A Proximal Multiplier Method for Convex Separable Symmetric Cone Optimization

Julio López[a,*], Erik Papa Quiroz[b]

[a]*Facultad de Ingeniería, Universidad Diego Portales, Ejército 441, Santiago, Chile*
[b]*Facultad de Ciencias Matemáticas, Universidad Nacional Mayor de San Marcos, Lima, Perú*

Abstract

This work is devoted to the study of a proximal decomposition algorithm for solving convex symmetric cone optimization with separable structures. The algorithm considered is based on the decomposition method proposed by Chen and Teboulle [1], and the proximal generalized distance defined by Auslender and Teboulle [2]. Under suitable assumptions, first a class of proximal distances is constructed, therefore some examples are given. Second, it is proven that each limit point of the primal-dual sequences generated by the algorithm solves the problem. Finally, the global convergence is established.

Keywords: Euclidean Jordan algebra, symmetric cone optimization, decomposition method, proximal distance, proximal multiplier method.

1. Introduction

Let $\mathbb{V}_1$ and $\mathbb{V}_2$ be two finite dimensional vectorial spaces on the real field $\mathbb{R}$ equipped with the inner product $\langle \cdot, \cdot \rangle_{\mathbb{V}_1}$ and $\langle \cdot, \cdot \rangle_{\mathbb{V}_2}$ respectively. Then consider the following Euclidean Jordan algebra $\mathbb{V}_1 = (\mathbb{V}_1, \circ_1, \langle \cdot, \cdot \rangle_{\mathbb{V}_1})$ and $\mathbb{V}_2 = (\mathbb{V}_2, \circ_2, \langle \cdot, \cdot \rangle_{\mathbb{V}_2})$, where $\circ_1$ and $\circ_2$ denote the Jordan product in $\mathbb{V}_1$ and $\mathbb{V}_2$, respectively. In this work, we consider the following convex *symmetric cone optimization* (SCP) problem with separable structure:

$$(P) \qquad \min\{f(x) + g(z) : \mathbb{A}x + \mathbb{B}z = b, x \in \mathcal{K}_1, z \in \mathcal{K}_2\},$$

where $f : \mathbb{V}_1 \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{V}_2 \to \mathbb{R} \cup \{+\infty\}$ are closed proper convex function, $\mathbb{A} : \mathbb{V}_1 \to \mathbb{R}^m$ and $\mathbb{B} : \mathbb{V}_2 \to \mathbb{R}^m$ are two linear mapping, $b \in \mathbb{R}^m$, and $\mathcal{K}_1 := \{x \circ_1 x : x \in \mathbb{V}_1\}$ and $\mathcal{K}_2 := \{z \circ_2 z : z \in \mathbb{V}_2\}$ denoting the sets of square elements (symmetric cones) in $\mathbb{V}_1$ and $\mathbb{V}_2$, respectively.

This problem provides a simple unified framework for various existing optimization problems such as nonlinear, second-order cone, and semidefinite programming, and hence has extensive applications in engineering, economics, game theory, management science, and other fields; see [3, 4, 5, 6, 7] and references therein. In particular, a class of problems of (P) have drawn recently a lot of attention due to their emerging applications of practical interest, as for example support vector machines or the sparse inverse covariance selection (see Section 4 for details).

---

*Corresponding author. Tel: +56-2-26762457.
Email addresses: `julio.lopez@udp.cl` (Julio López), `erikpapa@gmail.com` (Erik Papa Quiroz)

Recently, SCP has attracted the attention of some researcher focusing on the development of interior-point methods similar to those for linear optimization (see [8, 9, 10, 11]), since the class of symmetric cones contains the positive orthant in $\mathbb{R}^n$, the second-order cone, and the cone of positive semidefinite symmetric matrices.

Various decomposition methods that exploit the separable structure of the objetive function of the optimization problems have been proposed for solving nonlinear minimization ones, for instance: alternating direction method of multipliers [12], partial inverse of Spingarn method [13], the predictor corrector proximal multiplier (PCPM) methods of Chen and Teboulle [1], among others. In this work, we are interested in the last method and its extensions. The idea of the PCPM one for solving (P) without conical constraints reads as follow

$$
\begin{cases}
p^{k+1} = y^k + \lambda_k (Ax^k + Bz^k - b), \\
x^{k+1} = \arg\min_{x \in \mathbb{R}^n} \{ f(x) + \langle p^{k+1}, Ax \rangle + \dfrac{1}{2\lambda_k} \|x - x^k\|^2, \\
z^{k+1} = \arg\min_{x \in \mathbb{R}^n} \{ g(z) + \langle p^{k+1}, Bx \rangle + \dfrac{1}{2\lambda_k} \|z - z^k\|^2 \}, \\
y^{k+1} = y^k + \lambda_k (Ax^{k+1} + Bz^{k+1} - b),
\end{cases}
\tag{1}
$$

where $\{\lambda_k\}$ is a sequence of positive parameters. Auslender and Teboulle [14], and Kyono and Fukushima [15], independently, extend this method by using the logarithmic-quadratic proximal distance [16], and the Bregman proximal distance [17], respectively, instead of the Euclidean distance. The first approach is designed to solve variational inequality problems with separable structure, while the second one is restricted to solve convex programs with the same structure.

Recently, Sarmiento et al. [18] develop an extension of the (PCPM) method for solving (P), when the Euclidean Jordan algebra is the well known Euclidean vectorial space, by using regularized proximal distances [2]. In order to cover a broad range of applications, as for example separable semidefinite and second-order cone optimization problems (see also our Section 4), which cannot be covered by the work [18], the purpose of this paper is to extend the (PCPM) method so that it can be used to solve the (CP) problems. Specifically, first we define a proximal distance (similarly to [2, 19]) with respect to the interior of the symmetric cone $\mathcal{K}$. This distance can be generated by an appropriate class of continuously differentiable strict convex functions on $\mathbb{R}_{++}$. Second, we propose a proximal decomposition algorithm using proximal distance for solving (P). Then, under some mild assumptions, we establish the global convergence of the algorithm proposed.

This paper is organized as follows. Section 2 recalls some basic notion and reviews some basic results on Euclidean Jordan algebras. In Section 3 we introduce the class of proximal distances defined in the symmetric cone, provide two ways to construct it, and present some examples. In Section 4, we describe the algorithm with proximal distance for solving (P) and establish its global convergence. Finally, concluding remarks are given in Section 5.

2. Preliminaries

The following notation and terminology are used throughout the paper. Let $\mathbb{V}$ be a vectorial space, for a closed proper convex function f, its effective domain is defined by $\mathrm{dom}(f) = \{x \in \mathbb{V} : f(x) < +\infty\}$, for some $\varepsilon \geq 0$, $\partial_\varepsilon f(x) = \{p \in \mathbb{V} : f(x) + \langle p, z - x \rangle - \varepsilon \leq f(z), \ \forall z \in \mathbb{V}\}$ denotes its ε-subdifferential at x, $\partial f = \partial_0 f$ its subdifferential [20]. For any closed convex set $S \subset \mathbb{R}^n$,

δ_S denotes the indicator function of S and $\mathcal{N}_S(x)$ denotes the normal cone to S at $x \in S$. For a symmetric cone $\mathcal{K}$, its interior and boundary are denoted by $\text{int}(\mathcal{K})$ and $\text{bd}(\mathcal{K})$, respectively. Given a linear mapping $\mathbb{A} : \mathbb{V} \to \mathbb{R}^m$, we denote its adjoint by $\mathbb{A}^*$ which is defined by $\langle \mathbb{A}x, y \rangle = \langle x, \mathbb{A}^*y \rangle$ for all $x \in \mathbb{V}$, $y \in \mathbb{R}^m$.

2.1. Euclidean Jordan algebra

In this subsection, we briefly describe some concepts, properties, and results from Euclidean Jordan algebras that are needed in this paper and they have become important in the study of conic optimization; see, e.g., Schmieta and Alizadeh [11]. Most of this material can be found in Faraut and Korányi [21].

A *Euclidean Jordan algebra* is a triple $(\mathbb{V}, \circ, \langle \cdot, \cdot \rangle_{\mathbb{V}})$, where $(\mathbb{V}, \langle \cdot, \cdot \rangle_{\mathbb{V}})$ is a finite-dimensional space over the real field $\mathbb{R}$ equipped with an inner product $\langle \cdot, \cdot \rangle_{\mathbb{V}}$, and the Jordan product $(x, y) \mapsto x \circ y : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is a bilinear mapping satisfying the following three conditions:

(i) $x \circ y = y \circ x$ for all $x, y \in \mathbb{V}$,
(ii) $x \circ (x^2 \circ y) = x^2 \circ (x \circ y)$ for all $x, y \in \mathbb{V}$ where $x^2 = x \circ x$, and
(ii) $\langle x \circ y, z \rangle_{\mathbb{V}} = \langle y, x \circ z \rangle_{\mathbb{V}}$ for all $x, y, z \in \mathbb{V}$,

and there exists a (unique) unitary element $e \in \mathbb{V}$ such that $x \circ e = x$ for all $x \in \mathbb{V}$. Henceforth, we simply say that $\mathbb{V}$ is a Euclidean Jordan algebra and $x \circ y$ is called the *Jordan product* of x and y. We say that an element $x \in \mathbb{V}$ is *invertible*, if there exists $w \in \mathbb{V}$ such that $x \circ w = e$.

In a Euclidean Jordan algebra $\mathbb{V}$, is known that the set of squares $\mathcal{K} = \{x^2 : x \in \mathbb{V}\}$ is a *symmetric cone* (see [21, Theorem III.2.1]). This means that $\mathcal{K}$ is a self-dual closed and 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 $\mathcal{T}(x) = y$.

The *rank* of $(\mathbb{V}, \circ, \langle \cdot, \cdot \rangle_{\mathbb{V}})$ is defined as $r = \max\{\deg(x) : x \in \mathbb{V}\}$, where $\deg(x)$ is the degree of $x \in \mathbb{V}$ given by $\deg(x) = \min\{k > 0 : \{e, x, x^2, \ldots, x^k\}$ is linearly dependent$\}$.

An element $c \in \mathbb{V}$ is an *idempotent* if $c^2 = c$; it is a *primitive idempotent* iff it is nonzero and cannot be written as a sum of two nonzero idempotents. We say that a finite set $\{e_1, \ldots, e_r\}$ of primitive idempotents in $\mathbb{V}$ is a *Jordan frame* iff

$$e_i^2 = e_i, \ e_i \circ e_j = 0 \ \text{for all} \ i \neq j, \ \text{and} \ \sum_{i=1}^{r} e_i = e. \tag{2}$$

Note that $\langle e_i, e_j \rangle_{\mathbb{V}} = \langle e_i \circ e_j, e \rangle_{\mathbb{V}} = 0$ whenever $i \neq j$.

The following theorem gives us a spectral decomposition for the elements in a Euclidean Jordan algebra (see Theorem III.1.2 of [21]).

Theorem 2.1 (Spectral decomposition theorem). *Suppose that $(\mathbb{V}, \circ, \langle \cdot, \cdot \rangle_{\mathbb{V}})$ is a Euclidean Jordan algebra with rank r. Then, for every $x \in \mathbb{V}$, there exist a Jordan frame $\{e_1(x), \ldots, e_r(x)\}$ and real numbers $\lambda_1(x), \ldots, \lambda_r(x)$, arranged in the decreasing order, such that $x = \lambda_1(x)e_1(x) + \cdots + \lambda_r(x)e_r(x)$.*

The numbers $\lambda_i(x)$ (counting multiplicities), which are uniquely determined by x, are called the *eigenvalues* of x. The *trace* of x, denoted as $\text{tr}(x)$, is defined by $\text{tr}(x) := \sum_{j=1}^{r} \lambda_j(x)$; whereas the *determinant* of x is defined by $\det(x) := \prod_{j=1}^{r} \lambda_j(x)$.

It is easy to show that $x \in \mathcal{K}$ (resp. $\text{int}(\mathcal{K})$) iff every eigenvalue $\lambda_i(x)$ of x is nonnegative (resp. positive). Moreover, an element $x \in \mathbb{V}$ is *invertible*, if $\det(x) \neq 0$, that is, if every eigenvalue of x is nonzero.

3

Example 2.1. Typical examples of Euclidean Jordan algebras are the following:

(i) *Euclidean Jordan algebra of n-dimensional vectors:*

$$\mathbb{V} = \mathbb{R}^n, \quad \mathcal{K} = \mathbb{R}^n_+, \quad r = n, \quad \langle x, y \rangle_{\mathbb{R}^n} = \sum_{i=1}^n x_i y_i, \quad x \circ y = x * y,$$

where $x * y$ denotes the componentwise product of vectors x and y. Here, the unitary element is $e = (1, \ldots, 1) \in \mathbb{R}^n$. On the other hand, the spectral decomposition of any $x \in \mathbb{R}^n$ is given by $x = \sum_{i=1}^n x_i e_i$, where e_i is a vector with 1 in the i-th entry and 0's elsewhere.

(ii) *Euclidean Jordan algebra of quadratic forms:*

$$\mathbb{V} = \mathbb{R}^n, \quad \mathcal{K} = \mathcal{L}^n_+ = \{x = (x_1, \bar{x}) \in \mathbb{R} \times \mathbb{R}^{n-1} : \|\bar{x}\| \le x_1\}, \quad r = 2,$$

$$\langle x, y \rangle_{\mathbb{R}^n} = \sum_{i=1}^n x_i y_i, \quad x \circ y = (x_1, \bar{x}) \circ (y_1, \bar{y}) = (\langle x, y \rangle, x_1 \bar{y} + y_1 \bar{x}),$$

where $\bar{x} = (x_2, \ldots, x_n) \in \mathbb{R}^{n-1}$, and $\|\cdot\|$ denotes the Euclidean norm. In this algebra, the cone of squares is called the Lorentz cone (or the second-order cone). Moreover, the unitary element is $e = (1, 0, \ldots, 0) \in \mathbb{R}^n$. On the other hand, the spectral decomposition of any $x \in \mathbb{R}^n$ associated with $\mathcal{L}^n_+$ is given by $x = \lambda_1(x)u_1(x) + \lambda_2(x)u_2(x)$, where $\lambda_i(x) = x_1 + (-1)^i \|\bar{x}\|$ and $u_i(x) = \frac{1}{2}(1, (-1)^i \frac{\bar{x}}{\|\bar{x}\|})$, for $i = 1, 2$, denote the eigenvalues and eigenvectors of x, respectively.

(iii) *Euclidean Jordan algebra of n-dimensional symmetric matrices:* Let $\mathcal{S}^n$ be the set of all $n \times n$ real symmetric matrices, and $\mathcal{S}^n_+$ be the cone of $n \times n$ symmetric positive semidefinite matrices.

$$\mathbb{V} = \mathcal{S}^n, \quad \mathcal{K} = \mathcal{S}^n_+, \quad r = n, \quad \langle X, Y \rangle_{\mathcal{S}^n} = \text{tr}(XY), \quad X \circ Y = \frac{1}{2}(XY + YX).$$

Here tr denotes the trace of a matrix $X = (x_{ij}) \in \mathcal{S}^n$. In this setting, the identity matrix $I \in \mathbb{R}^{n \times n}$ is the unit element e. On the other hand, the spectral decomposition of any $X \in \mathcal{S}^n$ is given by $X = \sum_{i=1}^n \lambda_i(X) q_i(X) q_i(X)^\top$, where $\lambda_i(X)$ and $q_i(X) \in \mathbb{R}^n$ denote the eigenvalue and eigenvector of X, respectively.

Other examples are the set of $n \times n$ hermitian positive semidefinite matrices made of complex numbers, the set of $n \times n$ positive semidefinite matrices with quaternion entries, the set of 3×3 positive semidefinite matrices with octonion entries, and the exceptional 27-dimensional Albert octonion cone (see [21, 22]).

By [9, Proposition III.1.5], a Jordan algebra over $\mathbb{R}$ with a unit element $e \in \mathbb{V}$ is Euclidean iff the symmetric bilinear form $\text{tr}(x \circ y)$ is positive definite. Hence, we may define another inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{V}$ by

$$\langle x, y \rangle := \text{tr}(x \circ y), \quad \forall x, y \in \mathbb{V}. \tag{3}$$

Furthermore, we can define the norm, induced by the inner product $\langle x, y \rangle$, on $\mathbb{V}$ by

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

3. Proximal distances over symmetric cones

In this section, first, we present the definition of proximal distances with respect to the open convex cone $\mathrm{int}(\mathcal{K})$ and, then provide two ways to construct proximal distances with respect to this open convex cone.

Definition 3.1. *Let $\mathbb{V}$ be a finite dimensional vectorial space. An extented-valued function $H : \mathbb{V} \times \mathbb{V} \to \mathbb{R} \cup \{+\infty\}$ is called a proximal distance with respect to $\mathrm{int}(\mathcal{K})$ iff it satisfies the following properties:*

(P1) $\mathrm{dom}(H(\cdot, \cdot)) = \mathcal{C}_1 \times \mathcal{C}_2$ *with* $\mathrm{int}(\mathcal{K}) \times \mathrm{int}(\mathcal{K}) \subset \mathcal{C}_1 \times \mathcal{C}_2 \subseteq \mathcal{K} \times \mathcal{K}$.

(P2) *For each $y \in \mathrm{int}(\mathcal{K})$, $H(\cdot, y)$ is continuous and strictly convex on $\mathcal{C}_1$, and it is continuously differentiable on $\mathrm{int}(\mathcal{K})$ with $\mathrm{dom}(\partial_x H(\cdot, y)) = \mathrm{int}(\mathcal{K})$, where $\partial_x H(\cdot, y)$ denotes the subdifferential of $H(\cdot, y)$ with respect to the first variable.*

(P3) $H(u, v) \geq 0$ *for all $u, v \in \mathbb{V}$, and $H(v, v) = 0$ for all $v \in \mathrm{int}(\mathcal{K})$.*

(P4) *For each $\gamma \in \mathbb{R}$, the lower level set $\{u \in \mathcal{C}_1 : H(u, v) \leq \gamma\}$ is bounded for any $v \in \mathcal{C}_2$.*

This definition was considered in [19] for a proximal distance with respect to interior of the second-order cone $\mathcal{L}^n_{++} := \mathrm{int}(\mathcal{L}^n_+)$, and it has a little difference from Definition introduced by Auslender and Teboulle (see [2, Definition 2.1]), since here $H(\cdot, y)$ is required to be strictly convex over $\mathcal{C}_1$ for any fixed $y \in \mathrm{int}(\mathcal{K})$. Let us denote by $\mathcal{D}(\mathrm{int}(\mathcal{K}))$ the family of functions H satisfying Definition 3.1.

Unfortunately, the above class is not sufficient to guarantee the convergence of the proposed algorithm, so we will use some of the following extra conditions on the proximal distance H:

(B1) For any $u, v \in \mathrm{int}(\mathcal{K})$ and $w \in \mathcal{C}_1$, $\langle \nabla_u H(u, v), w - u \rangle \leq H(w, v) - H(w, u)$.

(B1′) For any $u, v \in \mathrm{int}(\mathcal{K})$ and $w \in \mathcal{C}_2$, $\langle \nabla_u H(u, v), w - u \rangle \leq H(v, w) - H(u, w)$.

(B2) For each $u \in \mathcal{C}_1$, $H(u, \cdot)$ is level bounded on $\mathcal{C}_2$.

(B3) For any $\{v^k\} \subseteq \mathrm{int}(\mathcal{K})$ such that $v^k \to v^*$, and $c \in \mathcal{C}_1$, we have that $H(c, v^k) \to H(c, v^*)$.

(B3′) For any $\{v^k\} \subseteq \mathrm{int}(\mathcal{K})$ such that $v^k \to v^*$, and $c \in \mathcal{C}_2$, we have that $H(v^k, c) \to H(v^*, c)$.

(B4) For any $\{v^k\} \subseteq \mathrm{int}(\mathcal{K})$ converging to $v^* \in \mathcal{K}$, we have that $H(v^*, v^k) \to 0$.

(B4′) For any $\{v^k\} \subseteq \mathrm{int}(\mathcal{K})$ converging to $v^* \in \mathcal{K}$, we have that $H(v^k, v^*) \to 0$.

3.1. Constructing proximal distances over SCs

Let $\phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a scalar-valued function. Following [23, 24], one can define a vector-valued function $\phi^{\mathrm{sc}} : \mathbb{V} \to \mathbb{V} \cup \{+\infty\}$ called *Löwner's operator*, by

$$\phi^{\mathrm{sc}}(x) := \phi(\lambda_1(x))e_1(x) + \cdots + \phi(\lambda_r(x))e_r(x), \quad \text{if } \lambda_i(x) \in \mathrm{dom}(\phi), \tag{4}$$

and $+\infty$ otherwise, where $x \in \mathbb{V}$ has the following spectral decomposition $x = \sum_{i=1}^r \lambda_i(x)e_i(x)$, with $\lambda_1(x) \geq \lambda_2(x) \geq \ldots \geq \lambda_r(x)$. Then, we can consider its corresponding *spectrally defined* function $\Phi : \mathbb{V} \to \mathbb{R} \cup \{+\infty\}$ given by

$$\Phi(x) := \mathrm{tr}(\phi^{\mathrm{sc}}(x)) = \sum_{i=1}^r \phi(\lambda_i(x)), \quad \text{if } \lambda_i(x) \in \mathrm{dom}(\phi). \tag{5}$$

The following result gives the first derivative of the functions Φ and ϕ^{sc}. Their proofs can be found in [23, Theorem 38] and [24, Theorem 3.2].

Lemma 3.1. *Let ϕ be a continuously differentiable function on a subset of $\mathrm{dom}(\phi)$ and $x \in \mathbb{V}$ with its spectral decomposition $x = \sum_{i=1}^{r} \lambda_i(x)e_i(x)$. Then*

(a) *Φ is continuously differentiable on $\mathrm{dom}((\phi')^{\mathrm{sc}})$ and*

$$\nabla\Phi(x) = \sum_{i=1}^{r} \phi'(\lambda_i(x))e_i(x) = (\phi')^{\mathrm{sc}}(x), \quad \forall x \in \mathrm{dom}((\phi')^{\mathrm{sc}}). \tag{6}$$

(b) *$(\phi)^{\mathrm{sc}}$ is continuously differentiable on $\mathrm{dom}((\phi')^{\mathrm{sc}})$ and*

$$\nabla(\phi)^{\mathrm{sc}}(x)x = \sum_{i=1}^{r} \psi'(\lambda_i(x))\lambda_i(x)e_i(x) = (\psi')^{\mathrm{sc}}(x) \circ x, \quad \forall x \in \mathrm{dom}((\phi')^{\mathrm{sc}}). \tag{7}$$

Next, we provide with two ways to construct a function H in terms of the function ϕ, and give some conditions on ϕ in order to show that H is a proximal distance. For the first approach, let us consider $\phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ a closed proper convex function with $\mathrm{dom}(\phi) \subseteq [0, +\infty)$, and $\mathrm{int}(\mathrm{dom}(\phi)) = (0, +\infty)$, define the function $d_\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ by

$$d_\phi(s,t) := \begin{cases} \phi(s) - \phi(t) - \phi'(t)(s-t) & , \quad s \in \mathrm{dom}(\phi),\, t \in \mathbb{R}_{++}, \\ +\infty & , \quad \text{otherwise}, \end{cases} \tag{8}$$

and, suppose the following conditions:

(S.1) ϕ is continuous and strictly convex on its domain.

(S.2) ϕ is continuously differentiable on $\mathbb{R}_{++}$.

(S.3) For each $\gamma \in \mathbb{R}$, the level sets $\{s \in \mathrm{dom}(\phi) : d_\phi(s,t) \le \gamma\}$ and $\{t \in \mathbb{R}_{++} : d_\phi(s,t) \le \gamma\}$ are bounded for any $t \in \mathbb{R}_{++}$ and $s \in \mathrm{dom}(\phi)$, respectively.

(S.4) If $\{t_k\} \subset \mathbb{R}_{++}$ is such that $\lim_{k \to +\infty} t_k = 0$, then $\lim_{k \to +\infty} \phi'(t_k)(s - t_k) = -\infty$, $\forall s \in \mathbb{R}_{++}$.

We denote by Σ the class of functions ϕ that satisfy the above assumptions, and we define the following function $H : \mathbb{V} \times \mathbb{V} \to \mathbb{R} \cup \{+\infty\}$ by

$$H(x,y) := \begin{cases} \Phi(x) - \Phi(y) - \langle \nabla\Phi(y), x - y \rangle & , \quad x \in \mathrm{dom}(\Phi),\, y \in \mathrm{int}(\mathcal{K}), \\ +\infty & , \quad \text{otherwise}. \end{cases} \tag{9}$$

Now, we study some important properties of the function H. This result extends [25, Proposition 3.1 and 3.3] to our context.

Proposition 3.2. *Given $\phi \in \Sigma$, let H be the function defined by (9). Then the following results hold:*

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

(b) *For any fixed $y \in \mathrm{int}(\mathcal{K})$, $H(\cdot, y)$ is continuously differentiable on $\mathrm{int}(\mathcal{K})$ with*

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

(c) *$H(x,y) \ge 0$ for any $x \in \mathrm{dom}(\Phi)$ and $y \in \mathrm{int}(\mathcal{K})$, and $H(x,y) = 0$ iff $x = y$.*

6

(d) $H(x, y) \geq \sum_{i=1}^{r} d_\phi(\lambda_i(x), \lambda_i(y)) \geq 0$, *for any $x \in \mathrm{dom}(\Phi)$ and $y \in \mathrm{int}(\mathcal{K})$.*

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

(f) *For any $x, y \in \mathrm{int}(\mathcal{K})$ and $z \in \mathrm{dom}(\Phi)$, the following three-point identity holds:*

$$\langle \nabla_x H(x, y), z - x \rangle = H(z, y) - H(z, x) - H(x, y).$$

(g) *If $\{y_k\} \subset \mathrm{int}(\mathcal{K})$ is such that $\lim_{k \to +\infty} y_k = \bar{y} \in \mathrm{bd}(\mathcal{K})$, then*

$$\lim_{k \to +\infty} \langle \nabla \Phi(y_k), x - y_k \rangle = -\infty, \quad \forall x \in \mathrm{int}(\mathcal{K}). \tag{10}$$

Proof. Statements (b), (c) and (f) follow directly from Lemma 3.1, Part (a), the strictly convex of the function Φ and the definition of H, respectively.

(a) Since $\phi^{\mathrm{sc}}(x)$, $(\phi')^{\mathrm{sc}}(y)$, $(\phi')^{\mathrm{sc}}(y) \circ (x - y)$ are continuous for any $x \in \mathrm{dom}(\phi^{\mathrm{sc}})$ (by [24, Theorem 3.2]), $y \in \mathrm{int}(\mathcal{K})$, and the trace function is also continuous, it follows that the function H is the same on $\mathrm{dom}(\Phi) \times \mathrm{int}(\mathcal{K})$. On the other hand, by using (S.1) and [23, Theorem 41] one has that the function Φ is strictly convex on $\mathrm{dom}(\Phi)$. Moreover, the expression $-\Phi(y) - \langle \nabla \Phi(y), . - y \rangle$ is clearly convex on $\mathcal{K}$, for any fixed $y \in \mathrm{int}(\mathcal{K})$. Hence, $H(\cdot, y)$ is strictly convex for any fixed $y \in \mathrm{int}(\mathcal{K})$.

(d) Let $y = \sum_{i=1}^{r} \lambda_i(y) e_i(y)$, and $x = \sum_{i=1}^{r} \lambda_i(x) e_i(x)$ be the spectral decomposition of y, and x, respectively. Using the definition of H, we have for any $x \in \mathrm{dom}(\Phi)$ and $y \in \mathrm{int}(\mathcal{K})$,

$$
\begin{aligned}
H(x, y) &= \Phi(x) - \Phi(y) - \mathrm{tr}(\nabla \Phi(y) \circ x) + \mathrm{tr}(\nabla \Phi(y) \circ y) \\
&\geq \Phi(x) - \Phi(y) - \sum_{i=1}^{r} \phi'(\lambda_i(y)) \lambda_i(x) + \mathrm{tr}(\nabla \Phi(y) \circ y) \\
&= \sum_{i=1}^{r} [\phi(\lambda_i(x)) - \phi(\lambda_i(y)) - \phi'(\lambda_i(y)) \lambda_i(x) + \phi'(\lambda_i(y)) \lambda_i(y)] \\
&= \sum_{i=1}^{r} d_\phi(\lambda_i(x), \lambda_i(y)),
\end{aligned}
$$

where the first equality is due to (3), the inequality follows from (6) and *Von Neumann inequality* [23, Theorem 23], the second equality is due to (6), (5), and (2), and the last one follows from (8). The nonnegativity of $d_\phi(t, s)$ is due to the strict convexity of ϕ on its domain.

(e) For any fixed $y \in \mathrm{int}(\mathcal{K})$ and $\gamma \geq 0$, from Part (d) we obtain $L_H(y, \gamma) \subseteq \{x \in \mathrm{dom}(\Phi) : \sum_{i=1}^{r} d_\phi(\lambda_i(x), \lambda_i(y)) \leq \gamma\}$. By (S.3) it follows that the set in the right-hand side is bounded. Then, $L_H(y, \gamma)$ is bounded for all $\gamma \geq 0$. Similarly, for fixed $x \in \mathrm{dom}(\Phi)$, the sets $L_H(x, \gamma)$ are bounded for all $\gamma \geq 0$.

(g) Let $y_k = \sum_{i=1}^{r} \lambda_i(y_k) e_i(y_k)$, $\bar{y} = \sum_{i=1}^{r} \lambda_i(\bar{y}) e_i(\bar{y})$, and $x = \sum_{i=1}^{r} \lambda_i(x) e_i(x)$ be the spectral decompositions of y_k, $\bar{y}$, and x, respectively. By using (2), (3), (6), and *Von Neumann inequality* [23, Theorem 23] we get

$$\langle \nabla \Phi(y_k), x - y_k \rangle \leq \sum_{i=1}^{r} \phi'(\lambda_i(y_k))(\lambda_i(x) - \lambda_i(y_k)), \quad \forall x \in \mathrm{int}(\mathcal{K}). \tag{11}$$

7

On the other hand, note that $x \in \text{int}(\mathcal{K})$, thus $\lambda_i(x) > 0$, for all $i = 1, \ldots, r$. Moreover, as $\bar{y} \in \text{bd}(\mathcal{K})$, there exists an $l \in \{1, \ldots, r\}$ such that $\lambda_i(\bar{y}) = 0$, for all $i = 1, \ldots, l$, and $\lambda_i(\bar{y}) > 0$, for all $i = l+1, \ldots, r$. Then, from (S.4) and the continuity of λ_i, one has that

$$\lim_{k \to +\infty} \phi'(\lambda_i(y_k))(\lambda_i(x) - \lambda_i(y_k)) = -\infty, \quad \forall i = 1, \ldots, l, \tag{12}$$

and that the expression $\lim_{k \to +\infty} \phi'(\lambda_i(y_k))(\lambda_i(x) - \lambda_i(y_k))$ is finite, for all $i = l+1, \ldots, r$. Hence, the result follows from (11) and (12). $\qquad \square$

This result shows that the function H defined by (9), with $\phi \in \Sigma$, is a proximal distance with respect to $\text{int}(\mathcal{K})$ and moreover, it satisfies the conditions $(B1)$-$(B2)$. Nevertheless, we should point out that the proximal distance H, with $\phi \in \Sigma$ and $\text{dom}(\phi) = \mathbb{R}_+$, generally does not satisfy $(B4)$ (see [25, 26] and Example below), even if ϕ satisfies the following condition:

(S.5) For any $\{t_k\} \subset \mathbb{R}_{++}$ such that $\lim_{k \to +\infty} = t^* \in \mathbb{R}_+$, one has $\lim_{k \to +\infty} d_\phi(t^*, t_k) = 0$.

Remark 3.1. *It follows from Proposition 3.2, Part (g) that if $\{y_k\} \subset \text{int}(\mathcal{K})$ is a sequence that convergent to a point on the boundary of $\mathcal{K}$, then $\lim_{k \to +\infty} H(x, y_k) = +\infty$, $\forall x \in \text{int}(\mathcal{K})$, that is, Φ is essentially smooth [20]. Moreover, if we suppose that H, defined by (9), satisfies the condition $(B4)$, by using the ideas of [18], one can to show that the condition $(B3)$ holds.*

Now, we present some examples that satisfies Definition 3.1.

Example 3.1. (Entropy-like proximal distance)
Let $\phi(t) = t \ln(t)$, if $t \geq 0$ (with the convention $0 \ln(0) = 0$), and $\phi(t) = +\infty$, if $t < 0$. It is easy to check that $\phi \in \Sigma$ with $\text{dom}(\phi) = \mathbb{R}_+$. Its spectrally defined function associated is $\Phi(x) = \text{tr}(x \circ \ln(x))$, if $x \in \mathcal{K}$. By [25, Lemma 3.2], $\nabla \Phi(x) = \ln(x) + e$, for $x \in \text{int}(\mathcal{K})$. Then, the function $H : \mathbb{V} \times \mathbb{V} \to (-\infty, +\infty]$ is given by

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

It follows from Proposition 3.2 (see also [25, Proposition 3.1 and 3.3]) that H is a proximal distance and that it satisfies $(B1)$ and $(B2)$ with $\mathcal{C}_1 = \mathcal{K}$ and $\mathcal{C}_2 = \text{int}(\mathcal{K})$.

(a) *Note that when $\mathbb{V} = \mathbb{R}^n$, $\mathcal{K} = \mathbb{R}^n_+$, the function H has the form*

$$H(x, y) = \begin{cases} \sum_{i=1}^n (x_i \ln(x_i/y_i) + y_i - x_i) & , \quad \forall x \in \mathbb{R}^n_+, y \in \mathbb{R}^n_{++}, \\ +\infty & , \quad \text{otherwise,} \end{cases}$$

which is is the so-called Kullback-Leibler relative entropy distance [27, 28]. This distance satisfies the condition $(B4)$ (see [28, Lemma 2.1]), then it satisfies $(B3)$ (cf. Remark 3.1).

(b) *For $\mathbb{V} = \mathbb{R}^n$, $\mathcal{K} = \mathcal{L}^n_+$, the function H has the form*

$$H(x, y) = \begin{cases} \text{tr}(x \circ \ln(x) - x \circ \ln(y) + y - x) & , \quad \forall x \in \mathcal{L}^n_+, y \in \mathcal{L}^n_{++}, \\ +\infty & , \quad \text{otherwise.} \end{cases}$$

Opposite to the previous case, this function does not satisfy the property $(B4)$. This fact is illustrated in [19, Example 4.7].

8

(c) *For $\mathbb{V} = \mathcal{S}^n$, $\mathcal{K} = \mathcal{S}_+^n$, the function H has the form*

$$H(X, Y) = \begin{cases} \mathrm{tr}(X \ln(X - Y) + Y - X) & , \quad \forall X \in \mathcal{S}_+^n, Y \in \mathcal{S}_{++}^n, \\ +\infty & , \quad otherwise. \end{cases}$$

As noticed in a counterexample given by Doljansky and Teboulle [26], property (B4) does not hold for this function.

Example 3.2. *For $p \in (0, 1)$, let us consider the family of functions $\phi(t) = (pt - t^p)/(1 - p)$ if $t \geq 0$, and $\phi(t) = +\infty$ if, $t < 0$ ([29]). It is not hard to verify that $\phi \in \Sigma$ with $\mathrm{dom}(\phi) = \mathbb{R}_+$. Then, its spectrally defined function associated is $\Phi(x) = \frac{1}{1-p}\mathrm{tr}(px - x^p)$, for $x \in \mathcal{K}$. By Lemma 3.1 it follows that $\nabla\Phi(x) = \frac{p}{1-p}(e - x^{p-1})$, for $x \in \mathrm{int}(\mathcal{K})$. Thus, we have that*

$$H(x, y) = \begin{cases} \frac{1}{1-p}\mathrm{tr}((1 - p)y^p + (py^{p-1} - x^{p-1}) \circ x) & , \quad \forall x \in \mathcal{K}, y \in \mathrm{int}(\mathcal{K}), \\ +\infty & , \quad otherwise. \end{cases}$$

From Proposition 3.2 we obtain that H is a proximal distance satisfying (B1)-(B2) with $\mathcal{C}_1 = \mathcal{K}$ and $\mathcal{C}_2 = \mathrm{int}(\mathcal{K})$.

Note that when $\mathbb{V} = \mathbb{R}^n$ and $\mathcal{K} = \mathbb{R}_+^n$, the proximal distance H takes the form

$$H(x, y) = \begin{cases} \sum_{i=1}^n (y_i^{p-1}(y_i + qx_i) - (1 + q)x_i^p) & , \quad \forall x \in \mathbb{R}_+^n, y \in \mathbb{R}_{++}^n, \\ +\infty & , \quad otherwise, \end{cases}$$

where $q = \frac{p}{1-p}$, and satisfies the condition (B4) (see [2, Example 3.1]).

Example 3.3. (Log-barrier proximal distance)
Let $\phi(t) = -\ln(t)$, for $t > 0$. It is easy to verify that $\phi \in \Sigma$ with $\mathrm{dom}(\phi) = \mathbb{R}_{++}$, and that its spectrally defined function associated is $\Phi(x) = -\mathrm{tr}(\ln(x)) = -\ln(\det(x))$, for $x \in \mathrm{int}(\mathcal{K})$. By Lemma 3.1 it follows that $\nabla\Phi(x) = -x^{-1}$, for $x \in \mathrm{int}(\mathcal{K})$. Then, one has that

$$H(x, y) = \begin{cases} \mathrm{tr}(\ln(y) - \ln(x) + y^{-1} \circ x) - r & , \quad \forall x, y \in \mathrm{int}(\mathcal{K}), \\ +\infty & , \quad otherwise. \end{cases}$$

This function extends the log-barrier proximal distance given in [2, 30, 26] to our context. Moreover, by using Proposition 3.2 one has that H is a proximal distance satisfying (B1)-(B2) with $\mathcal{C}_1 = \mathcal{C}_2 = \mathrm{int}(\mathcal{K})$.

Secondly, let us consider $\psi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ a closed proper convex function with $\mathrm{dom}(\psi) = \mathbb{R}_+$, satisfying the following conditions [31, 19]:

(C.1) ψ is continuous and strictly convex on $\mathbb{R}_+$, and it is continuously differentiable on a subset of $\mathrm{dom}(\psi)$, where $\mathrm{dom}(\psi') \subseteq \mathrm{dom}(\psi)$ and $\mathrm{int}(\mathrm{dom}(\psi')) = \mathbb{R}_{++}$.

(C.2) ψ is twice continuously differentiable on $\mathbb{R}_{++}$ and $\lim_{t \to 0^+} \psi''(t) = +\infty$.

(C.3) $\xi(t) = \psi'(t)t - \psi(t)$ is convex on $\mathrm{dom}(\psi')$, and ψ' is strictly concave on $\mathrm{dom}(\psi')$.

(C.4) $(\psi')^{\mathrm{sc}}$ is concave with respect to $\mathcal{K}$, that is,

$$(\psi')^{\mathrm{sc}}(\beta x + (1 - \beta)y) - \beta(\psi')^{\mathrm{sc}}(x) - (1 - \beta)(\psi')^{\mathrm{sc}}(y) \in \mathcal{K}, \quad \forall x, y \in \mathrm{dom}((\psi')^{\mathrm{sc}}) \text{ and } \beta \in [0, 1].$$

In the sequel, we denote by $\hat{\Sigma}$ the class of functions ψ satisfying the above assumptions. We define the function $\hat{H} : \mathbb{V} \times \mathbb{V} \to \mathbb{R} \cup \{+\infty\}$ by

$$\hat{H}(x,y) := \begin{cases} \Psi(y) - \Psi(x) - \langle \nabla\Psi(x), y - x \rangle & , \quad x \in \text{dom}((\psi')^{\text{sc}}), \ y \in \mathcal{K}, \\ +\infty & , \quad \text{otherwise.} \end{cases} \tag{13}$$

The following result extends [31, Lemma 4.2] to our context.

Lemma 3.3. *Let $\psi \in \hat{\Sigma}$. Then*

(a) *The function $\Xi(x) = \langle \nabla\Psi(x), x \rangle - \Psi(x) = \text{tr}(\xi^{\text{sc}}(x))$, with $\xi^{\text{sc}}(x) = (\psi')^{\text{sc}}(x) \circ x - (\psi)^{\text{sc}}(x)$, is convex in $\text{dom}((\psi')^{\text{sc}})$ and continuously differentiable on $\text{int}(\mathcal{K})$.*

(b) *For any fixed $y \in \mathbb{V}$, the function $\rho_y(x) = \langle \nabla\Psi(x), y \rangle = \langle (\psi')^{\text{sc}}(x), y \rangle$ is continuously differentiable on $\text{int}(\mathcal{K})$ with $\nabla\rho_y(x) = \nabla(\psi')^{\text{sc}}(x)y$, and moreover, it is strictly concave over $\text{dom}((\psi')^{\text{sc}})$ whenever $y \in \text{int}(\mathcal{K})$.*

Proof. (a) This part follows by (C.2)-(C.3), and [23, Corollary 39, Theorem 41].

(b) The first statement follows from (C.2) and [24, Theorem 3.2]. Now, we prove the second one. By (C.3) and [23, Theorem 41] we obtain that the function $\text{tr}((\psi')^{\text{sc}}(x))$ is strictly concave on $\text{dom}((\psi')^{\text{sc}})$, that is, for any $x, z \in \text{dom}((\psi')^{\text{sc}})$ with $x \neq z$, and $\beta \in (0,1)$ one has that

$$\text{tr}((\psi')^{\text{sc}}(\beta x + (1-\beta)z)) > \beta\text{tr}((\psi')^{\text{sc}}(x)) + (1-\beta)\text{tr}((\psi')^{\text{sc}}(z)),$$

whence $(\psi')^{\text{sc}}(\beta x + (1-\beta)z) - \beta(\psi')^{\text{sc}}(x) - (1-\beta)(\psi')^{\text{sc}}(z) \neq 0$. Then, from this relation, (C.4) and [8, Theorem 5.13] we have that

$$\langle (\psi')^{\text{sc}}(\beta x + (1-\beta)z) - \beta(\psi')^{\text{sc}}(x) - (1-\beta)(\psi')^{\text{sc}}(z), y \rangle > 0, \quad \text{for any } y \in \text{int}(\mathcal{K}).$$

This shows that the function $\rho_y(x)$ is strictly concave over $\text{dom}((\psi')^{\text{sc}})$, for any fixed $y \in \text{int}(\mathcal{K})$. $\square$

The next results present some properties of the function $\hat{H}$. The proof of the first result can be obtained by using the conditions (C.1)-(C.4), Lemma 3.1, Lemma 3.3 and, following the same arguments as for [31, Proposition 4.1].

Proposition 3.4. *Let $\psi \in \hat{\Sigma}$ and $\hat{H}$ be defined by (13). Then*

(a) *$\hat{H}(x,y) \geq 0$ for any $(x,y) \in \text{dom}((\psi')^{\text{sc}}) \times \mathcal{K}$, and $\hat{H}(x,y) = 0$ iff $x = y$.*

(b) *$\hat{H}(\cdot,\cdot)$ is continuous on $\text{dom}((\psi')^{\text{sc}}) \times \mathcal{K}$ and, for any fixed $y \in \text{int}(\mathcal{K})$, the function $\hat{H}(\cdot,y)$ is strictly convex on $\text{dom}((\psi')^{\text{sc}})$.*

(c) *For any fixed $y \in \mathcal{K}$, $\hat{H}(\cdot,y)$ is continuously differentiable on $\text{int}(\mathcal{K})$ with*

$$\nabla_1\hat{H}(x,y) = \nabla(\psi')^{\text{sc}}(x)(x-y). \tag{14}$$

(d) *$\hat{H}(x,y) \geq \sum_{i=1}^{r} d_\psi(\lambda_i(y), \lambda_i(x)) \geq 0$, for any $(x,y) \in \text{dom}((\psi')^{\text{sc}}) \times \mathcal{K}$, where d_ψ is defined in (8).*

(e) *For all $\gamma \geq 0$, the level sets $L_{\hat{H}}(y, \gamma) = \{x \in \text{dom}((\psi')^{\text{sc}}) : \hat{H}(x,y) \leq \gamma\}$ and $L_{\hat{H}}(x, \gamma) = \{y \in \mathcal{K} : \hat{H}(x,y) \leq \gamma\}$ are bounded, for any fixed $y \in \mathcal{K}$ and $x \in \text{dom}((\psi')^{\text{sc}})$, respectively.*

Proposition 3.5. *Let $\psi \in \hat{\Sigma}$ and $\hat{H}$ be defined by (13). Then,*

(a) *for all $x, y \in \text{int}(\mathcal{K})$ and $z \in \mathcal{K}$, we have*

$$\langle \nabla_1 \hat{H}(y, x), z - y \rangle \leq \hat{H}(x, z) - \hat{H}(y, z). \tag{15}$$

(b) *if $\text{dom}(\psi) = \text{dom}(\psi') = \mathbb{R}_+$, $\hat{H}$ satisfies condition (B4').*

Proof. (a) From the definition of $\hat{H}$, Ξ and $\rho_z(\cdot)$, we have for any $x, y \in \text{int}(\mathcal{K})$ and $z \in \mathcal{K}$

$$
\begin{aligned}
\hat{H}(x, z) - \hat{H}(y, z) &= \Xi(x) - \Xi(y) + \rho_z(y) - \rho_z(x) \\
&\geq \langle \nabla \Xi(y), x - y \rangle + \langle \nabla \rho_z(y), y - x \rangle \\
&= \langle \nabla \Xi(y) - \nabla (\psi')^{\text{sc}}(y) z, x - y \rangle,
\end{aligned}
\tag{16}
$$

where the inequality follows from the convexity of Ξ and the strictly concave of ρ_z (see Lemma 3.3), and the last equality is due to Lemma 3.3, Part(b). On the other hand, first, we note that from Lemma 3.3, Part(a), Lemma 3.1, Part(a), (4) and the definition of ξ, one has that $\nabla \Xi(y) = (\psi'')^{\text{sc}}(y) \circ y$. Second, by Lemma 3.1, Part(b) we obtain that $\nabla (\psi')^{\text{sc}}(y) y = (\psi'')^{\text{sc}}(y) \circ y$. Then,

$$\nabla (\psi')^{\text{sc}}(y) z = \nabla (\psi')^{\text{sc}}(y)(z - y) + \nabla (\psi')^{\text{sc}}(y) y = \nabla (\psi')^{\text{sc}}(y)(z - y) + \nabla \Xi(y).$$

Using this relation in (16) we have

$$\hat{H}(x, z) - \hat{H}(y, z) \geq \langle \nabla (\psi')^{\text{sc}}(y)(y - z), x - y \rangle = \langle \nabla (\psi')^{\text{sc}}(y)(y - x), z - y \rangle,$$

where the equality is due to symmetry of $\nabla (\psi')^{\text{sc}}(y)$. The result follows by using (14).

(b) This part follows from the continuity of $\hat{H}(\cdot, v^*)$ on $\mathcal{K}$. $\qquad\square$

Example 3.4. *Let $\psi(t) = \frac{t^{q+1}}{q+1}$, if $t \geq 0$, and $\phi(t) = +\infty$, if $t < 0$, where $q \in (0, 1)$. From [19, Example 4.9] it follows that ψ satisfies (C.1)-(C.3) with $\text{dom}(\psi) = \text{dom}(\psi') = \mathbb{R}_+$. Since, $\Psi(x) = \frac{1}{q+1} \text{tr}(x^{q+1})$ and $\nabla \Psi(x) = (\psi')^{\text{sc}}(x) = x^q$, for $x \in \mathcal{K}$, by using the same arguments of [32, Proposition 3.7], one can prove that (C.4) holds. Then, the function $\hat{H}$ given by*

$$\hat{H}(x, y) = \begin{cases} \frac{1}{q+1} \text{tr}(y^{q+1}) + \frac{q}{q+1} \text{tr}(x^{q+1}) - \text{tr}(x^q \circ y) & , \quad \forall x, y \in \mathcal{K}, \\ +\infty & , \quad otherwise, \end{cases}$$

is a proximal distance with $\mathcal{C}_1 = \mathcal{C}_2 = \mathcal{K}$ and satisfies (B1'), (B2) and (B4') (cf. Propositions 3.4 and 3.5).

4. Inexact algorithm with proximal generalized distance

In this section, we introduce an inexact proximal algorithm with generalized distance in order to solve the problem (P), that is,

$$(P) \qquad \min\{f(x) + g(z) : \mathbb{A}x + \mathbb{B}z = b, x \in \mathcal{K}_1, z \in \mathcal{K}_2\},$$

where $f : \mathbb{V}_1 \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{V}_2 \to \mathbb{R} \cup \{+\infty\}$ are two closed proper convex functions defined on the Euclidean Jordan algebra $\mathbb{V}_1 = (\mathbb{V}_1, \circ_1, \langle \cdot, \cdot \rangle_1)$ and $\mathbb{V}_2 = (\mathbb{V}_2, \circ_2, \langle \cdot, \cdot \rangle_2)$, respectively, $\mathbb{A} : \mathbb{V}_1 \to \mathbb{R}^m$ and $\mathbb{B} : \mathbb{V}_2 \to \mathbb{R}^m$ are two linear mapping, $b \in \mathbb{R}^m$ and $\mathcal{K}_1 := \{x \circ_1 x : x \in \mathbb{V}_1\}$ and $\mathcal{K}_2 := \{z \circ_2 z : z \in \mathbb{V}_2\}$ denoting the sets of square elements in $\mathbb{V}_1$ and $\mathbb{V}_2$, respectively.

Remark 4.1. *Note that for $\mathbb{V} = \mathbb{R}^n$, the linear application $\mathbb{A} : \mathbb{R}^n \to \mathbb{R}^m$ is defined as $\mathbb{A}x = Ax$ with $A \in \mathbb{R}^{m \times n}$, while for the case $\mathbb{V} = \mathcal{S}^n$ the linear application $\mathbb{A} : \mathcal{S}^n \to \mathbb{R}^m$ is given by $\mathbb{A}X = (\langle A_1, X \rangle, \ldots, \langle A_m, X \rangle)$, where $A_i \in \mathcal{S}^n$ for $i = 1, \ldots, m$.*

A class of problems of (P) have drawn recently a lot of attentions due to their emerging applications of practical interest, for example:

1. **Support vector machines (SVM) for binary classification :** Given a set of instances with their respective labels $(\mathbf{x}_i, y_i)$, where $\mathbf{x}_i \in \mathbb{R}^n$, $i = 1, \ldots, m$ and $y_i \in \{-1, +1\}$, the SVM determines an optimal hyperplane, namely $H(\mathbf{w}, b) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{w}^\top \mathbf{x} + b = 0\}$, that separates the reduced convex hulls of both training patterns. This optimal hyperplane can be found by solving the following Quadratic Programming problem [33]:

$$
\begin{aligned}
\min_{\mathbf{w}, b, \boldsymbol{\xi}} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i \\
\text{s.t.} \quad & y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \qquad i = 1, \ldots, m, \\
& \xi_i \geq 0, \qquad i = 1, \ldots, m,
\end{aligned}
\tag{17}
$$

   where $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m) \in \mathbb{R}^m$ and $C > 0$ is a penalty parameter.
   Let us denote by $\mathbf{z} = (\mathbf{w}, b) \in \mathbb{R}^{n+1}$, by $\text{Diag}(\mathbf{y})$ a diagonal matrix which main diagonal is the vector $\mathbf{y} = (y_1, \ldots, y_m)$, by $\mathbf{1}$ a vector of ones in $\mathbb{R}^m$, and by

$$
\hat{X} = \begin{pmatrix} \mathbf{x}_1^\top & 1 \\ \vdots & \vdots \\ \mathbf{x}_m^\top & 1 \end{pmatrix} \in \mathbb{R}^{m \times n+1},
$$

   then the first inequality constraint can be written as $\text{Diag}(\mathbf{y})\hat{X}\mathbf{z} + \boldsymbol{\xi} - \mathbf{1} \geq 0$. By introducing the variables $\mathbf{u} = \text{Diag}(\mathbf{y})\hat{X}\mathbf{z} + \boldsymbol{\xi} - \mathbf{1}$ and $\mathbf{v} = (\boldsymbol{\xi}, \mathbf{u}) \in \mathbb{R}^{2m}$, the problem (17) can be rewritten as

$$
\min_{\mathbf{z}, \mathbf{v}} \{f(\mathbf{z}) + g(\mathbf{v}) : A\mathbf{z} + B\mathbf{v} = b, \ \mathbf{v} \geq 0\},
\tag{18}
$$

   where $f(\mathbf{z}) = \frac{1}{2}\|\mathbf{w}\|^2$, $g(\mathbf{v}) = C\mathbf{1}^\top \boldsymbol{\xi}$, $A = \text{Diag}(\mathbf{y})\hat{X} \in \mathbb{R}^{m \times n+1}$, $B = (I \ -I) \in \mathbb{R}^{m \times 2m}$, and $b = \mathbf{1}$. Hence, this model has the form of (P).

2. **Sparse inverse covariance selection (SICS)**: Gaussian Graphical models are of certain interesting in statistical learning [34, 35], and are modelled by the following convex optimization problem:

$$
\min_{X \in \mathcal{S}^n} \{\langle S, X \rangle - \ln \det(X) + \rho \|X\|_1 : X \in \mathcal{S}^n_+\}
\tag{19}
$$

   where $\rho > 0$, $S \in \mathcal{S}^n_+$ denotes the empirical covariance matrix, and $\|X\|_1$ the l_1-norm of the matrix X defined by $\|X\|_1 := \sum_{i=1}^n \sum_{j=1}^n |x_{ij}|$. Defining $f(X) = \langle S, X \rangle - \ln \det(X)$ and $g(X) = \rho\|X\|_1$, we obtain $\min_{X \in \mathcal{S}^n} \{f(X) + g(X) : X \in \mathcal{S}^n_+\}$, which can be rewritten as:

$$
\min_{X, Y \in \mathcal{S}^n} \{f(X) + g(Y) : X - Y = 0, \ X, Y \in \mathcal{S}^n_+\}.
\tag{20}
$$

   This model is a special case of (P) when $\mathbb{V}_1 = \mathbb{V}_2 = \mathcal{S}^n$, $\mathcal{K}_1 = \mathcal{K}_2 = \mathcal{S}^n_+$, and $\mathbb{A}, \mathbb{B} : \mathcal{S}^n \to \mathbb{R}^{\tilde{m}}$, where $\tilde{m} = \frac{1}{2}n(n+1)$, $\mathbb{A}(X) = \text{svec}(X)$, and $\mathbb{B}(Y) = -\text{svec}(Y)$ with $\text{svec}(X) = (x_{11}, \sqrt{2}x_{12}, \ldots, \sqrt{2}x_{1n}, x_{22}, \sqrt{2}x_{23}, \ldots, \sqrt{2}x_{2n}, \ldots, x_{nn})$ (see [36]).

4.1. Proximal Multiplier Algorithm

For each $i, = 1, 2$, let $H_i : \mathbb{V}_i \times \mathbb{V}_i \to \mathbb{R} \cup \{+\infty\}$, be a proximal distance, that is, $H_i \in \mathcal{D}(\text{int}(\mathcal{K}_i))$ and $\theta_i > 0$ a fixed parameter. We define

$$H_{\theta_1}(x_1, x_2) := H_1(x_1, x_2) + \frac{\theta_1}{2} \|x_1 - x_2\|^2. \tag{21}$$

$$H_{\theta_2}(z_1, z_2) := H_2(z_1, z_2) + \frac{\theta_2}{2} \|z_1 - z_2\|^2. \tag{22}$$

It is easy to show that for each $i = 1, 2$, H_{θ_i}, is also a proximal distance with respect to $\text{int}(\mathcal{K}_i)$, that is $H_{\theta_i} \in \mathcal{D}(\text{int}(\mathcal{K}_i))$.

The algorithm with proximal distance for solving the problem (P) is defined by:

Algorithm SC-PMA: Let $H_i \in \mathcal{D}(\text{int}(\mathcal{K}_i))$, $\theta_i > 0$, $i = 1, 2$, *tol* > 0 and $\{\varepsilon_k\}, \{\zeta_k\}, \{\lambda_k\}$ be sequences of positive scalars.

Step 0: Start with some initial point $w^0 = (x^0, z^0, y^0) \in \text{int}(\mathcal{K}_1) \times \text{int}(\mathcal{K}_2) \times \mathbb{R}^m$. Set $k = 0$.

Step 1: Compute
$$p^{k+1} = y^k + \lambda_k(\mathbb{A}x^k + \mathbb{B}z^k - b), \tag{23}$$

Step 2: Find $(x^{k+1}, z^{k+1}) \in \text{int}(\mathcal{K}_1) \times \text{int}(\mathcal{K}_2)$ and $(g_1^{k+1}, g_2^{k+1}) \in \mathbb{V}_1 \times \mathbb{V}_2$ such that

$$\begin{aligned} g_1^{k+1} &\in \partial_{\varepsilon_k} f(x^{k+1}) \\ g_1^{k+1} + \mathbb{A}^* p^{k+1} &+ \frac{1}{\lambda_k} \nabla_x H_{\theta_1}(x^{k+1}, x^k) = 0, \end{aligned} \tag{24}$$

$$\begin{aligned} g_2^{k+1} &\in \partial_{\zeta_k} g(z^{k+1}) \\ g_2^{k+1} + \mathbb{B}^* p^{k+1} &+ \frac{1}{\lambda_k} \nabla_x H_{\theta_2}(z^{k+1}, z^k) = 0. \end{aligned} \tag{25}$$

Step 3: Compute
$$y^{k+1} = y^k + \lambda_k(\mathbb{A}x^{k+1} + \mathbb{B}z^{k+1} - b), \tag{26}$$

Step 4: Set $w^{k+1} = (x^{k+1}, z^{k+1}, y^{k+1})$. If $\|w^{k+1} - w^k\| \le tol$, stop; otherwise replace k by $k + 1$ and go to Step 1.

The following result gives some estimates in proximal algorithms and its proof is similar to [2, Theorem 2.1] and [37, Lemma 2.2].

Lemma 4.1. *Let $F : \mathbb{V} \to \mathbb{R} \cup \{+\infty\}$ be a closed proper convex function, $H \in \mathcal{D}(\text{int}(\mathcal{K}))$ satisfying $(B1)$ or $(B1')$ and $\lambda_k > 0$. If $\{(u^k, g_1^k)\}$ and $\{(v^k, g_2^k)\}$ are two sequence satisfying*

$$\begin{aligned} g_1^{k+1} &\in \partial_{\varepsilon_k} F(u^{k+1}) \\ g_1^{k+1} + \lambda_k^{-1} \nabla_x H(u^{k+1}, u^k) &= 0, \end{aligned}$$

$$\begin{aligned} \tilde{g}_2^{k+1} &\in \partial_{\zeta_k} F(v^{k+1}) \\ \tilde{g}_2^{k+1} + \lambda_k^{-1}(v^{k+1} - v^k) &= 0, \end{aligned}$$

then for any $k \ge 0$, we get

(i) $\lambda_k(F(u^{k+1}) - F(u)) \le H(u, u^k) - H(u, u^{k+1}) + \lambda_k \varepsilon_k, \quad \forall u \in \mathcal{C}_1$, *if $(B1)$ holds;*

(ii) $\lambda_k(F(u^{k+1}) - F(u)) \le H(u^k, u) - H(u^{k+1}, u) + \lambda_k \varepsilon_k, \quad \forall u \in \mathcal{C}_2$, *if $(B1')$ holds;*

(iii) $2\lambda_k(F(v^{k+1}) - F(v)) \le \|v^k - v\|^2 - \|v^{k+1} - v\|^2 - \|v^{k+1} - v^k\|^2 + 2\lambda_k \zeta_k, \quad \forall z \in \mathbb{R}^n$.

13

4.2. Convergence analysis of algorithm SC-PMA

In this section, under appropriate assumptions, we will prove that the sequences generated by the algorithm SC-PMA are well-defined, and we will establish the global convergence of the algorithm proposed.

From now on, we suppose that the following assumptions hold true:

(A1) Problem (P) admits at least an optimal solution (x^*, z^*).

(A2) $\exists x \in \text{int}(\mathcal{K}_1) \cap \text{ri}(\text{dom}(f))$, $z \in \text{int}(\mathcal{K}_2) \cap \text{ri}(\text{dom}(g))$ such that $\mathbb{A}x + \mathbb{B}z = b$.

(A3) The sequences $\{\zeta_k\}$ and $\{\varepsilon_k\}$ are nonnegative, and $\sum_{k=0}^{\infty}(\zeta_k + \varepsilon_k) < \infty$.

Remark 4.2. *Assumption (A2) ensures the existence of an optimal dual Lagrange multiplier y^*. Hence, under the assumptions (A1)-(A2), (x^*, z^*, y^*) is a saddle point of L, that is,*

$$L(x^*, z^*, y) \leq L(x^*, z^*, y^*) \leq L(x, z, y^*), \quad \forall x \in \mathcal{K}_1 \cap dom(f), \; z \in \mathcal{K}_2 \cap dom(g), \; y \in \mathbb{R}^m,$$

where
$$L(x, z, y) = f(x) + g(z) + \langle y, \mathbb{A}x + \mathbb{B}z - b \rangle, \quad x \in \mathcal{K}_1, \; z \in \mathcal{K}_2, \; y \in \mathbb{R}^m.$$

denotes the Lagrangian function for problem (P).
Assumption (A3) is a natural condition when we work with inexact iterations.

The following result shows that the algorithm SC-PMA is well-defined.

Proposition 4.2. *Under assumptions (A1), (A2), $H_i \in \mathcal{D}(\text{int}(\mathcal{K}_i)), i = 1, 2$, and $(x^k, z^k, y^k) \in \text{int}(\mathcal{K}_1) \times \text{int}(\mathcal{K}_2) \times \mathbb{R}^m$, there exists a unique point $(x^{k+1}, z^{k+1}) \in \text{int}(\mathcal{K}_1) \times \text{int}(\mathcal{K}_2)$ satisfying (24)-(25).*

Proof. Set $F_k(x) = f(x) + \langle \mathbb{A}^* p^{k+1}, x \rangle + \frac{1}{\lambda_k} H_\theta(x, x^k) + \delta_{\mathcal{K}_1}(x)$, then from assumption (A2) we have that $\text{dom}(F_k) = \text{dom}(f) \cap \text{dom}(H_{\theta_1}(\cdot, x^k)) \cap \mathcal{K}_1 \neq \emptyset$. Define $S_k = \arg\min_{x \in \mathcal{K}_1}\{F_k(x)\}$, we will prove that the set S_k is not empty. In fact, because $f(\cdot)$, $\langle \mathbb{A}^* p^{k+1}, \cdot \rangle$ and $(1/\lambda_k)H_{\theta_1}(\cdot, x^k)$ are closed proper convex functions, then F_k is proper closed and convex too, thus, it suffices to show that, see [20, Theorem 27.1(d)], $(F_k)_\infty(d) > 0, \forall d \neq 0$, where $(F_k)_\infty$ denotes the recession function of F_k. Let $d \neq 0$, then we have that

$$(F_k)_\infty(d) = f_\infty(d) + (\langle \mathbb{A}^* p^{k+1}, \cdot \rangle)_\infty(d) + \frac{1}{\lambda_k}(H_{\theta_1})_\infty(d, x^k) + (\delta_{\mathcal{K}_1})_\infty(d). \tag{27}$$

It is easy to prove that

$$f_\infty(d) > 0, \quad (\delta_{\mathcal{K}_1})_\infty(d) \geq 0, \quad (\langle \mathbb{A}^* p^{k+1}, \cdot \rangle)_\infty(d) = \langle \mathbb{A}^* p^{k+1}, d \rangle,$$

then, we will prove that $(H_{\theta_1})_\infty(d, x^k) = +\infty$. Since $H_{\theta_1}(\cdot, z)$ is proper closed and convex one has

$$(H_{\theta_1})_\infty(d, x^k) = (H_1)_\infty(d, x^k) + \frac{\theta_1}{2}(\| \cdot - x^k \|^2)_\infty(d).$$

Due that H_1 is coercive, $(H_1)_\infty(d, x^k) > 0$ and as $(\| \cdot - x^k \|^2)_\infty(d) = +\infty$, we obtain that $(H_{\theta_1})_\infty(d, x^k) = +\infty$, and so from (27) we also obtain

$$(F_k)_\infty(d) = +\infty > 0, \quad \forall d \neq 0.$$

14

Hence, there exists $\bar{x} := x^{k+1} \in \mathbb{V}_1$ such that

$$0 \in \partial\left(f(\cdot) + \langle \mathbb{A}^* p^{k+1}, \cdot \rangle + \frac{1}{\lambda_k} H_{\theta_1}(\cdot, x^k) + \delta_{\mathcal{K}_1}(\cdot)\right)(x^{k+1}).$$

Also, the strict convexity of F_k (cf. Definition 3.1, Part (P2)) guarantees the uniqueness of x^{k+1}. Now by assumption **(A2)**, we have that

$$0 \in \partial f(x^{k+1}) + \mathbb{A}^* p^{k+1} + (1/\lambda_k)\partial_x H_{\theta_1}(., x^k)(x^{k+1}) + \partial\delta_{\mathcal{K}_1}(x^{k+1}),$$

From Definition, we have that $\mathrm{dom}(\partial_x H_{\theta_1}(\cdot, y)) = \mathrm{int}(\mathcal{K}_1)$, then $x^{k+1} \in \mathrm{int}(\mathcal{K}_1)$, hence $\partial\delta_{\mathcal{K}_1}(x^{k+1}) = \mathcal{N}_{\mathcal{K}_1}(x^{k+1}) = \{0\}$. Thus, there exists $g_1^{k+1} \in \partial f(x^{k+1})$ such that (24) holds with $\varepsilon_k = 0$. Since $\partial f(x^{k+1}) \subset \partial_{\varepsilon_k} f(x^{k+1})$, (24) holds for $\varepsilon_k > 0$.
The existence and uniqueness of z^{k+1} and (25) are shown in a similar form. $\square$

The next result establishes estimates for the sequences generated by the algorithm (SC-PMA).

Proposition 4.3. *Let $\{x^k, z^k, y^k, p^k\}$ be the sequence generated by algorithm (SC-PMA), (x^*, z^*) be an optimal solution of (P) and y^* be a corresponding Lagrange multiplier. Assume that $H_i \in \mathcal{D}(\mathrm{int}(\mathcal{K}_i))$, for $i = 1, 2$, and $(B1)$ or $(B1')$ hold. Then, the following inequalities hold for all $k \geq 0$*

$$
\begin{aligned}
H_{\theta_1}(x^*, x^{k+1}) + H_{\theta_2}(z^*, z^{k+1}) \leq \;& H_{\theta_1}(x^*, x^k) + H_{\theta_2}(z^*, z^k) - \lambda_k\langle p^{k+1} - y^*, \mathbb{A}x^{k+1} + \mathbb{B}z^{k+1} - b\rangle \\
& + \lambda_k(\zeta_k + \varepsilon_k), \qquad \text{if } (B1) \text{ holds;} \qquad (28)
\end{aligned}
$$

$$
\begin{aligned}
H_{\theta_1}(x^{k+1}, x^*) + H_{\theta_2}(z^{k+1}, z^*) \leq \;& H_{\theta_1}(x^k, x^*) + H_{\theta_2}(z^k, z^*) - \lambda_k\langle p^{k+1} - y^*, \mathbb{A}x^{k+1} + \mathbb{B}z^{k+1} - b\rangle \\
& + \lambda_k(\zeta_k + \varepsilon_k), \qquad \text{if } (B1') \text{ holds.} \qquad (29)
\end{aligned}
$$

Proof. Assume that $(B1)$ holds. Using Lemma 4.1, Part (i) with $F(\cdot) := f(\cdot) + \langle p^{k+1}, \mathbb{A}\cdot\rangle$ and with $F(\cdot) := g(\cdot) + \langle p^{k+1}, \mathbb{B}\cdot\rangle$ we have that, for all $x \in \mathcal{C}_1$:

$$\lambda_k(f(x^{k+1}) + \langle p^{k+1}, \mathbb{A}x^{k+1}\rangle - f(x) - \langle p^{k+1}, \mathbb{A}x\rangle) \leq H_{\theta_1}(x, x^k) - H_{\theta_1}(x, x^{k+1}) + \lambda_k\varepsilon_k,$$

and for all $z \in \mathcal{C}_1$:

$$\lambda_k(g(z^{k+1}) + \langle p^{k+1}, \mathbb{B}z^{k+1}\rangle - g(z) - \langle p^{k+1}, \mathbb{B}z\rangle) \leq H_{\theta_2}(z, z^k) - H_{\theta_2}(z, z^{k+1}) + \lambda_k\zeta_k.$$

Adding the inequalities above, we obtain

$$
\begin{aligned}
\lambda_k(L(x^{k+1}, z^{k+1}, p^{k+1}) - L(x, z, p^{k+1})) \leq \;& H_{\theta_1}(x, x^k) - H_{\theta_1}(x, x^{k+1}) + \\
& H_{\theta_2}(z, z^k) - H_{\theta_2}(z, z^{k+1}) + \lambda_k(\zeta_k + \varepsilon_k). \quad (30)
\end{aligned}
$$

On the other hand, since (x^*, z^*, y^*) is saddle point of L we have

$$\lambda_k(L(x^*, z^*, p^{k+1}) - L(x^{k+1}, z^{k+1}, y^*)) \leq 0.$$

Using (30) with $x = x^*$, $z = z^*$ and adding with the above inequality we obtain (28) after rearranging terms. The inequality (29) it follows by using the arguments above and the property $(B1')$. $\square$

15

Proposition 4.4. *Let $\{x^k, z^k, y^k, p^k\}$ be the sequence generated by algorithm (SC-PMA), (x^*, z^*) be an optimal solution of (P) and y^* be a corresponding Lagrange multiplier. Then, the following inequalities hold for all $k \geq 0$*

$$\lambda_k \langle \mathbb{A}x^{k+1} + \mathbb{B}z^{k+1} - b, y^* - y^{k+1} \rangle \leq \frac{1}{2} \left(\|y^k - y^*\|^2 - \|y^{k+1} - y^*\|^2 - \|y^{k+1} - y^k\|^2 \right), \quad (31)$$

$$\lambda_k \langle \mathbb{A}x^k + \mathbb{B}z^k - b, y^{k+1} - p^{k+1} \rangle \leq \frac{1}{2} \left(\|y^k - y^{k+1}\|^2 - \|p^{k+1} - y^{k+1}\|^2 - \|p^{k+1} - y^k\|^2 \right). \quad (32)$$

Proof. The inequalities (31)-(32) follow directly from [14, Proposition 2] or [15, Lemma 4.1]. $\square$

For any vector $w_1 = (x_1, z_1, y_1) \in \mathcal{C}_1 \times \mathcal{C}_1 \times \mathbb{R}^m$ and $w_2 = (x_2, z_2, y_2) \in \mathcal{C}_2 \times \mathcal{C}_2 \times \mathbb{R}^m$ we define

$$\hat{H}_\theta(w_1, w_2) = H_{\theta_1}(x_1, x_2) + H_{\theta_2}(z_1, z_2) + \frac{1}{2}\|y_1 - y_2\|^2. \quad (33)$$

Lemma 4.5. *Let $\{x^k, z^k, y^k, p^k\}$ be the sequence generated by algorithm (SC-PMA) and let $w^* = (x^*, z^*, y^*)$ with (x^*, z^*) an optimal solution of (P) and y^* its corresponding Lagrange multiplier. Assume that $H_i \in \mathcal{D}(\mathrm{int}(\mathcal{K}_i))$, for $i = 1, 2$. If $(B1)$ holds, then*

$$
\begin{aligned}
\hat{H}_\theta(w^*, w^{k+1}) \leq \; & \hat{H}_\theta(w^*, w^k) - \frac{1}{2}(\theta_1 - 4\lambda_k^2\|\mathbb{A}\|^2)\|x^{k+1} - x^k\|^2 - \frac{1}{2}(\theta_2 - 4\lambda_k^2\|\mathbb{B}\|^2)\|z^{k+1} - z^k\|^2 \\
& - \frac{1}{2}\|p^{k+1} - y^{k+1}\|^2 - \frac{1}{2}\|p^{k+1} - y^k\|^2 + \lambda_k(\zeta_k + \varepsilon_k);
\end{aligned} \quad (34)
$$

and if $(B1')$ holds, then

$$
\begin{aligned}
\hat{H}_\theta(w^{k+1}, w^*) \leq \; & \hat{H}_\theta(w^k, w^*) - \frac{1}{2}(\theta_1 - 4\lambda_k^2\|\mathbb{A}\|^2)\|x^{k+1} - x^k\|^2 - \frac{1}{2}(\theta_2 - 4\lambda_k^2\|\mathbb{B}\|^2)\|z^{k+1} - z^k\|^2 \\
& - \frac{1}{2}\|p^{k+1} - y^{k+1}\|^2 - \frac{1}{2}\|p^{k+1} - y^k\|^2 + \lambda_k(\zeta_k + \varepsilon_k).
\end{aligned} \quad (35)
$$

Proof. Suppose that $(B1)$ holds. Adding the inequalities (31)-(32), we obtain

$$
\begin{aligned}
\frac{1}{2}\|y^{k+1} - y^*\|^2 \leq \; & \frac{1}{2}\|y^k - y^*\|^2 - \frac{1}{2}\left(\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2 \right) \\
& - \lambda_k \left(\langle \mathbb{A}x^k + \mathbb{B}z^k - b, y^{k+1} - p^{k+1} \rangle + \langle \mathbb{A}x^{k+1} + \mathbb{B}z^{k+1} - b, y^* - y^{k+1} \rangle \right)
\end{aligned}
$$

Let $w^{k+1} = (x^{k+1}, z^{k+1}, y^{k+1})$ and $w^k = (x^k, z^k, y^k)$. Then, adding (28) with the above inequality we get

$$
\begin{aligned}
\hat{H}_\theta(w^*, w^{k+1}) \leq \; & \hat{H}_\theta(w^*, w^k) - \frac{\theta_1}{2}\|x^{k+1} - x^k\|^2 - \frac{\theta_2}{2}\|z^{k+1} - z^k\|^2 \\
& - \frac{1}{2}\left(\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2 \right) + \rho_k + \lambda_k(\zeta_k + \varepsilon_k), \quad (36)
\end{aligned}
$$

where $\rho_k = \lambda_k \langle y^{k+1} - p^{k+1}, \mathbb{A}(x^{k+1} - x^k) + \mathbb{B}(z^{k+1} - z^k) \rangle$. Now, by using (23) and (26) it follows that

$$\rho_k = \lambda_k^2 \|\mathbb{A}(x^{k+1} - x^k) + \mathbb{B}(z^{k+1} - z^k)\|^2 \leq 2\lambda_k^2(\|\mathbb{A}\|^2\|x^{k+1} - x^k\|^2 + \|\mathbb{B}\|^2\|z^{k+1} - z^k\|^2).$$

Hence, by using this inequality in (36) the result follows. The inequality (35) it follows by using the same arguments above and the property $(B1')$. $\square$

Theorem 4.6. *Let $\{x^k, z^k, y^k, p^k\}$ be the sequence generated by algorithm (SC-PMA) and let $w^* = (x^*, z^*, y^*)$ with (x^*, z^*) an optimal solution of (P) and y^* its corresponding Lagrange multiplier. Assume that $H_i \in \mathcal{D}(\text{int}(\mathcal{K}_i))$, for $i = 1, 2$, and $(B1)$-$(B3)$ or $(B1')$,$(B3')$ hold. If $\{\lambda_k\}$ satisfies*

$$\lambda_k \geq \eta, \quad \lambda_k \|\mathbb{A}\| \leq \frac{1}{2}(\theta_1 - \vartheta)^{1/2}, \quad \lambda_k \|\mathbb{B}\| \leq \frac{1}{2}(\theta_2 - \vartheta)^{1/2}, \ \forall k \geq 0, \tag{37}$$

for some $\eta > 0$ and $0 < \vartheta < \min\{\theta_1, \theta_2\}$, then the following hold:

(i) *The sequence $w^k = (x^k, z^k, y^k)$ is bounded and every limit point of w^k is a saddle point of the Lagrangian.*

(ii) *Furthermore, if $(B4)$ or $(B4')$ hold then the sequence $\{(x^k, z^k, y^k)\}$ globally converges to a solution of problem (P).*

Proof. (i) Assume that $(B1)$-$(B2)$ hold. Since λ_k satisfies (37), from (34) we have that

$$\begin{aligned}
\hat{H}_\theta(w^*, w^{k+1}) &\leq \hat{H}_\theta(w^*, w^k) - \frac{\vartheta}{2}\left(\|x^{k+1} - x^k\|^2 + \|z^{k+1} - z^k\|^2\right) \\
&\quad - \frac{1}{2}\left(\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\right) + \lambda_k(\zeta_k + \varepsilon_k).
\end{aligned} \tag{38}$$

This implies that $\{w^k\} \subset \{w \in \text{int}(\mathcal{K}) \times \mathbb{R}^p \times \mathbb{R}^m : \hat{H}_\theta(w^*, w) \leq \bar{\alpha}\}$, with $\bar{\alpha} = \hat{H}_\theta(w^*, w^0) + \sum_{k=0}^\infty \lambda_k(\zeta_k + \varepsilon_k)$. By assumption (B2) and the fact that $\sum_{k=0}^\infty \lambda_k(\zeta_k + \varepsilon_k) < \infty$ (cf. Assumption (A3) and (37)), it follows that the sequence $\{w^k\}$ is bounded. Moreover, (38) together with $\hat{H}_\theta(w^*, w^{k+1}) \geq 0$ implies that there exists $l(w^*) \geq 0$ such that

$$\lim_{k \to \infty} \hat{H}_\theta(w^*, w^k) = l(w^*). \tag{39}$$

Therefore, by taking the limits on both sides of (38), we obtain

$$\|x^{k+1} - x^k\| \to 0, \ \|z^{k+1} - z^k\| \to 0, \ \|p^{k+1} - y^{k+1}\| \to 0, \ \|p^{k+1} - y^k\| \to 0. \tag{40}$$

On the other hand, since $\{w^k\}$ is bounded, there exists a subsequence $\{w^{k_j} = (x^{k_j}, z^{k_j}, y^{k_j})\}$ and a limit point $w^\infty = (x^\infty, z^\infty, y^\infty)$ such that $w^{k_j} \to w^\infty$. We now proceed to show that w^∞ is a saddle point of L. First, since $\lambda_k \geq \eta$, passing to the limit in (30) on the subsequence, and using (B3) and (40), we get

$$L(x^\infty, z^\infty, y^\infty) \leq L(x, z, y^\infty), \quad \forall x \in \mathcal{C}_1, \ \forall z \in \mathcal{C}_1. \tag{41}$$

Second, by applying Lemma 4.1, Part (iii) (in its exact form) with $F(\cdot) := -L(x^{k+1}, z^{k+1}, \cdot)$ we have

$$\lambda_k(L(x^{k+1}, z^{k+1}, y) - L(x^{k+1}, z^{k+1}, y^{k+1})) \leq \frac{1}{2}(\|y^k - y\|^2 - \|y^{k+1} - y\|^2), \quad \forall y \in \mathbb{R}^m.$$

Taking the limit on the subsequence in the above inequality and using (40), we have

$$L(x^\infty, z^\infty, y) \leq L(x^\infty, z^\infty, y^\infty), \quad \forall y \in \mathbb{R}^m. \tag{42}$$

It follows from (42) that $Ax^\infty + Bz^\infty = b$. Finally, since $\{x^k\} \subset \text{int}(\mathcal{K}_1)$ and $\{z^k\} \subset \text{int}(\mathcal{K}_2)$, passing to the limit one has that $(x^\infty, z^\infty) \in \mathcal{K}_1 \times \mathcal{K}_2$. Hence, from (41)-(42) the result follows. In

17

the case that $(B1'), (B3')$ hold, the proof is similar as above and by using (P4) of Definition 3.1 instead of (B2).

(ii) Suppose that (B4) holds. Let w^∞ be the limit of a subsequence $\{w^{k_j}\}$ of $\{w^k\}$, that is, $w^{k_j} \to w^\infty$. Then, by (B4) we have

$$\lim_{j \to \infty} \hat{H}_\theta(w^\infty, w^{k_j}) = 0. \tag{43}$$

Since w^∞ is a saddle point of L (by Part (i)), it follows from (39) and (43) that $l(w^\infty) = 0$. Hence, by using (21) and (33) we get that the sequence $\{w^k\}$ converges to w^∞. Now, if (B4') holds the result it follows by using the same arguments as before. □

5. Concluding remark and future works

In this paper, we have extended the proximal decomposition method given in [1, 14, 15, 18] for solving convex symmetric cone programming problems. The extension is based on proximal distance given in [2]. A class of this distance can be generated by a function ϕ that belongs to the set Σ or $\hat{\Sigma}$. Some examples of this distance have been presented. Under mild assumptions on the proximal distance, we have established that each limit point is a saddle point of the Lagrangian associated of the problem. Moreover, the global convergence has been obtained.

In some practical situations, f and/or g is nondifferentiable, so (CP) is a nonsmooth convex minimization problem. In most cases it is impossible to solve problems (24) and/or (25). The family of Bundle methods [38, Ch. 9] is perhaps the most practical computational tool for nonsmooth optimization. A future research will provide a convenient framework for coupling the decomposition scheme (23)-(26) with appropriate Bundle techniques for solving the nonsmooth problems. Moreover, to extend the algorithm SC-PMA in order to solve more general problems, such as, quasi-convex optimization and variational inequalities.

Acknowledgements

References

[1] G. Chen and M. Teboulle. A proximal-based decomposition method for convex minimization problems. *Mathematical Programming*, 64(1-3):81–101, 1994.

[2] A. Auslender and M. Teboulle. Interior gradient and proximal methods for convex and conic optimization. *SIAM Journal on Optimization*, 16(3):697–725, 2006.

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

[4] F. Alvarez, J. López, and H. Ramírez C. Interior proximal algorithm with variable metric for second-order cone programming: applications to structural optimization and support vector machines. *Optimization Methods Software*, 25(6):859–881, 2010.

[5] A. Li and L. Tuncel. Some applications of symmetric cone programming in financial mathematics. *Transactions in Operational Research*, 17:1–19, 2006.

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

[7] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38:49–95, 1994.

[8] F. Faybusovich. Linear systems in jordan algebras and primal-dual interior-point algorithms. *Journal of Computational and Applied Mathematics*, 86(1):149 – 175, 1997.

[9] L. Faybusovich. Euclidean jordan algebras and interior-point algorithms. *Positivity*, 1(4):331–357, 1997.

[10] Y. Lu and Y. Yuan. An interior-point trust-region algorithm for general symmetric cone programming. *SIAM Journal on Optimization*, 18(1):65–86, 2007.

[11] S.H. Schmieta and F. Alizadeh. Extension of primal-dual interior point algorithms to symmetric cones. *Mathematical Programming*, 96(3):409–438, 2003.

[12] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17 – 40, 1976.

[13] J.E. Spingarn. Applications of the method of partial inverses to convex programming: Decomposition. *Mathematical Programming*, 32(2):199–223, 1985.

[14] A. Auslender and M. Teboulle. Entropic proximal decomposition methods for convex programs and variational inequalities. *Mathematical Programming*, 91(1):33–47, 2001.

[15] M. Kyono and M. Fukushima. Nonlinear proximal decomposition method for convex programming. *Journal of Optimization Theory and Applications*, 106(2):357–372, 2000.

[16] A. Auslender, M. Teboulle, and S. Ben-Tiba. A logarithmic-quadratic proximal method for variational inequalities. *Computational Optimization and Applications*, pages 31–40, 1999.

[17] K.C. Kiwiel. Proximal minimization methods with generalized bregman functions. *SIAM Journal on Control and Optimization*, 35(4):1142–1168, 1997.

[18] O. Sarmiento, E. Papa Quiroz, and P. Oliveira. A proximal multiplier method for separable convex minimization, 2014. Available on Optimization Online, submitted for publication.

[19] S. Pan and J.S. Chen. Interior proximal methods and central paths for convex second-order cone programming. *Nonlinear Analysis: Theory, Methods & Applications*, 73(9):3083 – 3100, 2010.

[20] R.T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[21] J. Faraut and A. Korányi. *Analysis on symmetric cones*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 1994. Oxford Science Publications.

[22] E.B. Vinberg. The theory of homogeneous convex cones. *Transactions of the Moscow Mathematical Society*, 12:340–403, 1963.

[23] M. Baes. Convexity and differentiability properties of spectral functions and spectral mappings on euclidean jordan algebras. *Linear Algebra and its Applications*, 422(2-3):664 – 700, 2007.

[24] D. Sun and J. Sun. Löwner's operator and spectral functions in euclidean jordan algebras. *Mathematics Operations Research*, 33(2):421–445, 2008.

[25] J.-S. Chen and S. Pan. An entropy-like proximal algorithm and the exponential multiplier method for convex symmetric cone programming. *Computational Optimization and Applications*, 47(3):477–499, 2010.

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

[27] F. Liese and I. Vajda. *Convex statistical distances*. Teubner, 1987.

[28] M. Teboulle. Convergence of proximal-like algorithms. *SIAM Journal on Optimization*, 7(4):1069–1083, 1997.

[29] M. Teboulle. Entropic proximal mappings with applications to nonlinear programming. *Mathematical Operations Research*, 17(3):670–690, 1992.

[30] I. Csiszar. An axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, 19(4):2032–2066, 1991.

[31] S. Pan and J.S. Chen. A class of interior proximal-like algorithms for convex second-order cone programming. *SIAM Journal on Optimization*, 19(2):883–910, 2008.

[32] J.S. Chen. The convex and monotone functions associated with second-order cone. *Optimization*, 55(4):363–385, 2006.

[33] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 1995.

[34] D. Goldfarb, S. Ma, and K. Scheinberg. Fast alternating linearization methods for minimizing the sum of two convex functions. *Mathematical Programming*, 141(1-2):349–382, 2013.

[35] S. Katya, M. Shiqian, and Donald G. Sparse inverse covariance selection via alternating linearization methods. *CoRR*, abs/1011.0097, 2010.

[36] E. de Klerk. *Aspects of semidefinite programming: interior point algorithms and selected applications*. Applied optimization. Kluwer Academic Publishers, 2002.

[37] O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29(2):403–419, 1991.

[38] J.F. Bonnans, J.C. Gilbert, C. Lemarechal, and C.A. Sagastizábal. *Numerical Optimization: Theoretical and Practical Aspects*. Universitext. Springer, 2006.