scheme*

$$x^k = \operatorname*{argmax}_{x\succeq_{\mathcal{K}}0} \left\{ D(x) - \mu_k^{-1} H(x, x^{k-1}) \right\}, \tag{32}$$

where $D(x) := \inf_{y\in\mathbb{R}^m} L(y,x)$ *is the dual objective function of* (DSCLP).

Proof First, we prove that $-A(y^k) \in \partial D(x^k)$. Using Lemma 5.1(c) and the optimality condition of (25), we obtain that

$$\begin{aligned} 0 &= -b_i + \left\langle a_i,\ \exp(-\mu_k A(y^k) + \ln(x^{k-1})) \right\rangle \\ &= -b_i + \left\langle a_i, x^k \right\rangle \\ &= -b_i + \operatorname{tr}(a_i \circ x^k), \quad i = 1, 2, \ldots, m, \end{aligned}$$

where the second equality is due to (26). This implies that y^k is also minimizing the Lagrangian $L(y, x^k)$, and consequently $D(x^k) = L(y^k, x^k)$. Now, we have that

$$\begin{aligned} D(x) &= \inf_{y\in\mathbb{R}^m} \left\{ -b^T y - \operatorname{tr}[x \circ A(y)] \right\} \\ &\le -b^T y^k - \operatorname{tr}[x \circ A(y^k)] \\ &= -b^T y^k - \operatorname{tr}[x^k \circ A(y^k)] - \operatorname{tr}[(x - x^k) \circ A(y^k)] \\ &= D(x^k) + \langle x - x^k,\ -A(y^k)\rangle. \end{aligned} \tag{33}$$

In view of the concavity of $D(x)$, the inequality (33) means that $-A(y^k) \in \partial D(x^k)$. Using formula (26), we then have $\mu_k^{-1}(\ln x^k - \ln x^{k-1}) \in \partial D(x^k)$. From Proposition 3.1, this is precisely the optimality condition of the maximum problem in (32). □

Now we are in a position to present the convergence results of the algorithm defined by (25)–(26). Their proofs are similar to those of [6, Theorem 5.1], and we here include them for completeness.

Proposition 5.1 *Let* $\{y^k\}_{k\in N}$ *and* $\{x^k\}_{k\in N}$ *be the sequences generated by* (25)–(26). *Then, under assumptions* (A1) *and* (A2), *the following results hold.*

(a) *The sequence* $\{x^k\}_{k\in N} \subset \operatorname{int}(\mathcal{K})$ *is bounded and each limit point is a dual solution.*
(b) $\operatorname{tr}[x^k \circ A(y^k)] \to 0$ *when* $k \to +\infty$.
(c) *Let* $\tilde{y}^k = \sum_{l=1}^k \eta_l y^l$ *with* $\eta_l := \mu_l/\nu_k > 0$ *and* $\nu_k := \sum_{l=1}^k \mu_l$. *Then*

$$\liminf_{k\to+\infty} \lambda_{\min}(A(\tilde{y}^k)) \ge 0.$$

(d) *Let* D^* *be the optimal value of* (DSCLP). *Then* $-b^T y^k \to D^*$ *and* $-b^T \tilde{y}^k \to D^*$.
(e) $\{\tilde{y}^k\}$ *is bounded and its every limit point is a solution of* (SCLP).

(f) $\lim_{k\to+\infty} -b^T y^k = \lim_{k\to+\infty} D(x^k) = -b^T y^*$, *where* y^* *is a solution of* (SCLP).

Proof (a) From Lemma 5.4, $\{x^k\}_{k\in N}$ is the sequence generated by applying the entropy-like proximal algorithm (11)–(12) to (DSCLP). Since under assumption (A2) the solution set of (DSCLP) is nonempty and compact, the result follows from Proposition 4.3 directly.

(b) Using the definition of H and noting that $-\mu_k A(y^k) = \ln x^k - \ln x^{k-1}$, we have

$$\begin{aligned} H(x^k, x^{k-1}) &= \mathrm{tr}[x^k \circ (\ln x^k - \ln x^{k-1}) + x^{k-1} - x^k] \\ &= -\mu_k \mathrm{tr}[x^k \circ A(y^k)] + \mathrm{tr}(x^{k-1} - x^k). \end{aligned}$$

From Proposition 4.2(e), we know that $H(x^k, x^{k-1}) \to 0$ and $\mathrm{tr}(x^{k-1} - x^k) \to 0$. Thus, by noting that $\mu_k > \bar{\mu} > 0$, the last equality implies $\mathrm{tr}[x^k \circ A(y^k)] \to 0$.

(c) From the linearity of $A(y)$ and the definition of $\tilde{y}^k$, we have that

$$\begin{aligned} A(\tilde{y}^k) = \sum_{l=1}^{k} \eta_l A(y^l) &= \sum_{l=1}^{k} \frac{\eta_l}{\mu_l} \Big[\ln x^{l-1} - \ln x^l\Big] \\ &= \nu_k^{-1} \sum_{l=1}^{k} \Big[\ln x^{l-1} - \ln x^l\Big] \\ &= \nu_k^{-1} (\ln x^0 - \ln x^k), \end{aligned}$$

where the second equality is due to (26). From Lemma 2.1, it then follows that

$$\lambda_{\min}(A(\tilde{y}^k)) = \lambda_{\min}\left(\frac{\ln x^0 - \ln x^k}{\nu_k}\right) \geq \frac{\lambda_{\min}(\ln x^0)}{\nu_k} + \frac{\lambda_{\min}(-\ln x^k)}{\nu_k}.$$

Since, as $\nu_k \to +\infty$, the first term of the right hand side tends to zero, it remains to prove that $\liminf_{k\to+\infty} \lambda_{\min}(-\ln x^k)/\nu_k \geq 0$. Notice that

$$\begin{aligned} \liminf_{k\to+\infty} \nu_k^{-1} \lambda_{\min}(-\ln x^k) &= -\limsup_{k\to+\infty} \nu_k^{-1} \lambda_{\max}(\ln x^k) \\ &= -\limsup_{k\to+\infty} \nu_k^{-1} \ln(\lambda_{\max}(x^k)). \end{aligned} \tag{34}$$

In addition, since $\{x^k\}_{k\in N} \subset \mathrm{int}(\mathcal{K})$ is bounded, we have $\lambda_{\max}(x^k) \leq \bar{\lambda}$ for some $\bar{\lambda} > 0$. This implies that

$$-\limsup_{k\to+\infty} \nu_k^{-1} \ln(\lambda_{\max}(x^k)) \geq -\limsup_{k\to+\infty} \nu_k^{-1} \ln \bar{\lambda} \geq 0.$$

Combining with (34) then yields the desired result.

(d) Since $\{x^k\}_{k\in N}$ is a feasible sequence of (DSCLP), there holds that

$$\begin{aligned}\operatorname{tr}[x^k \circ A(y^k)] &= \operatorname{tr}\left[x^k \circ c - \sum_{i=1}^{m} y_i^k x^k \circ a_i\right]\\ &= \operatorname{tr}[x^k \circ c] - b^T y^k\\ &= -b^T y^k - D(x^k).\end{aligned}$$

Noting that $\operatorname{tr}[x^k \circ A(y^k)] \to 0$ and $D(x^k) \to h^*$ by Proposition 4.3(b), we readily obtain the result from the last equation.

(e) Suppose that $\{\tilde{y}^k\}$ is unbounded. Let $\hat{y}^*$ be the element with the maximum norm from the solution set of (SCLP). The existence of $\hat{y}^*$ is guaranteed by the boundedness of the solution set of (SCLP). Define

$$\alpha_k = 1 - \frac{4\|\hat{y}^*\|}{\|\tilde{y}^k - \hat{y}^*\|}.$$

Since $\|\tilde{y}^k\| \to +\infty$, there must exist an k_0 such that $0 < \alpha_k < 1$ for all $k \geq k_0$. Let $z^k = \alpha_k \hat{y}^* + (1-\alpha_k)\tilde{y}^k$. It is easy to verify that

$$3\|\hat{y}^*\| \leq \|z^k\| \leq 9\|\hat{y}^*\|.$$

This means that the sequence $\{z^k\}$ is bounded. We next prove that each limit point of $\{z^k\}$ is an optimal solution to (SCLP), which together with the last inequality contradicts the fact that $\hat{y}^*$ is an element of the maximum norm in the solution set of (SCLP). Let z^* be a limit point of $\{z^k\}$. Without loss of generality, we assume that $z^k \to z^*$. Noting that $A(z^k) = \alpha_k A(\hat{y}^*) + (1-\alpha_k)A(\tilde{y}^k)$, $\alpha_k \to 1$ and $\liminf_{k\to+\infty} \lambda_{\min}(A(\tilde{y}^k)) \geq 0$, we have $A(z^*) \succeq_{\mathcal{K}} 0$, i.e., z^* is a feasible point of (SCLP), which in turn means that $b^T\hat{y}^* \geq b^T z^*$. On the other hand, since $-b^T\tilde{y}^k \to D^* \leq -b^T\hat{y}^*$ by part (d) and the weak duality, we get

$$b^T z^* = \lim_{k\to+\infty} b^T z^k = \lim_{k\to+\infty} [\alpha_k b^T \hat{y}^* + (1-\alpha_k) b^T \tilde{y}^k] \geq b^T \hat{y}^*.$$

Thus, we have $b^T z^* = b^T \hat{y}^*$, and consequently z^* is an optimal solution of (SCLP).

Let $\tilde{y}^*$ be a limit point of $\{\tilde{y}^k\}$. Since $\liminf_{k\to+\infty} \lambda_{\min}(A(\tilde{y}^k)) \geq 0$ by part (d), $\tilde{y}^*$ be a feasible solution of (SCLP). Therefore, $-b^T\tilde{y}^* \geq -b^T y^*$, where y^* be a solution of (SCLP) (its existence is guaranteed by assumption (A1)). On the other hand, from part (d) and the weak duality, it follows that $-b^T\tilde{y}^* = D^* \leq -b^T y^*$. The two sides show that

$$-b^T\tilde{y}^* = -b^T y^* = D^*. \tag{35}$$

Consequently, $\tilde{y}^*$ is an optimal solution of (SCLP).

(f) The first equality is due to part (d) and the second follows from (35). □

Observe that the above convergence properties of the algorithm (25)–(26) are similar to the ones obtained by [27] for convex programs over nonnegative orthant cones, except that the global convergence of the dual sequence to an optimal dual solution is not guaranteed. The main reason is that under the setting of symmetric cones, when $\mathrm{int}(\mathcal{K}) \supset \{x^k\} \to \bar{x}^* \in \mathcal{K}$, $H(x^k, \bar{x}^*) \to 0$ does not hold. (A counterexample can be found for the semidefinite program in [6].) However, one still has convergence in terms of function values, and moreover, by applying Proposition 4.3(a) with $x = x^*$, where x^* is a solution of (DSCLP), one has the global convergence rate estimate:

$$\mathrm{tr}(c \circ (x^* - x^k)) \leq \left(\sum_{l=1}^{k} \mu_l\right)^{-1} H(x^*, x^0).$$

6 Conclusions

We have developed an entropy-like proximal algorithm for the CSCP (1). The algorithm is based on the distance-like function $H(\cdot,\cdot)$ defined on the symmetric cone $\mathcal{K}$ of the Euclidean Jordan algebra. We showed that the proposed algorithm is well-defined and established its convergence properties. Also, we presented a dual application of the algorithm to the symmetric cone linear programming problem (SCLP), leading to a multiplier method for this class of symmetric cone optimization problems. The multiplier method was shown to share many similar properties with the exponential multiplier method developed by [27] for convex minimization with non-negative orthant cone constraints.

Acknowledgements The authors would like to thank the editor and two anonymous referees for their helpful comments and suggestions on the revision of this paper.

References

1. Alizadeh, F., Goldfarb, D.: Second-order cone programming. Math. Program. **95**, 3–51 (2003)
2. Baes, M.: Convexity and differentiability properties of spectral functions and spectral mappings on Euclidean Jordan algebras. Linear Algebra Appl. **422**, 664–700 (2007)
3. Bertsekas, D.: Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York (1982)
4. Chen, G., Teboulle, M.: Convergence analysis of a proximal-like minimization algorithm using Bregman functions. SIAM J. Optim. **3**, 538–543 (1993)
5. Censor, Y., Zenios, S.A.: The proximal minimization algorithm with D-functions. J. Optim. Theory Appl. **73**, 451–464 (1992)
6. Doljansky, M., Teboulle, M.: An interior proximal algorithm and the exponential multiplier method for semidefinite programming. SIAM J. Optim. **9**, 1–13 (1998)
7. Eckstein, J.: Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. Math. Oper. Res. **18**, 206–226 (1993)
8. Eggermont, P.B.: Multilicative iterative algorithms for convex programming. Linear Algebra Appl. **130**, 25–42 (1990)
9. Faybusovich, L.: Euclidean Jordan algebras and interior point algorithm. Positivity **1**, 331–357 (1997)
10. Faybusovich, L.: Linear systems in Jordan algebras and primal-dual interior point algorithm. J. Comput. Appl. Math. **86**, 149–175 (1997)
11. Faraut, J., Korányi, A.: Analysis on Symmetric Cones. Oxford Mathematical Monographs. Oxford University Press, New York (1994)
12. Guler, O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**, 403–419 (1991)

13. Iusem, A., Teboulle, M.: Convergence rate analysis of nonquadratic proximal method for convex and linear programming. Math. Oper. Res. **20**, 657–677 (1995)
14. Iusem, A., Svaiter, B., Teboulle, M.: Entropy-like proximal methods in convex programming. Math. Oper. Res. **19**, 790–814 (1994)
15. Koecher, M.: In: Brieg, A., Walcher, S. (eds.) The Minnesota Notes on Jordan Algebras and Their Applications. Springer, Berlin (1999)
16. Lobo, M.S., Vandenberghe, L., Boyd, S., Lebret, H.: Application of second-order cone programming. Linear Algebra Appl. **284**, 193–228 (1998)
17. Martinet, B.: Perturbation des methodes d'optimisation. Appl. RAIRO Numer. Anal. (1978)
18. Moreau, J.J.: Promimité et dualité dans un espace hilbertien. Bull. Soc. Math. Fr. **93**, 273–299 (1965)
19. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
20. Rockafellar, R.T.: Augmented lagrangians and applications of proximal point algorithm in convex programming. Math. Oper. Res. **1**, 97–116 (1976)
21. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
22. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
23. Schmieta, S.H., Alizadeh, F.: Extension of primal-dual interior point algorithms to symmetric cones. Math. Program. **96**, 409–438 (2003)
24. Sun, D., Sun, J.: Löwner's operator and spectral functions on Euclidean Jordan algebras, Math. Oper. Res. (2004, submitted)
25. Teboulle, M.: Entropic proximal mappings with applications to nonlinear programming. Math. Oper. Res. **17**, 670–690 (1992)
26. Teboulle, M.: Convergence of proximal-like algorithms. SIAM J. Optim. **7**, 1069–1083 (1997)
27. Tseng, P., Bertsekas, D.: On the convergence of the exponential multiplier method for convex programming. Math. Program. **60**, 1–19 (1993)
28. Vandenberghe, L., Boyd, S.: A primal-dual potential reduction method for problems involving matrix inequalities. Math. Program. **69**, 205–236 (1995)