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

Jein-Shan Chen [1]
Department of Mathematics
National Taiwan Normal University
Taipei 11677, Taiwan

Shaohua Pan [2]
School of Mathematical Sciences
South China University of Technology
Guangzhou 510640, China

April 28, 2006
(revised May 16, 2007)

Abstract. We introduce an entropy-like proximal algorithm for the problem of minimizing a closed proper convex function subject to the symmetric cone constraint. 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 algorithm for convex programming with nonnegative orthant cone constraint, 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 enjoy properties similar to the exponential multiplier method in [29].

Key words. Symmetric cone optimization, proximal-like method, entropy-like distance, exponential multiplier method.

1 Introduction

Symmetric cone programming (SCP) provides a unified framework for various mathematical programming including the linear programming (LP), the second-order cone programming (SOCP) and the semidefinite programming (SDP). Such programming problems arise

---

[1] Member of Mathematics Division, National Center for Theoretical Sciences, Taipei Office. The author's work is partially supported by National Science Council of Taiwan. E-mail: jschen@math.ntnu.edu.tw.

[2] E-mail:shhpan@scut.edu.cn.

naturally in a wide range of applications in engineering, economics, optimal control, management science, combinatorial optimization, and other fields; see [1, 16, 30] and references therein. Recently, the symmetric cone optimization problem, especially the symmetric cone linear programming (SCLP), has recently attracted the attention of some researchers with a focus on the development of interior point methods similar to those used in linear programming; see [9, 10, 25]. Although interior-point methods were successfully applied for solving the SCLP, it is worthwhile to explore other solution methods for more 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 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 $\mathbb{V}$ (will be introduced in Sec. 2). In this paper, we consider the following convex symmetric cone programming (CSCP):

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

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 $x - y \in \mathrm{int}(\mathcal{K})$. A function is closed if and only if it is lower semi-continuous (l.s.c. for short) and a function is proper if $f(x) < \infty$ for at least one $x \in \mathbb{V}$ and $f(x) > -\infty$ for all $x \in \mathbb{V}$.

Notice that the CSCP is indeed a special class of convex programs, and therefore it can in principle 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 $\{x^k\}_{k\in N}$ by the iterative scheme as below:

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

where μ_k is a sequence of positive numbers and $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$. This method was originally introduced by Martinet [17], based on the Moreau proximal approximation of f (see [18]), and then further developed and studied by Rockafellar [22, 23]. Later, several generalizations of the proximal algorithm have been considered where the usual quadratic proximal term in (2) is replaced by nonquadratic distance-like functions; see, e.g. [5, 7, 8, 14, 27]. Among others, the proximal algorithms using an entropy-like distance [27, 14, 13, 28], designed for minimizing a convex function $f(x)$ subject to nonnegative orthant constraints $x \in \mathbb{R}^n_+$, generate the iterates by

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

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, 27, 28]. An important choice of φ is the case that $\varphi(t) = t \ln t - t + 1$ 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 well known Kullback-Leibler entropy from statistics and that is the "entropy" terminology stems from. One key feature of entropic proximal methods is that they automatically generate a sequence staying the interior of $\mathbb{R}^n_+$, i.e. $\{x^k\}_{k \in N} \subset \mathbb{R}^n_{++}$, and hence eliminates the combinatorial nature of the problem. One of the main applications of these proximal methods is to the dual of smooth convex programs, yielding twice continuously differentiable nonquadratic augmented Lagrangians and thus allowing the use of Newton's method.

The main motivation of this paper is to develop unified proximal-type algorithms and their corresponding dual augmented Lagrangian methods for the convex symmetric cone optimization problems. Specifically, by using the Euclidean Jordan algebraic techniques, we introduce an interior proximal-type algorithm for the CSCP which can be viewed as an extension of the entropy-like proximal algorithm defined by (3)-(4) with $\varphi(t) = t \ln t - t + 1$. For the proposed algorithm, we establish its convergence properties, and also present a dual application to the SCLP, leading to an exponential multiplier method which is shown to possess properties analogous to the method proposed by [3, 29] for convex programming over nonnegative orthant cones.

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 Section 3, we introduce a distance-like function H to measure how close between two points in the symmetric cone $\mathcal{K}$ and show some related properties. In addition, we also outline a basic proximal-like algorithm with the measure function H. The convergence analysis of the basic algorithm is studied in Section 4. In Section 5, we consider a dual application of the algorithm to the SCLP and establish the convergence results for the resulted multiplier method. We close with some final remarks in Section 6.

2 Preliminaries on Euclidean Jordan Algebra

In this section, we recall 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 the monograph by Faraut and Korányi [11].

Let $\mathbb{V}$ be a finite-dimensional vector space endowed with a bilinear mapping $(x, y) \to x \circ y$ from $\mathbb{V} \times \mathbb{V}$ into $\mathbb{V}$. 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 usual matrix multiplication of X and Y and $\text{Tr}(XY)$ is the trace of matrix XY. Then, $(\mathbb{S}^n, \circ, \langle \cdot, \cdot \rangle_{\mathbb{S}^n})$ is a Euclidean Jordan algebra with the Jordan product given 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 (denoted as $\mathbb{S}^n_+$) 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 $e = (1, 0)^T$ and the set of squares $\mathcal{K} = \{x = (x_1, x_2) \in$

$\mathbb{R} \times \mathbb{R}^{n-1} : \|x_2\| \leq x_1\}$.

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 $\{c_1, c_2, \cdots, c_k\}$ is a complete system of orthogonal idempotents if

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

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 second version of the 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 $\{c_1, c_2, \cdots, c_r\}$ and real numbers $\lambda_1(x), \lambda_2(x), \cdots, \lambda_r(x)$, 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*. We let $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the smallest and the largest eigenvalues, respectively. The *trace* of x, denoted as $\mathrm{tr}(x)$, is given by $\mathrm{tr}(x) := \sum_{j=1}^r \lambda_j(x)$.

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

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

Now we can present two important inequalities of the eigenvalue function $\lambda_{\min}(\cdot)$.

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

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

Proof. The first inequality can be found in [25, Lemma 14]. Thus, we only prove the second inequality. From [25, Lemma 13], we know

$$\lambda_{\min}(x) = \min_{u \in \mathbb{V}} \frac{\langle u, x \circ u \rangle}{\langle u, u \rangle} \quad \text{and} \quad \lambda_{\max}(x) = \max_{u \in \mathbb{V}} \frac{\langle u, x \circ u \rangle}{\langle u, u \rangle}.$$

Hence, there holds

$$\begin{aligned}
\lambda_{\min}(x+y) &= \min_{u \in \mathbb{V}} \frac{\langle u, (x+y) \circ u \rangle}{\langle u, u \rangle} = \min_{u \in \mathbb{V}} \frac{\langle u, x \circ u \rangle + \langle u, y \circ u \rangle}{\langle u, u \rangle} \\
&\leq \min_{u \in \mathbb{V}} \left\{ \frac{\langle u, x \circ u \rangle}{\langle u, u \rangle} + \max_{u \in \mathbb{V}} \frac{\langle u, y \circ u \rangle}{\langle u, u \rangle} \right\} \\
&= \lambda_{\min}(x) + \lambda_{\max}(y).
\end{aligned}$$

Consequently, the second inequality follows. $\square$

Let $\phi : \mathbb{R} \to \mathbb{R}$ 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

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

where $x \in \mathbb{V}$ has the spectral decomposition $x = \sum_{j=1}^{r} \lambda_j(x) c_j$. This function ϕ^{sc} is called the Löwner operator and shown to enjoy the following property [26].

Lemma 2.2 *[26, Theorem 13] For any $x = \sum_{j=1}^{r} \lambda_j(x) c_j$, let ϕ^{sc} be defined as in (6). Then ϕ^{sc} is (continuously) differentiable at x if and only if for each $j \in \{1, 2, \cdots, r\}$, ϕ is (continuously) differentiable at $\lambda_j(x)$. The derivative of ϕ^{sc} at x, for any $h \in \mathbb{V}$, is given by*

$$(\phi^{\mathrm{sc}})'(x)h = \sum_{j=1}^{r} [\phi^{[1]}(\lambda(x))]_{jj} \langle c_j, h \rangle c_j + \sum_{1 \leq j < l \leq r} 4 [\phi^{[1]}(\lambda(x))]_{jl} c_j \circ (c_l \circ h)$$

with

$$[\phi^{[1]}(\lambda(x))]_{ij} := \begin{cases} \dfrac{\phi(\lambda_i(x)) - \phi(\lambda_j(x))}{\lambda_i(x) - \lambda_j(x)} & \text{if } \lambda_i(x) \neq \lambda_j(x) \\ \phi'(\lambda_i(x)) & \text{if } \lambda_i(x) = \lambda_j(x) \end{cases}, \quad i, j = 1, 2, \cdots, r.$$

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

$$(\phi^{\mathrm{sc}})'(x) = \sum_{j=1}^{r} \varphi'(\lambda_j(x)) \mathcal{Q}(c_j) + 2 \sum_{i,j=1, i \neq j}^{r} [\phi^{[1]}(\lambda(x))]_{ij} \mathcal{L}(c_j) \mathcal{L}(c_i), \tag{7}$$

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 $\mathbb{V}$.

We next introduce the real-valued spectral functions on the Euclidean Jordan algebra. Let $\mathbb{P}$ be 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 $\mathbb{P}$ (we adopt the same notations used in [2])).

Definition 2.1 *Let $S \subseteq \mathbb{R}^r$ be a symmetric set. A real-valued function $f : S \to \mathbb{R}$ is symmetric if for every permutation $P \in \mathbb{P}$ and each $s \in S$, we have $f(Ps) = f(s)$.*

For any $x \in \mathbb{V}$ with the spectral decomposition $x = \sum_{j=1}^{r} \lambda_j(x) c_j$, we define

$$K := \{x \in \mathbb{V} \mid \lambda(x) \in S\},$$

where $\lambda(x) = (\lambda_1(x), \cdots, \lambda_r(x))^T$ is the spectral vector of x and $S \subseteq \mathbb{R}^r$ is a symmetric set. Let $f : S \to \mathbb{R}$ is a symmetric function. Then the function $F : K \to \mathbb{R}$ given by

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

is the *spectral function generated by* f. Moreover, from [2, Theorem 41], we know that F is (strictly) convex if f is (strictly) convex.

Unless otherwise stated, in the sequel, $\mathbb{A} = (\mathbb{V}, \circ, \langle \cdot, \cdot \rangle)$ 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 the domain of f by $\mathrm{dom}(f) := \{x \in \mathbb{V} \mid f(x) < +\infty\}$. The subdifferential of f at $x_0 \in \mathbb{V}$ is the convex set

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

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

$$\partial f(x_0) = \{\xi \in \mathbb{V} \mid f(x) \geq f(x_0) + \mathrm{tr}(\xi \circ (x - x_0)) \quad \forall\, x \in \mathbb{V}\}. \tag{10}$$

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

3 An entropy-like proximal algorithm for SCP

To solve the convex symmetric optimization problem CSCP, we in this section suggest an entropy-like proximal minimization algorithm defined as follows:

$$\left\{ \begin{array}{lcl} x^0 & \succ_{\mathcal{K}} & 0 \\ x^k & = & \displaystyle\operatorname*{argmin}_{x \,\succeq_{\mathcal{K}} 0} \left\{ f(x) + \frac{1}{\mu_k} H(x, x^{k-1}) \right\}, \end{array} \right. \tag{11}$$

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

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

This algorithm is indeed a proximal-type one, except that, instead of using the classical quadratic term $\|x - x^{k-1}\|^2$, we adopt the distance-like function H to guarantee that the generated sequence $\{x^k\}_{k \in N}$ satisfies $x^k \in \mathrm{int}(\mathcal{K})$ for all k, thus leading to an "interior" proximal method (see Proposition 4.1).

By the definition of Löwner operator in Section 2, it is clear that the function $H(x,y)$ is well-defined for all $x, y \in \operatorname{int}(\mathcal{K})$. Moreover, the domain of $x \in \operatorname{int}(\mathcal{K})$ can be extended to $x \in \mathcal{K}$ by adopting the convention that $0 \ln 0 \equiv 0$. The function H is a natural extension of the distance-like entropy function given as 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 given as 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 $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ being the standard dot product $\langle a, b \rangle_{\mathbb{R}^n} = \sum_{i=1}^n a_i b_i$ for every $a, b \in \mathbb{R}^n$. As shown by Proposition 3.1 below, most of the important properties, but not all, hold for $d_\varphi(\cdot, \cdot)$ also hold for $H(\cdot, \cdot)$.

In what follows, we present several technical lemmas that will be used in investigating some favorable properties of the distance measure H. We start with the extension to Euclidean Jordan algebras of *Von Neumann inequality*, whose proof can be found in [2].

Lemma 3.1 *For any $x, y \in \mathbb{V}$, we have $\operatorname{tr}(x \circ y) \leq \sum_{j=1}^r \lambda_j(x)\lambda_j(y) = \lambda(x)^T \lambda(y)$, 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 $\Phi(x) := \operatorname{tr}(x \circ \ln x)$. Then, we have the following results.*

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

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

(b) *$\Phi(x)$ is continuously differentiable on $x \in \operatorname{int}(\mathcal{K})$ with $\nabla \Phi(x) = \ln x + e$.*

(c) *The function $\Phi(x)$ is strictly convex over $x \in \mathcal{K}^n$.*

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

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

Then, from Section 2, we know that the vector-valued function $x \circ \ln x$ is exactly the Löwner function $\phi^{\mathrm{sc}}(x)$, i.e., $\phi^{\mathrm{sc}}(x) = x \circ \ln x$. Clearly, ϕ^{sc} is well-defined for any $x \in \mathcal{K}$ and

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

Therefore

$$\Phi(x) = \operatorname{tr}(\phi^{\mathrm{sc}}(x)) = \sum_{j=1}^r \lambda_j(x) \ln(\lambda_j(x)) = \widehat{\Phi}(\lambda(x)),$$

where $\widehat{\Phi} : \mathbb{R}^r_+ \to \mathbb{R}$ is given by (13). By Definition 2.1, clearly, the function $\widehat{\Phi}$ is symmetric. Consequently, $\Phi(x)$ is the spectral function generated by the symmetric function $\widehat{\Phi}$ in view of (8).

(b) From Lemma 2.2, we know that $\phi^{\rm sc}(x) = x \circ \ln x$ is continuously differentiable on $x \in \mathrm{int}(\mathcal{K})$. Thus, $\Phi(x)$ is also continuously differentiable on $x \in \mathrm{int}(\mathcal{K})$ because Φ is the composition of the trace function (clearly continuously differentiable) and $\phi^{\rm sc}$. 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 the 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, it is easy to compute that

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

This together with equation (14) yields the desired results.

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

Now we are in a position to summarize some of favorable properties of the distance-like function H. These properties play an important role in the convergence analysis (as will be seen in the next section) of the algorithm (11)-(12).

Proposition 3.1 *Let $H(x, y)$ be given as in (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) *$H(x, y) \ge 0$ for any $x \in \mathcal{K}$ and $y \in \mathrm{int}(\mathcal{K})$; moreover, $H(x, y) = 0$ if and only if $x = y$.*

(d) *$H(x,y) \geq d(\lambda(x), \lambda(y))$ 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) *The level sets of $H(\cdot, y)$ and $H(x, \cdot)$ are respectively bounded for any fixed $y \in \mathrm{int}(\mathcal{K})$ and $x \in \mathcal{K}$.*

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, it gives that H is continuous over $\mathcal{K} \times \mathrm{int}(\mathcal{K})$. On the other hand, from definition of H, it follows that

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

Notice that $\Phi(x)$ is strictly convex over $\mathcal{K}$ by Lemma 3.2 (c) and the other terms in the above expression 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 as (16) and Lemma 3.2 (b), clearly, 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),

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

for any $x \in \mathcal{K}$ and $y \in \mathrm{int}(\mathcal{K})$. On the other hand, the strict convexity of Φ shown in Lemma 3.2 (c) 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) First, from Lemma 3.1, it follows that

$$\mathrm{tr}(x \circ \ln y) \leq \sum_{j=1}^{r} \lambda_j(x) \lambda_j(\ln y) = \sum_{j=1}^{r} \lambda_i(x) \ln(\lambda_j(y)), \quad \forall\, x \in \mathbb{V}, y \in \mathrm{int}(\mathcal{K}).$$

Applying this inequality, we thus obtain for all $x \in \mathcal{K}$ and $y \in \mathrm{int}(\mathcal{K})$,

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

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

(e) For a fixed $y \in \mathrm{int}(\mathcal{K})$, let $L_\gamma(x) := \{x \in \mathcal{K} \mid H(x, y) \leq \gamma\}$ where $\gamma \geq 0$. By part (c), we have $L_\gamma(x) \subseteq \{x \in \mathcal{K} \mid d(\lambda(x), \lambda(y)) \leq \gamma\}$. Since $d(\cdot, v)$ has bounded level sets for any $v \in \mathbb{R}^r_{++}$ (see [13, Proposition 4]) and $\lambda(\cdot)$ is continuous, $L_\gamma(x)$ are bounded for all $\gamma \geq 0$. Similarly, $H(x, \cdot)$ has bounded level sets for any $x \in \mathcal{K}$. $\square$

Proposition 3.2 *Let $H(x, y)$ be given as in (12). Suppose $\{x^k\}_{k \in N} \subseteq \mathcal{K}$ and $\{y^k\}_{k \in N} \subset \mathrm{int}(\mathcal{K})$ 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* $j = 1, 2, \cdots, r$.

(b) $\mathrm{tr}(x^k - y^k) \to 0$.

Proof. (a) By Proposition 3.1 (d) and the nonnegativity of $d(\cdot, \cdot)$,

$$H(x^k, y^k) \geq d(\lambda(x^k), \lambda(y^k)) \geq 0.$$

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. Therefore, $d(\lambda(x^k), \lambda(y^k)) \to 0$ implies that

$$\lambda_j(y^k) \varphi\Big(\lambda_j(x^k)/\lambda_j(y^k)\Big) \to 0, \quad j = 1, 2, \cdots, r,$$

which is also equivalent to

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

Since $\{\lambda_j(x^k)\}$ and $\{\lambda_j(y^k)\}$ are bounded, using [6, Lemma A.1] yields that $\lambda_j(x^k) - \lambda_j(y^k) \to 0$ for all $j = 1, 2, \cdots, r$.

(b) Since $\mathrm{tr}(x^k - y^k) = \sum_{j=1}^{r} (\lambda_j(x^k) - \lambda_j(y^k))$, the desired results follows by part (a). $\square$

Finally, we present two useful relations for the function H, which can be easily verified by direct substitution, using the definition of H given as in (12), and recalling the nonnegativity of H (see Proposition 3.1 (c)).

Proposition 3.3 *Let $H(x,y)$ be given in (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) $\mathrm{tr}\Big((z-y) \circ (\ln y - \ln x)\Big) = H(z,x) - H(z,y) - H(y,x) \leq H(z,x) - H(z,y)$.

4 Convergence analysis of the algorithm

The analysis of entropy-like proximal algorithm for the CSCP defined by (11)-(12) is similar to that developed for convex minimization problem by Güler [12] for quadratical proximal methods and its extension derived by Chen and Teboulle [4] for the proximal-like algorithm using Bregman functions. In what follows, we present a global convergence rate estimate for the entropy-like proximal algorithm (11)-(12) in terms of function values.

We make the following assumptions for the problem CSCP:

(A1) $\inf\{f(x) \mid x \succeq_{\mathcal{K}} 0\} := f_* > -\infty$,

(A2) $\mathrm{dom}(f) := \{x \in \mathbb{V} \mid f(x) < \infty\} \bigcap \mathrm{int}(\mathcal{K}) \neq \emptyset$.

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

Before proving the convergence results, we have to show that the algorithm given in (11)-(12) is well-defined, i.e., it generates a sequence $\{x^k\}_{k \in N} \subset \mathrm{int}(\mathcal{K})$. A technical lemma as below is needed for the proof.

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. It is clear that $\mathrm{tr}(x \circ y) = 0$ implies $\langle x, y \rangle = 0$. From the self-duality of $\mathcal{K}$ and [11, Proposition I. 1.4], we know that

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

This immediately implies the desired result that $x = 0$. $\Box$

Proposition 4.1 *Consider the proximal algorithm defined as in (11)-(12), for any $y \in \mathrm{int}(\mathcal{K})$ and $\mu > 0$, we have the following results.*

(a) *Under assumption (A1), the function* $f(\cdot) + \mu^{-1}H(\cdot, y)$ *has bounded level sets.*

(b) *If, in addition, assumption (A2) holds, then there exists a unique* $x(y) \in \mathrm{int}(\mathcal{K})$ *satisfying*

$$x(y) = \operatorname*{argmin}_{x\ \succeq_{\mathcal{K}} 0} \left\{ f(x) + \frac{1}{\mu} H(x, y) \right\}, \tag{18}$$

and the minimum is attained at $x(y) \in \mathrm{int}(\mathcal{K})$ *satisfying*

$$\mu^{-1}(\ln y - \ln x(y)) \in \partial f(x(y)), \tag{19}$$

where ∂f *is the subdifferential of* f *given as in (9).*

Proof. (a) Fix a $y \in \mathrm{int}(\mathcal{K})$ and $\mu > 0$. Let $F_y(x) : \mathbb{V} \to \mathbb{R}$ be a function given by

$$F_y(x) := f(x) + \mu^{-1}H(x, y).$$

By Proposition 3.1 (e), the level sets

$$\mathcal{L}_\gamma(x) := \{\ x \in \mathcal{K} \mid H(x, y) \le \gamma\} \tag{20}$$

are bounded for all $\gamma \ge 0$ and any fixed $y \in \mathrm{int}(\mathcal{K})$. This implies that $F_y(x)$ has bounded level sets by assumption (A1) (since, suppose not, there exists some γ such that $\mathcal{L}_\gamma(x)$ is not bounded).

(b) From part (a), $F_y(x)$ has bounded level sets, which implies the existence of the minimum point $x(y)$. Moreover, the strict convexity of $F_y(x)$ by Proposition 3.1 (a) guarantees the uniqueness. Under assumption (A2), using Proposition 3.1 (b) and the optimality condition for (18) gives that

$$0 \in \partial f(x(y)) + \mu^{-1}\Big(\ln x(y) - \ln y\Big) + \partial \delta(x(y) \mid \mathcal{K}), \tag{21}$$

where $\delta(z|\ \mathcal{K}) = 0$ if $z \succeq_{\mathcal{K}} 0$ and $+\infty$ otherwise. We will show that $\partial\delta(x(y)|\ \mathcal{K}) = 0$ and hence the desired result follows. We observe 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,

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

This by [21, pp. 251] means that $H(\cdot, y)$ is essentially smooth, and hence $\partial_x H(x, y) = \emptyset$ for all $x \in \mathrm{bd}(\mathcal{K})$ by [21, Theorem 26.1]. Thus, from (21), we must have $x(y) \in \mathrm{int}(\mathcal{K})$ (since, if not, there is $\partial_x H(x, y) = \emptyset$ which contradicts (21)). Furthermore, from [21, 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, we then obtain $\partial\delta(x(y)|\,\mathcal{K}) = \{0\}$. Thus, the proof is completed. $\Box$

The next result provides the key properties of the algorithm (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 as in (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) *$\mu_k(f(x^k) - f(x)) \le H(x, x^{k-1}) - H(x, x^k)$ for all $x \in \mathcal{K}$.*

(c) *$\sigma_m(f(x^m) - f(x)) \le H(x, x^0) - H(x, x^m)$ 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 $\operatorname{tr}(x^k - x^{k-1}) \to 0$, if the optimal set $\mathcal{X}^* \neq \emptyset$.*

Proof. (a) By the definition of x^k given as in (11), we know that

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

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

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

(b) From (19), we have $-\mu_k^{-1}(\ln x^k - \ln x^{k-1}) \in \partial f(x^k)$. Plugging this into the formula of $\partial f(x^k)$ given as in (10), we obtain

$$f(x) \ge f(x^k) - \mu_k^{-1}\cdot\operatorname{tr}\Big((x - x^k)\circ(\ln x^k - \ln x^{k-1})\Big) \quad \forall\, x \in \mathcal{K}.$$

This yields

$$\begin{array}{rcl} \mu_k\cdot(f(x^k) - f(x)) & \le & \operatorname{tr}\Big((x - x^k)\circ(\ln x^k - \ln x^{k-1})\Big) \\ & = & H(x, x^{k-1}) - H(x, x^k) - H(x^k, x^{k-1}) \\ & \le & H(x, x^{k-1}) - H(x, x^k), \end{array} \tag{22}$$

where the equality and the last inequality hold due to Proposition 3.3 (b).

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

$$\sum_{k=1}^{m}\mu_k f(x^k) - \sigma_m f(x) \le H(x, x^0) - H(x, x^m) \quad \forall\, x \in \mathcal{K}. \tag{23}$$

On the other hand, note that $f(x^m) \le f(x^k)$ from part(a). Multiplying this inequality by μ_k gives

$$\mu_k f(x^m) \le \mu_k f(x^k).$$

Then summing up leads to

$$\sum_{k=1}^{m} \mu_k f(x^k) \geq \sum_{k=1}^{m} \mu_k f(x^m) = \sigma_m f(x^m). \tag{24}$$

Now, combining (23) and (24) yields the desired result.

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

(e) From part (d), we know that 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,

$$H(x, x^{k-1}) - H(x, x^k) \to 0 \quad \forall\, x \in \mathcal{X}^*. \tag{25}$$

On the other hand, from (22) in the proof of part (b) as well as part (d), we have

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

Hence,

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

Using the equation (25) and the nonnegativity of $H(x^k, x^{k-1})$, the first result is 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). $\square$

By now, we have proved that the algorithm (11)-(12) is well-defined and satisfies some favorable properties. With these properties, we are ready to show the convergence results of the proposed 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 (A1) and (A2), the following results hold.*

(a) *$f(x^m) - f(x) \leq \sigma_m^{-1} H(x, x^0)$ 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 $\mathcal{X}^* \neq \emptyset$, the sequence $\{x^k\}_{k\in N}$ is bounded and every accumulation point is a solution of the problem CSCP.*

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 to the limit in part (a), we have

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

This together with the fact $f(x^m) \geq f_*$ yields the desired 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 [24, page 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 CSCP. $\square$

Proposition 4.3 (a) indicates an estimate for global rate of convergence which is similar to the one obtained by proximal-like algorithms in convex minimization. Analogously, as remarked in [6, Remark 4.1], global convergence of $\{x^k\}$ itself to an optimal 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 The exponential multiplier method for SCLP

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

$$\text{(SCLP)} \quad \min\left\{-b^Ty: \quad c - \sum_{i=1}^{m} y_i a_i \succeq_{\mathcal{K}} 0, \quad y \in \mathbb{R}^m\right\}, \tag{26}$$

where $c, a_i \in \mathbb{V}$ for $i = 1, 2, \cdots, m$ and $b \in \mathbb{R}^m$. For convenience, we denote the feasible set of (SCLP) by $S = \{y \in \mathbb{R}^m : A(y) \succeq_{\mathcal{K}} 0\}$, where $A(y) : \mathbb{R}^m \to \mathbb{V}$ is defined by

$$A(y) := c - \textstyle\sum_{i=1}^{m} y_i a_i. \tag{27}$$

We make the following assumptions for the problem (SCLP):

(S1) The optimal solution set of (SCLP) is nonempty and bounded;

(S2) Slater's condition holds, i.e., there exists a $\hat{y} \in \mathbb{R}^m$ such that $A(\hat{y}) \succ_{\mathcal{K}} 0$.

The Lagrangian function associated with (SCLP) is given as follows:

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

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

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

which can be explicitly written as

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

By standard convex analysis arguments from [21], we know that under the assumption (S2), there is no duality gap between (SCLP) and (DSCLP). Moreover, the set of optimal solution of (DSCLP) is nonempty and compact.

To solve the problem (SCLP), we introduce the following multiplier-type algorithm:

Algorithm 5.1 *Given an $x^0 \in \text{int}(\mathcal{K})$, generate the sequence $\{y^k\}_{k\in N} \subseteq \mathbb{R}^m$ and $\{x^k\}_{k\in N} \subset \text{int}(\mathcal{K})$ by*

$$y^k = \operatorname*{argmin}_{y\in\mathbb{R}^m} \left\{-b^T y + \mu_k^{-1}\text{tr}\left[\exp\Big(-\mu_k A(y) + \ln x^{k-1}\Big)\right]\right\}, \tag{28}$$

$$x^k = \exp\left(-\mu_k A(y^k) + \ln x^{k-1}\right), \tag{29}$$

where the penalty 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, 29]) to symmetric cones. In the rest of this section, we will show that the algorithm (28)-(29) shares similar properties.

We first present two technical lemmas. One of them is to collect some properties of the Löwner operator $\exp(z)$, while the other is to characterize the recession cone of the set S.

Lemma 5.1 **(a)** *For any $z \in \mathbb{V}$, there always holds that $\exp(z) \in \text{int}(\mathcal{K})$.*

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

$$\nabla_y \text{tr}\Big[\exp(\textstyle\sum_{i=1}^m y_i a_i)\Big] = A^T \exp(\textstyle\sum_{i=1}^m y_i a_i)$$

where $y \in \mathbb{R}^m$ and $a_i \in \mathbb{V}$ for $i = 1, 2, \cdots, m$, and $A^T \in \mathbb{R}^{m\times n}$ be the matrix corresponding to the linear transformation that maps $x \in \mathbb{V}$ to the m-vector whose i^{th} component is $\langle a_i, x\rangle$.

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 $j = 1, 2, \cdots, r$, are strictly positive.

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

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

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

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

(c) From the continuous differentiability of trace function and part (b), we immediately obtain the first part. Notice that

$$\mathrm{tr}\Big[\exp(\textstyle\sum_{i=1}^m y_i a_i)\Big] = \Big\langle \exp(\textstyle\sum_{i=1}^m y_i a_i),\ e\Big\rangle.$$

Therefore, applying the differential chain rule readily yields the desired result. $\square$

Lemma 5.2 *Let S_∞ denote the recession cone of the feasible set S. Then,*

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

Proof. Assume that $d \in \mathbb{R}^m$ such that $A(d) - c \succeq_{\mathcal{K}} 0$. If $d = 0$, clearly, $d \in S_\infty$. If $d \neq 0$, we take any $y \in S$. Then $A(y) \succeq_{\mathcal{K}} 0$, and moreover, 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 [21, pp. 61], shows that $d \in S_\infty$. Therefore, $\{d \in \mathbb{R}^m : A(d) - c \succeq_{\mathcal{K}} 0\} \subseteq S_\infty$.

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

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

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

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

which contradicts to $\lambda_{\min}[A(y + \tau d)] \ge 0$. Thus, we have that $A(d) - c \succeq_{\mathcal{K}} 0$. Combining with the above discussions, we get $S_\infty = \{d \in \mathbb{R}^m \ | A(d) - c \succeq_{\mathcal{K}} 0\}$. $\square$

Now we begin with the convergence of the algorithm (28)-(29). We first prove that the sequence generated by (28)-(29) 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*

$$F(y) := -b^T y + \mu^{-1}\mathrm{tr}\left[\exp(-\mu A(y) + \ln x^{k-1})\right] \tag{30}$$

where $\mu > 0$. Then under assumption (S1) the minimum set of F is nonempty and bounded.

Proof. By assumption (S1), it suffices to prove $F(y)$ has bounded level sets. Suppose not, then there exists a sequence $\{y^k\} \subseteq \mathbb{R}^m$ such that

$$\|y^k\| \to +\infty, \quad \lim_{k\to+\infty} \frac{y^k}{\|y^k\|} = d \neq 0 \quad \text{and} \quad F(y^k) \le \delta. \tag{31}$$

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

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

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

$$\begin{gathered} -b^T y^k < \delta, \\ \sum_{j=1}^r \exp\left[\lambda_j\Big(-\mu A(y^k) + \ln x^{k-1}\Big)\right] \le \mu(\delta + b^T y^k). \end{gathered}$$

From this, we immediately obtain that

$$-b^T d \le 0, \tag{32}$$

$$\exp\left[\lambda_j\Big(-\mu A(y^k) + \ln x^{k-1}\Big)\right] \le \mu(\delta + b^T y^k), \quad \forall\, j = 1, 2, \cdots, r. \tag{33}$$

Using the monotonicity of $\ln t$ ($t > 0$) and the homogeneity of $\lambda_j(\cdot)$, we can rewrite (33) as

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

where the second inequality is due to $\ln t \le t - 1$ ($t > 0$). Passing to the limit on the both sides of the inequality and using the homogeneity of $\lambda_j(\cdot)$ again, we then get

$$\lambda_j(A(d) - c) \ge 0, \quad \forall\, j = 1, 2, \cdots, r,$$

which in turn means that

$$A(d) - c \succeq_{\mathcal{K}} 0. \tag{34}$$

From Lemma 5.2, we know that $\{d \in \mathbb{R}^m : A(d) - c \succeq_{\mathcal{K}} 0\} = S_\infty$. Combining with (32), we thus show that there exists a nonzero $d \in \mathbb{R}^m$ such that $d \in S_\infty$ but $-b^T d \le 0$. This clearly contradicts the assumption (S1). Hence, the proof is completed. $\Box$

To analyze the convergence of the algorithm (28)-(29), we also need the following result which states that the sequence $\{x^k\}_{k\in N}$ generated by (28)-(29) is exactly the sequence produced by the algorithm (11)-(12) when applied to the dual problem (DSCLP).

Lemma 5.4 *The sequence $\{x^k\}_{k\in N}$ generated by the multiplier method (28)-(29) can be obtained via the following iterate scheme*

$$x^k = \operatorname*{argmax}_{x\succeq_{\mathcal{K}}0} \left\{h(x) - \mu_k^{-1}H(x, x^{k-1})\right\}, \tag{35}$$

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

Proof. First, we show that $-A(y^k) \in \partial h(x^k)$. Using Lemma 5.1 (c) and the optimality condition of (28), we obtain that

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

where the second equality is due to (29). This implies that y^k is also minimizing the Lagrangian $L(y, x^k)$, and consequently $h(x^k) = L(y^k, x^k)$. Now, we have that

$$\begin{aligned} h(x) &= \inf_{y\in\mathbb{R}^m} \left\{-b^T y - \operatorname{tr}[x \circ A(y)]\right\} \\ &\le -b^T y^k - \operatorname{tr}[x \circ A(y^k)] \\ &= -b^T y^k - \operatorname{tr}[x^k \circ A(y^k)] - \operatorname{tr}[(x - x^k) \circ A(y^k)] \\ &= h(x^k) + \langle x - x^k,\ -A(y^k)\rangle. \end{aligned} \tag{36}$$

Now, considering that $h(x)$ is concave, the inequality (36) means that $-A(y^k) \in \partial h(x^k)$. Combining with (29), we therefore have

$$\mu_k^{-1}(\ln x^k - \ln x^{k-1}) \in \partial h(x^k).$$

From Proposition 3.1, we see that this is precisely the optimality condition of the maximum problem in (35). Thus, the proof is completed. $\Box$

Now we are ready to present the convergence results of the algorithm defined as (28)-(29). Their proof techniques 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 (28)-(29). Then, under assumptions (S1) and (S2), the following results hold.*

(a) *The dual sequence $\{x^k\}_{k\in N} \subset \operatorname{int}(\mathcal{K})$ is bounded and all of its limit points are optimal dual solutions.*

(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 h^* be the optimal value of the dual problem (*DSCLP*). Then $-b^T y^k \to h^*$, and consequently, $-b^T \tilde{y}^k \to h^*$.*

(e) *$\{\tilde{y}^k\}$ is bounded and its every limit point is an optimal solution of (*SCLP*).*

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

Proof. (a) From Lemma 5.4, we know that $\{x^k\}_{k\in N}$ is the sequence generated by the entropy-like proximal algorithm (11)-(12) applied to (DSCLP). Since under assumption (S2) the set of optimal solutions of (DSCLP) is nonempty and compact, the result directly follows from Proposition 4.3.

(b) The equation (29) implies that $-\mu_k A(y^k) = \ln x^k - \ln x^{k-1}$. On the other hand,

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

From Proposition 3.2 (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 that $\mathrm{tr}[x^k \circ A(y^k)] \to 0$.

(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 (29). 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

$$\begin{aligned} \liminf_{k\to+\infty} \nu_k^{-1}\lambda_{\min}(-\ln x^k) &= -\limsup_{k\to+\infty} \nu_k^{-1}\lambda_{\max}(\ln x^k) \\ &= -\limsup_{k\to+\infty} \nu_k^{-1}\ln(\lambda_{\max}(x^k)). \end{aligned} \tag{37}$$

Since $\{x^k\}_{k\in N} \subset \mathrm{int}(\mathcal{K})$ is bounded, $\lambda_{\max}(x^k) \leq \lambda_0$ for some $\lambda_0 > 0$, and thus

$$-\limsup_{k\to+\infty} \nu_k^{-1}\ln(\lambda_{\max}(x^k)) \geq -\limsup_{k\to+\infty} \nu_k^{-1}\ln\lambda_0 \geq 0.$$

Then, combining with (37) yields the desired result.

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

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

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

(e) Suppose that $\{\tilde{y}^k\}$ is unbounded. Since the optimal solution set of (SCLP) is bounded, we let $\hat{y}^*$ be its element with the maximum norm. 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

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

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). Without loss of generality, let z^* be a limit point of $\{z^k\}$ and $z^k \to z^*$. 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 then obtain that $A(z^*) \succeq_{\mathcal{K}} 0$, i.e., z^* is a feasible point for (SCLP), which in turn means that $b^T\hat{y}^* \geq b^T z^*$. On the other hand, since $-b^T\tilde{y}^k \to h^* \leq -b^T\hat{y}^*$ 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 (S1)). On the other hand, from part (d) and the weak duality, it follows that

$$-b^T\tilde{y}^* = h^* \leq -b^T y^*.$$

At last, combining the last two equations yields that

$$-b^T\tilde{y}^* = -b^T y^* = h^*. \tag{38}$$

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

(f) The first equality is due to part (d) and the second is from (38). $\square$

We observe that the above convergence properties of the algorithm (28)-(29) are very similar to the ones obtained by [29] 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 function values, and moreover, by applying Proposition 4.3 (a) with $x = x^*$, one has the global convergence rate estimate:

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

where x^* is an optimal solution of (DSCLP).

6 Conclusions and final remarks

In this paper, we have developed a unified entropy-like proximal algorithm and the corresponding dual augmented Lagrangian method for convex symmetric optimization problems. 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. In addition, we also 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 method was shown to share many similar properties with the exponential multiplier method developed by [29] for convex minimization with nonnegative orthant cone constraints.

The earlier version of this paper is titled "A proximal-like algorithm for convex second-order cone programming". During the reviewing process, several directions as suggested by two referees were considered in other new manuscripts [19, 20]. We therefore follow the direction suggested by the editor which is to extend the proposed algorithm to more symmetric cone programming. We have considered the linear constraints in the second algorithm whereas we only focus on the constraint $x \succeq_{\mathcal{K}} 0$ in the first algorithm. Nonetheless, by similar modifications as done in [20], one can relax the constraint to $Ax + b \succeq_{\mathcal{K}} 0$ which may have more realistic applications.

There have some topics worthy of further investigation. For instance, we may consider to expand choice of $H(\cdot,\cdot)$ given as in (12) to more general classes of distance functions. Another direction for future study is to present inexact entropy-like proximal algorithms for the convex symmetric cone programs and establish the corresponding convergence analysis, which are very important from the viewpoint of applications. We leave them as future research topics.

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

References

[1] F. Alizadeh and D. Goldfarb, *Second-order cone programming*, Mathematical Programming, vol. 95, pp. 3–51, 2003.

[2] M. Baes, *Convexity and differentiability properties of spectral functions and spectral mappings on Euclidean Jordan algebras*, Linear Algebra and its Applications, vol. 422, pp. 664–700, 2007.

[3] D. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academuc Press, New York, 1982.

[4] G. Chen and M. Teboulle, *Convergence analysis of a proximal-like minimization algorithm using Bregman functions*, SIAM Journal on Optimization, vol. 3, pp. 538–543, 1993.

[5] Y. Censor and S. A. Zenios *The proximal minimization algorithm with D-functions*, Journal of Optimization Theory and Applications, vol. 73, pp. 451–464, 1992.

[6] M. Doljansky and M. Teboulle, *An interior proximal algorithm and the exponential multiplier method for semidefinite programming*, SIAM Journal on Optimization, vol. 9, pp. 1–13, 1998.

[7] J. Eckstein, *Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming*, Mathematics of Operation Research, vol. 18, pp. 206–226, 1993.

[8] P. B. Eggermont, *Multilicative iterative algorithms for convex programming*, Linear Algebra and its Applications, vol. 130, pp. 25–42, 1990.

[9] L. Faybusovich, *Euclidean Jordan algebras and interior point algorithm*, Positivity, vol. 1, pp. 331–357, 1997.

[10] L. Faybusovich, *Linear systems in Jordan algebras and primal-dual interior point algorithm*, Journal of Compuation and Applied Mathematics, vol. 86, pp. 149–175, 1997.

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

[12] O. Guler, *On the convergence of the proximal point algorithm for convex minimization*, SIAM Journal on Control Optimization, vol. 29, pp. 403–419, 1991.

[13] A. Iusem and M. Teboulle, *Convergence rate analysis of nonquadratic proximal method for convex and linear programming*, Mathematics of Operation Research, vol. 20, pp. 657–677, 1995.

[14] A. Iusem, B. Svaiter, and M. Teboulle, *Entropy-Like proximal methods in convex programming*, Mathematics of Operation Research, vol. 19, pp. 790-814, 1994.

[15] M. Koecher, *The Minnesota Notes on Jordan Algebras and Their Applications*, edited and annotated by A. Brieg and S. Walcher, Springer, Berlin, 1999.

[16] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, *Application of second-order cone programming*, Linear Algebra and its Applications, vol. 284, pp. 193-228, 1998.

[17] B. Martinet, *Perturbation des methodes d'Optimisation*, Application, R.A.I.R.O. Numer. Anal., 1978.

[18] J. J. Moreau, *Promimité et Dualité dans un Espace Hilbertien*, Bull. Soc. Math. France, vol. 93, pp. 273–299, 1965.

[19] S-H. Pan and J-S. Chen, *A class of interior proximal-like algorithms for convex second-order cone programming*, Submitted, 2007.

[20] S-H. Pan and J-S. Chen, *A proximal-like algorithm using quasi D-function for convex second-order cone programming*, Submitted, 2007.

[21] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[22] R. T. Rockafellar, *Augmented lagrangians and applications of proximal point algorithm in convex programming*, Mathematics of Operation Research, vol. 1, pp. 97-116, 1976.

[23] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization, vol. 14, pp. 877-898, 1976.

[24] R. T. Rockafellar and R. J-B. Wets, *Variational Analysis*, Springer-Verlag, 1998.

[25] S. H. Schmieta and F. Alizadeh, *Extension of primal-dual interior point algorithms to symmetric cones*, Mathematical Programming, vol. 96, pp. 409–438, 2003.

[26] D. Sun and J. Sun, *Löwner's operator and spectral functions on Euclidean Jordan algebras*, Submitted to Mathematics of Operations Research, 2004.

[27] M. Teboulle, *Entropic proximal mappings with applications to nonlinear programming*, Mathematics of Operation Research , vol. 17, pp. 670–690, 1992.

[28] M. Teboulle, *Convergence of proximal-like algorithms*, SIAM Journal on Optimization, vol. 7, pp. 1069-1083, 1997.

[29] P. Tseng and D. Bertsekas, *On the convergence of the exponential multiplier method for convex programming*, Mathematical Programming, vol. 60, pp. 1-19, 1993.

[30] L. Vandenberghe and S. Boyd, *A primal-dual potential reduction method for problems involving matrix inequalities*, Mathematical Programming, vol. 69, pp. 205–236, 1995.