Linear and Nonlinear Analysis
Yokohama Publishers
ISSN 2188-8167

Volume 5, Number 3, 2019, 421–437

A SEMI-DISTANCE ASSOCIATED WITH SYMMETRIC CONE AND A NEW PROXIMAL DISTANCE FUNCTION ON SECOND-ORDER CONE

XIN-HE MIAO*, CHIEN-HAO HUANG, YONGDO LIM†, AND JEIN-SHAN CHEN‡

Abstract. In this paper, we introduce a new semi-distance on a second-order cone satisfying all proximal properties in Auslender and Teboulle [1]. To our best knowledge, it may be the only proximal distance which is not induced from Bregman distance, φ-divergence, or distance-like entropy function. With this new discovery, some algorithms based on proximal distance, for example, proximal point algorithm and proximal-like algorithm can be applied accordingly to solve second-order cone optimizations.

1. Introduction

We consider the following nonlinear symmetric cone programming (SCP):

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

where $\mathbb{V}$ is a Euclidean Jordan algebra, $\mathcal{K} \subset \mathbb{V}$ denotes the associated symmetric cone of invertible squares, $f : \mathbb{V} \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous (l.s.c. in short) convex function. A popular approach to deal with SCP is the proximal point algorithm, which generates a sequence $\{x^n\}$ via the following iterative scheme:

$$x^{k+1} = \arg\min_{x \in \mathcal{K}} \{f(x) + \lambda_k D(x, x^k)\}.$$

Here $D(\cdot,\cdot)$ is a certain function satisfying some desirable properties and $\{\lambda_k\}_{k\in\mathbb{N}}$ a positive sequence. The choice of $D(\cdot,\cdot)$ is important and examples of $D(\cdot,\cdot)$ are the distances induced by Euclidean norm, quasi-distance, Bregman distance, φ-divergence, and proximal distance, etc..

2010 *Mathematics Subject Classification.* 90C25, 15B99, 33E99.

Key words and phrases. Second-order cone, symmetric cone, proximal distance, convexity.

*The author's work is supported by National Natural Science Foundation of China (No. 11471241).

†The author's work is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MEST) No.2015R1A3A2031159.

‡Corresponding author. The author's work is supported by Ministry of Science and Technology, Taiwan.

Recall that a *distance function* (simply *distance* or called *metric*) on a set X is a function $d : X \times X \to \mathbb{R}_+$ satisfying that, for all $x, y, z \in X$,

(D1)	$d(x,y) \geq 0$;	(Nonnegative)
(D2)	$d(x,y) = 0 \Longleftrightarrow x = y$;	(Identity of indiscernibles)
(D3)	$d(x,y) = d(y,x)$;	(Symmetry)
(D4)	$d(x,z) \leq d(x,y) + d(y,z)$.	(Triangle inequality)

There are several ways of relaxing the axioms of distance. For example, a *semi-distance* is defined as a function that satisfies all axioms for a distance with the possible exception of (D4). A *quasi-distance* is defined as a function that satisfies all axioms for a distance with the exception of (D3).

The Bregman distance (also called Bregman divergence) was introduced by Bregman [5], whose definition is similar to a metric, but does not satisfy the triangle inequality. More specifically, let $\phi : S \subseteq \mathbb{R}^d \to \mathbb{R}$ be a real-valued convex function with gradient $\nabla\phi$, the *Bregman distance* (or *Bregman divergence*) $d_\phi : S \times \mathrm{ri}(S) \to \mathbb{R}$ is defined as

$$d_\phi(x,y) = \phi(x) - \phi(y) - \langle x - y, \nabla\phi(y)\rangle.$$

Here $\mathrm{ri}(S)$ denotes the relative interior of S. In particular, when taking $\phi(x) = \|x\|^2$, the corresponding Bregman distance is

$$d_\phi(x,y) = \|x - y\|^2.$$

For $\phi(x) = \sum_{i=1}^n (x_i \ln x_i - x_i)$, the corresponding Bregman distance is

$$d_\phi(x,y) = \sum_{i=1}^n \left(x_i \ln \frac{x_i}{y_i} - x_i + y_i \right). \tag{1.1}$$

For a collection of popular Bregman distances, please refer to [3].

The definition in above can be naturally extended to the space $\mathcal{S}^n$ of $n \times n$ real symmetric matrices. In other words, given a convex function $\phi : \mathcal{S}^n \to \mathbb{R}$, the Bregman distance (or Bregman matrix divergence) is defined to be

$$\begin{aligned} d_\phi(X,Y) &= \phi(X) - \phi(Y) - \langle X - Y, \nabla\phi(Y)\rangle \\ &= \phi(X) - \phi(Y) - \mathrm{tr}(X - Y)\nabla\phi(Y) \end{aligned}$$

Again, when taking $\phi(X) = \|X\|_F^2$, the corresponding Bregman distance is exactly the squared Frobenius norm $\|X - Y\|_F^2$. In addition, let $X \in \mathcal{S}$ with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, taking

$$\phi(X) = \sum_{i=1}^n (\lambda_1 \ln \lambda_i - \lambda_i), \quad \text{alternatively expressed as} \quad \phi(X) = \mathrm{tr}(X \ln X - X),$$

leads to the resulting Bregman distance

$$d_\phi(X,Y) = \mathrm{tr}\,(X \ln X - X \ln Y - X + Y).$$

The Bregman distance is widely used in clustering [3], co-clustering [4], low-rank matrix approximation [12, 17, 20], online learning [27], probability measure [25], and matrix nearness problem [13], etc.. Moreover, it is also employed to various proximal point algorithms and proximal-like algorithms, see [6, 9, 10, 18, 23] and references therein.

In the context of proximal methods, another kind of distance function, the so-called φ-divergence proposed by Teboulle [26], was also considered. More specifically, this type of distance is based on some properties as below. Let $\varphi : \mathbb{R}_{++} \to \mathbb{R}$ be a closed and proper convex function that satisfies the following properties:

(φ1) φ is twice continuously differentiable;
(φ2) φ is strictly convex;
(φ3) $\varphi(1) = \varphi'(1) = 0$ and $\varphi''(1) > 0$
(φ4) $\lim_{t\to 0^+} \varphi'(t) = -\infty$.

The *φ-divergence* $d_\varphi : \mathbb{R}^n_{++} \times \mathbb{R}^n_{++} \to \mathbb{R}$ is defined by

$$d_\varphi(x, y) = \sum_{i=1}^{n} y_i \varphi\left(\frac{x_i}{y_i}\right),$$

where φ satisfies (i)-(iv). Note that, when $\varphi(t) := t - \ln t - 1$, the corresponding φ-divergence is

$$d_\varphi(x, y) = \sum_{i=1}^{n} \left(y_i \ln \frac{y_i}{x_i} + x_i - y_i \right) \overset{(1.1)}{=} d_\phi(y, x).$$

Both Bregman distances and φ-divergences satisfy (D1)-(D2), that is, they are the *pre-distances*, but they do not satisfy the (D3)-(D4) in general. However, these distance functions satisfy some other desirable properties, e.g., (φ1)-(φ2).

In this paper, we introduce a semi-distance function associated with symmetric cone. In light of this semi-distance, we construct a proximal distance in the setting of second-order cone. To our best knowledge, it may be the only proximal distance which is not induced from Bregman distance, φ-divergence, or distance-like entropy function. This provides a good contribution to the literature. As mentioned, some algorithms based on proximal distance like proximal point algorithm and proximal-like algorithm can be applied accordingly to solve second-order cone optimizations. The outline of this paper is as follows. In section 2, we review some basic concepts regarding symmetric cone and second-order cone, and then propose a new semi-distance associated with symmetric cone and investigate its convexity. In section 3, according to the semi-distance, a proximal distance with respect to second-order cone is constructed. Finally, we draw a conclusion in section 4.

Throughout this paper, $\mathbb{R}^n$ denotes n-dimensional Euclidean space endowed with the canonical inner product $\langle \cdot\,, \cdot\rangle$ and the norm of x given by $\|x\| = \langle x, x\rangle^{\frac{1}{2}}$ is the Euclidean norm. In addition, for any subset C of $\mathbb{R}^n$, the interior of C is denoted by int$C$, the closure of C is denoted by $\bar{C}$ and the boundary of C is denoted by bd$C$. We also adopt the standard natation of convex analysis [24]. For a proper convex and l.s.c. function $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, its effective domain is defined by $\mathrm{dom}F = \{x \,|\, F(x) < +\infty\}$, $\nabla F(x)$ denotes the gradient of F at x whenever F is differentiable at x, and for $\epsilon \geq 0$ its ϵ-subdifferential at x is defined by

$$\partial_\epsilon F(X) = \{\xi \in \mathbb{R}^n \,|\, \forall z \in \mathbb{R}^n,\ F(z) + \epsilon \geq F(x) + \langle \xi, z - x\rangle\},$$

which coincides with the usual subdifferential $\partial F \equiv \partial_0 F$ whenever $\epsilon = 0$. In addition, we denote $\mathrm{dom}\,\partial F := \{x \in \mathbb{R}^n \,|\, \partial F(x) \neq \emptyset\}$.

2. A semi-distance associated with symmetric cone

In this section, we propose a new semi-distance function on a symmetric cone and construct a proximal distance in the setting of second-order cone. To this end, we review some basic concepts and properties on symmetric cones and second-order cones, which are needed in the subsequent analysis.

A *Euclidean Jordan algebra* [15] is a finite dimensional inner product space $(\mathbb{V}, \langle\cdot,\cdot\rangle)$ ($\mathbb{V}$ for short) over the field of real numbers $\mathbb{R}$ equipped with a bilinear map $(x, y) \mapsto x \circ y : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$, which satisfies the following 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 $x, y \in \mathbb{V}$;
: (iii) $\langle x \circ y, z\rangle = \langle x, y \circ z\rangle$ for all $x, y, z \in \mathbb{V}$,

where $x^2 := x \circ x$, and $x \circ y$ is called the *Jordan product* of x and y. If a Jordan product only satisfies the conditions (i) and (ii) in the above definition, the algebra $\mathbb{V}$ is said to be a *Jordan algebra*. Moreover, if there is an (unique) element $e \in \mathbb{V}$ such that $x \circ e = x$ for all $x \in \mathbb{V}$, the element e is called the *identity element* in $\mathbb{V}$. Note that a Jordan algebra does not necessarily have an identity element. Throughout this paper, we assume that $\mathbb{V}$ is a Euclidean Jordan algebra with an identity element e.

In a given Euclidean Jordan algebra $\mathbb{V}$, the set of squares $\mathcal{K} := \{x^2 \,|\, x \in \mathbb{V}\}$ is a *symmetric cone* [15, Theorem III.2.1]. This means that $\mathcal{K}$ is a self-dual closed convex cone and, for any two elements $x, y \in \mathrm{int}(\mathcal{K})$, there exists an invertible linear transformation $\Gamma : \mathbb{V} \to \mathbb{V}$ such that $\Gamma(x) = y$ and $\Gamma(\mathcal{K}) = \mathcal{K}$. We introduce the second-order cone in $\mathbb{R}^n$, an important example of symmetric cones, which is defined as follows:

$$\mathcal{K}^n := \left\{x = (x_0, \bar{x}) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid x_0 \geq \|\bar{x}\|\right\},$$

and the corresponding Jordan product of x and y in $\mathbb{R}^n$ with $x = (x_0, \bar{x}), y = (y_0, \bar{y}) \in \mathbb{R} \times \mathbb{R}^{n-1}$ is given by

$$x \circ y := \begin{bmatrix} x^T y \\ x_0\bar{y} + y_0\bar{x} \end{bmatrix}.$$

We note that $e = (1, 0) \in \mathbb{R} \times \mathbb{R}^{n-1}$ acts as the Jordan identity.

For any given $x \in \mathbb{V}$, we denote $m(x)$ the *degree* of the minimal polynomial of x, that is,

$$m(x) := \left\{ k > 0 \,|\, \{e, x, \ldots, x^k\} \text{ is linearly dependent} \right\}.$$

Since $m(x) \leq \dim(\mathbb{V})$ where $\dim(\mathbb{V})$ is the dimension of $\mathbb{V}$, the *rank* of $\mathbb{V}$ is well-defined by $r := \max\{m(x) \,|\, x \in \mathbb{V}\}$. In Euclidean Jordan algebra $\mathbb{V}$, an element $e^i \in \mathbb{V}$ is an *idempotent* if $(e^i)^2 = e^i$, and it is a *primitive idempotent* if it is nonzero and cannot be written as a sum of two nonzero idempotents. The idempotents e^i and e^j are said to be *orthogonal* if $e^i \circ e^j = 0$. In addition, we say that a finite set $\{e^1, e^2, \ldots, e^r\}$ of primitive idempotents in $\mathbb{V}$ is a *Jordan frame* if

$$e^i \circ e^j = 0 \quad \text{for} \quad i \neq j, \quad \text{and} \quad \sum_{i=1}^{r} e^i = e.$$

Note that $\langle e^i, e^j \rangle = \langle e^i \circ e^j, e \rangle$ whenever $i \neq j$.

With the above, there have the spectral decomposition and Peirce decomposition of an element x in $\mathbb{V}$.

Theorem 2.1 (The Spectral Decomposition Theorem [15, Theorem III.1.2])**.** *Let $\mathbb{V}$ be a Euclidean Jordan algebra. Then there is a number r such that, for every $x \in \mathbb{V}$, there exists a Jordan frame $\{e^1, \ldots, e^r\}$ and real numbers $\lambda_1(x), \ldots, \lambda_r(x)$ with*

$$x = \lambda_1(x)e^1 + \cdots + \lambda_r(x)e^r.$$

Here, the numbers $\lambda_i(x)$ $(i = 1, \ldots, r)$ are the eigenvalues of x, the expression $\lambda_1(x)e^1 + \cdots + \lambda_r(x)e^r$ is the spectral decomposition of x. Moreover, $\operatorname{tr} x := \sum_{i=1}^{r} \lambda_i(x)$ *is called the trace of x, and* $\det(x) = \lambda_1(x)\lambda_2(x) \ldots \lambda_r(x)$.

In the setting of second-order cone in $\mathbb{R}^n$, the spectral decomposition of $x = (x_0, \bar{x}) \in \mathbb{R} \times \mathbb{R}^{n-1}$ with $\bar{x} \neq 0$ reduces to

$$x = \lambda_1(x)e^1 + \lambda_2(x)e^2,$$

where $\lambda_i(x) = x_0 + (-1)^i \|\bar{x}\|$ and $e^i = \frac{1}{2}\left(1, (-1)^i \frac{\bar{x}}{\|\bar{x}\|}\right)$. Accordingly, we know that $\operatorname{tr} x = \lambda_1(x) + \lambda_2(x) = 2x_0$ and $\det(x) = \lambda_1(x)\lambda_2(x) = x_0^2 - \|\bar{x}\|^2$, see [8, 11, 14].

We point out that different elements x, y have their own Jordan frames in the spectral decomposition, which are not easy to handle when we need to do operations for x and y. Thus, we need another so-called Peirce decomposition to conquer such difficulty. In other words, in the Peirce decomposition, two different elements x, y share the same Jordan frame. We elaborate them more as below.

The Peirce decomposition: Fix a Jordan frame $\{e^1, e^2, \ldots, e^r\}$ in a Euclidean Jordan algebra $\mathbb{V}$. For $i, j \in \{1, 2, \ldots, r\}$, we define the following eigen-spaces

$$\mathbb{V}_{ii} := \{x \in \mathbb{V} \mid x \circ e^i = x\} = \mathbb{R}e^i$$

and

$$\mathbb{V}_{ij} := \left\{ x \in \mathbb{V} \mid x \circ e^i = \frac{1}{2}x = x \circ e^j \right\} \quad \text{for} \quad i \neq j.$$

Theorem 2.2 ([15, Theorem IV.2.1])**.** *The space $\mathbb{V}$ is the orthogonal direct sum of spaces $\mathbb{V}_{ij} (i \leq j)$. Furthermore,*

$$\begin{aligned}
&\mathbb{V}_{ij} \circ \mathbb{V}_{ij} \subset \mathbb{V}_{ii} + \mathbb{V}_{jj}, \\
&\mathbb{V}_{ij} \circ \mathbb{V}_{jk} \subset \mathbb{V}_{ik}, \quad \text{if} \quad i \neq k, \\
&\mathbb{V}_{ij} \circ \mathbb{V}_{kl} = \{0\}, \quad \text{if} \quad \{i,j\} \cap \{k,l\} = \emptyset.
\end{aligned}$$

Hence, given any Jordan frame $\{e^1, e^2, \ldots, e^r\}$, we can write any element $x \in \mathbb{V}$ as

$$x = \sum_{i=1}^{r} x_i e^i + \sum_{i<j} x_{ij},$$

where $x_i \in \mathbb{R}$ and $x_{ij} \in \mathbb{V}_{ij}$. The expression $\sum_{i=1}^r x_i e^i + \sum_{i<j} x_{ij}$ is called the Peirce decomposition of x.

Given a Euclidean Jordan algebra $\mathbb{V}$ with $\dim(\mathbb{V}) = n > 1$, from Proposition III 4.4-4.5 and Theorem V.3.7 in [15], we know that any Euclidean Jordan algebra $\mathbb{V}$ and its corresponding symmetric cone $\mathcal{K}$ are, in a unique way, a direct sum of simple Euclidean Jordan algebras and the constituent symmetric cones therein, respectively, i.e.,

$$\mathbb{V} = \mathbb{V}_1 \times \cdots \times \mathbb{V}_m \quad \text{and} \quad \mathcal{K} = \mathcal{K}^1 \times \cdots \times \mathcal{K}^m,$$

where every $\mathbb{V}_i$ is a simple Euclidean Jordan algebra (that cannot be a direct sum of two Euclidean Jordan algebras) with the corresponding symmetric cone $\mathcal{K}^i$ for $i = 1, \ldots, m$, and $n = \sum_{i=1}^m n_i$ (n_i is the dimension of $\mathbb{V}_i$). Therefore, for any $x = (x_1, \ldots, x_m)^T$ and $y = (y_1, \ldots, y_m)^T \in \mathbb{V}$ with $x_i, y_i \in \mathbb{V}_i$, we have

$$x \circ y = (x_1 \circ y_1, \ldots, x_m \circ y_m)^T \in \mathbb{V} \quad \text{and} \quad \langle x, y \rangle = \langle x_1, y_1 \rangle + \cdots + \langle x_m, y_m \rangle.$$

For simplicity, we focus on the single symmetric cone $\mathcal{K}$ because all the analysis can be carried over to the setting of Cartesian product.

In a Euclidean Jordan algebras $\mathbb{V}$, for any $x \in \mathbb{V}$, A linear transformation $L(x) : \mathbb{V} \to \mathbb{V}$ is called *Lyapunov transformation*, which is defined as $L(x)(y) := x \circ y$ for all $y \in \mathbb{V}$. The so-called *quadratic representation* $P(x)$ is define by $P(x) := 2L^2(x) - L(x^2)$. For any $x \in \mathbb{V}$, the endomorphisms $L(x)$ and $P(x)$ are self-adjoint. We say that two elements x and y of a Euclidean Jordan algebra $\mathbb{V}$ *operator commute* if $x \circ (y \circ z) = y \circ (x \circ z)$ for all $z \in \mathbb{V}$, which is equivalent to stating that $L(x)L(y) = L(y)L(x)$. For the quadratic representation $P(x)$, if x is invertible, then we have

$$P(x)\mathcal{K} = \mathcal{K} \quad \text{and} \quad P(x)\text{int}(\mathcal{K}) = \text{int}(\mathcal{K}).$$

In the setting of second-order cone, for any $x = (x_0, \bar{x}) \in \mathbb{R} \times \mathbb{R}^{n-1}$, we know that

$$L(x) = \begin{bmatrix} x_0 & \bar{x}^T \\ \bar{x} & x_0 I \end{bmatrix}$$

with I being the identity matrix in $\mathbb{R}^{(n-1)\times(n-1)}$, and the quadratic representation $P(x)$ can be expressed as

$$P(x) = 2xx^T - \det(x)J,$$

where $J := \begin{bmatrix} 1 & 0^T \\ 0 & -I \end{bmatrix}$ with I being the identity matrix in $\mathbb{R}^{(n-1)\times(n-1)}$, see [15, 21]. In addition, it is easy to see that the vectors x and y with $x = (x_0, \bar{x})$ and $y = (y_0, \bar{y})$ operator commute if and only if either $\bar{y}$ is a multiple of $\bar{x}$ or $\bar{x}$ is a multiple of $\bar{y}$.

Now, based on the symmetric cone trace function $\operatorname{tr}(\cdot)$, we propose a new semi-distance function associated with symmetric cone as bellow:

$$\mathrm{d}(x, y) := \operatorname{tr}(x + y) - 2 \operatorname{tr} \left(P(x^{\frac{1}{2}})y\right)^{\frac{1}{2}}, \quad \text{for } x, y \in \mathcal{K}. \tag{2.1}$$

In fact, when the symmetric cone reduces to the semi-definite positive matrix cone, the function $\mathrm{d}(A, B) := \operatorname{tr}(A + B) - 2 \operatorname{tr}(A^{\frac{1}{2}} B A^{\frac{1}{2}})^{\frac{1}{2}}$ appears in the Wasserstein distance between Gaussian measures [2, 16, 19].

The following theorem shows the properties of the function $\mathrm{d}(\cdot, \cdot)$.

Theorem 2.3. *Let $\mathrm{d}(\cdot, \cdot)$ be defined as in (2.1). For any $x, y \in \mathcal{K}$, assume that x and y operator commute. Then, we have*

- **(a):** $\mathrm{d}(x, y) \geq 0$;
- **(b):** $\mathrm{d}(x, y) = 0$ *if and only if* $x = y$;
- **(c):** $\mathrm{d}(x, y) = \mathrm{d}(y, x)$.

Proof. (a) Since $x, y \in \mathcal{K}$ and x and y operator commute, it follows from [15, Lemma X.2.2] that x and y have the same Jordan frame. In other words, there exists a Jordan frame $\{e^1, e^2, \ldots, e^r\}$ and the spectral decomposition of x and y can be expressed as below, respectively,

$$\begin{aligned} x &= \lambda_1(x)\, e^1 + \cdots + \lambda_r(x)\, e^r, \\ y &= \mu_1(y)\, e^1 + \cdots + \mu_r(y)\, e^r. \end{aligned}$$

From this, we obtain $x^{\frac{1}{2}} = \sqrt{\lambda_1(x)}\, e^1 + \cdots + \sqrt{\lambda_r(x)}\, e^r$, which implies that

$$P(x^{\frac{1}{2}})(y) = x \circ y = \lambda_1(x)\mu_1(y)e^1 + \cdots + \lambda_r(x)\mu_r(y)e^r.$$

Thus, we have $\left(P(x^{\frac{1}{2}})(y)\right)^{\frac{1}{2}} = \sqrt{\lambda_1(x)\mu_1(y)}\, e^1 + \cdots + \sqrt{\lambda_r(x)\mu_r(y)}\, e^r$, which further leads to

$$\operatorname{tr} \left(P(x^{\frac{1}{2}})(y)\right)^{\frac{1}{2}} = \sum_{i=1}^{r} \sqrt{\lambda_i(x)\mu_i(y)}.$$

Then, it follows that

$$\begin{aligned}
\mathrm{d}(x,y) &= \mathrm{tr}(x+y) - 2\,\mathrm{tr}\left(P(x^{\frac{1}{2}})(y)\right)^{\frac{1}{2}} \\
&= \sum_{i=1}^{r}(\lambda_i + \mu_i) - 2\sum_{i=1}^{r}\sqrt{\lambda_i(x)\mu_i(y)} \\
&= \sum_{i=1}^{r}\left(\sqrt{\lambda_i(x)} - \sqrt{\mu_i(y)}\right)^2 \\
&= \left\|x^{\frac{1}{2}} - y^{\frac{1}{2}}\right\|^2,
\end{aligned}$$

which proves $\mathrm{d}(x,y) \geq 0$.

(b) From the proof of part (a), we know that $\mathrm{d}(x,y) = \left\|x^{\frac{1}{2}} - y^{\frac{1}{2}}\right\|^2$. Hence, it is clear to see that $\mathrm{d}(x,y) = 0$ if and only if $x = y$.

(c) From the expression of $\mathrm{d}(x,y)$, $\mathrm{d}(x,y) = \mathrm{d}(y,x)$ is obvious. $\square$

We make some remarks regarding Theorem 2.3 here.

(1) By the fundamental theorem for Jordan algebras [15, Proposition II. 3.3]), the symmetric property in Theorem 2.3(c) holds for general $x, y \in \mathcal{K}$. Indeed, $P(P(x^{\frac{1}{2}})y) = P(x^{\frac{1}{2}})P(y)P(x^{\frac{1}{2}})$ is similar to $P(y^{\frac{1}{2}})P(x)P(y^{\frac{1}{2}}) = P(P(y^{\frac{1}{2}})x)$.

(2) For the cone of positive definite matrices of fixed size, the properties in Theorem 2.3(a)-(b) hold from the following relation to the extremal problem (see [2, Theorem 1]):

$$\min_{UU^*=I}\|A^{1/2} - B^{1/2}U\|_2^2 = \mathrm{d}(A,B).$$

Nonetheless, in general, we do not know whether properties in Theorem 2.3(a)-(b) hold for a general symmetric cone (without assuming operator commute).

Theorem 2.4. *Let* $\mathrm{d}(\cdot,\cdot)$ *be defined as in (2.1). Then,* $\mathrm{d}(x,y)$ *is convex, for any a fixed* x *or* y*.*

Proof. First, we fix the element x. From [7, Theorem 3.2], we know that for any $z \in \mathcal{K}$, the trace function $\mathrm{tr}(z)^{\frac{1}{2}}$ is concave. Moreover, by [15, Proposition III 2.2], we have that $P(x^{\frac{1}{2}})(y) \in \mathcal{K}$ for $x, y \in \mathcal{K}$, and $P(x^{\frac{1}{2}})(y)$ with respect to y is linear mapping. Combining with the concavity of the trace function $\mathrm{tr}(\cdot)^{\frac{1}{2}}$, it is easy to verify that the trace function $\mathrm{tr}(P(x^{\frac{1}{2}})(y))^{\frac{1}{2}}$ with respect to y is concave. Therefore, $-2\,\mathrm{tr}(P(x^{\frac{1}{2}})y)^{\frac{1}{2}}$ with respect to y is convex. Since the trace function $\mathrm{tr}(x+y)$ is convex, it is not hard to obtain that $\mathrm{d}(x,y)$ with respect to y is also convex.

If we fix the element y, with the same arguments, we have that $\mathrm{d}(y,x)$ with respect to y is convex. Applying part (c) of Theorem 2.3, we know that $\mathrm{d}(x,y)=\mathrm{d}(y,x)$. Hence, we obtain that $\mathrm{d}(x,y)$ with respect to y is also convex. $\Box$

Next, we discuss the corresponding results in the setting of second-order cone $\mathcal{K}^n$. For any $x\in\mathcal{K}^n$ with $x:=(x_0,\bar{x})\in\mathbb{R}\times\mathbb{R}^{n-1}$, we know

$$x^{\frac{1}{2}}=\begin{bmatrix}\frac{\sqrt{x_0-\|\bar{x}\|}+\sqrt{x_0+\|\bar{x}\|}}{2}\\ \frac{\sqrt{x_0+\|\bar{x}\|}-\sqrt{x_0-\|\bar{x}\|}}{2}\frac{\bar{x}}{\|\bar{x}\|}\end{bmatrix}=\begin{bmatrix}\sqrt{\frac{x_0+\sqrt{x_0^2-\|\bar{x}\|^2}}{2}}\\ \frac{1}{\sqrt{2\left(x_0+\sqrt{x_0^2-\|\bar{x}\|^2}\right)}}\bar{x}\end{bmatrix}.$$

Since $P(z)=2zz^T-\det(z)J$ for any $z\in\mathbb{R}^n$ in the setting of second-order cone, where $J=\begin{bmatrix}1&0\\0&-I\end{bmatrix}$, it follows that for any $y\in\mathcal{K}^n$ with $y:=(y_0,\bar{y})\in\mathbb{R}\times\mathbb{R}^{n-1}$,

$$\begin{aligned}P(x^{\frac{1}{2}})(y)&=\left(2x^{\frac{1}{2}}(x^{\frac{1}{2}})^T-\det(x^{\frac{1}{2}})J\right)(y)\\&=\begin{bmatrix}x_0y_0+\bar{x}^T\bar{y}\\ y_0\bar{x}+\frac{\bar{x}^T\bar{y}\bar{x}}{x_0+\sqrt{x_0^2-\|\bar{x}\|^2}}+\sqrt{x_0^2-\|\bar{x}\|^2}\bar{y}\end{bmatrix}\\&:=\begin{bmatrix}x_0y_0+\bar{x}^T\bar{y}\\ \bar{w}\end{bmatrix}.\end{aligned}$$

This yields that

$$\left(P(x^{\frac{1}{2}})(y)\right)^{\frac{1}{2}}=\begin{bmatrix}\sqrt{\frac{x_0y_0+\bar{x}^T\bar{y}+\sqrt{(x_0y_0+\bar{x}^T\bar{y})^2-\|\bar{w}\|^2}}{2}}\\ \frac{1}{\sqrt{2\left(x_0y_0+\bar{x}^T\bar{y}+\sqrt{(x_0y_0+\bar{x}^T\bar{y})^2-\|\bar{w}\|^2}\right)}}\bar{w}\end{bmatrix}.$$

Note that

$$\begin{aligned}&\|\bar{w}\|^2\\=\;&\langle\bar{w},\bar{w}\rangle\\=\;&y_0^2\|\bar{x}\|^2+\frac{(\bar{x}^T\bar{y})^2\|\bar{x}\|^2}{(x_0+\sqrt{x_0^2-\|\bar{x}\|^2})^2}+(x_0^2-\|\bar{x}\|^2)\|\bar{y}\|^2+\frac{2(\bar{x}^T\bar{y})^2\sqrt{x_0^2-\|\bar{x}\|^2}}{x_0+\sqrt{x_0^2-\|\bar{x}\|^2}}\\&+\frac{2y_0\bar{x}^T\bar{y}\|\bar{x}\|^2}{x_0+\sqrt{x_0^2-\|\bar{x}\|^2}}+2y_0\bar{x}^T\bar{y}\sqrt{x_0^2-\|\bar{x}\|^2}\\=\;&y_0^2\|\bar{x}\|^2+x_0^2\|\bar{y}\|^2-\|\bar{x}\|^2\|\bar{y}\|^2+2y_0\bar{x}^T\bar{y}\left(\frac{\|\bar{x}\|^2}{x_0+\sqrt{x_0^2-\|\bar{x}\|^2}}+\sqrt{x_0^2-\|\bar{x}\|^2}\right)\\&+\frac{(\bar{x}^T\bar{y})^2}{x_0+\sqrt{x_0^2-\|\bar{x}\|^2}}\left(\frac{\|\bar{x}\|^2}{x_0+\sqrt{x_0^2-\|\bar{x}\|^2}}+2\sqrt{x_0^2-\|\bar{x}\|^2}\right)\\=\;&y_0^2\|\bar{x}\|^2+x_0^2\|\bar{y}\|^2-\|\bar{x}\|^2\|\bar{y}\|^2+2x_0y_0\bar{x}^T\bar{y}+(\bar{x}^T\bar{y})^2.\end{aligned}$$

Hence, we have

$$
\begin{aligned}
\operatorname{tr}\left(P(x^{\frac{1}{2}})(y)\right)^{\frac{1}{2}} &= 2\sqrt{\frac{x_0y_0+\bar{x}^T\bar{y}+\sqrt{(x_0y_0+\bar{x}^T\bar{y})^2-\|\bar{w}\|^2}}{2}} \\
&= \sqrt{2(x_0y_0+\bar{x}^T\bar{y})+2\sqrt{(x_0^2-\|\bar{x}\|^2)(y_0^2-\|\bar{y}\|^2)}}. \qquad (2.2)
\end{aligned}
$$

With the above expression (2.2), we show that the function $\mathrm{d}(\cdot,\cdot)$ on the second-order cone is a semi-distance.

Theorem 2.5. *Let the function* $\mathrm{d}(\cdot,\cdot)$ *be defined by (2.1). Then, for any* $x, y \in \mathcal{K}^n$*, we have*

- **(a):** $\mathrm{d}(x,y) \geq 0$*;*
- **(b):** $\mathrm{d}(x,y) = 0$ *if and only if* $x = y$*;*
- **(c):** $\mathrm{d}(x,y) = \mathrm{d}(y,x)$*.*

Proof. (a) For any $x, y \in \mathcal{K}^n$ with $x := (x_0, \bar{x}), y := (y_0, \bar{y}) \in \mathbb{R}\times\mathbb{R}^{n-1}$, by expression (2.2), we have

$$
\begin{aligned}
\operatorname{tr}\left(P(x^{\frac{1}{2}})(y)\right)^{\frac{1}{2}} &= \sqrt{2(x_0y_0+\bar{x}^T\bar{y})+2\sqrt{(x_0^2-\|\bar{x}\|^2)(y_0^2-\|\bar{y}\|^2)}} \\
&\leq \sqrt{2(x_0y_0+\|\bar{x}\|\|\bar{y}\|)+2\sqrt{(x_0^2-\|\bar{x}\|^2)(y_0^2-\|\bar{y}\|^2)}} \\
&= \sqrt{(x_0+\|\bar{x}\|)(y_0+\|\bar{y}\|)}+\sqrt{(x_0-\|\bar{x}\|)(y_0-\|\bar{y}\|)},
\end{aligned}
$$

which gives

$$
\begin{aligned}
\mathrm{d}(x,y) &= \operatorname{tr}(x+y) - 2\operatorname{tr}\left(P(x^{\frac{1}{2}})(y)\right)^{\frac{1}{2}} \\
&\geq 2(x_0+y_0) - 2\left(\sqrt{(x_0+\|\bar{x}\|)(y_0+\|\bar{y}\|)}+\sqrt{(x_0-\|\bar{x}\|)(y_0-\|\bar{y}\|)}\right) \\
&= (\sqrt{x_0+\|\bar{x}\|}-\sqrt{y_0+\|\bar{y}\|})^2 + (\sqrt{x_0-\|\bar{x}\|}-\sqrt{y_0-\|\bar{y}\|})^2 \\
&\geq 0.
\end{aligned}
$$

(b) If $x = y$, it is easy to verify that $\mathrm{d}(x,y) = 0$. It remains to show the other direction. Suppose that $\mathrm{d}(x,y) = 0$. From the proof of part (a), we see that

$$
\begin{aligned}
0 &= \mathrm{d}(x,y) \\
&= 2(x_0+y_0) - 2\sqrt{2(x_0y_0+\bar{x}^T\bar{y})+2\sqrt{(x_0^2-\|\bar{x}\|^2)(y_0^2-\|\bar{y}\|^2)}} \\
&\geq 2(x_0+y_0) - 2\sqrt{2(x_0y_0+\|\bar{x}\|\|\bar{y}\|)+2\sqrt{(x_0^2-\|\bar{x}\|^2)(y_0^2-\|\bar{y}\|^2)}} \\
&= \left(\sqrt{x_0+\|\bar{x}\|}-\sqrt{y_0+\|\bar{y}\|}\right)^2 + \left(\sqrt{x_0-\|\bar{x}\|}-\sqrt{y_0-\|\bar{y}\|}\right)^2 \qquad (2.3) \\
&\geq 0.
\end{aligned}
$$

This says that $\bar{x}^T\bar{y} = \|\bar{x}\|\|\bar{y}\|$, $\sqrt{x_0 + \|\bar{x}\|} = \sqrt{y_0 + \|\bar{y}\|}$, $\sqrt{x_0 - \|\bar{x}\|} = \sqrt{y_0 - \|\bar{y}\|}$, which leads to $\bar{x} = \bar{y}$. Combining with $\sqrt{x_0 + \|\bar{x}\|} = \sqrt{y_0 + \|\bar{y}\|}$, this yields that $x_0 = y_0$. Then, it follows that $x = y$.

(c) From the expression of $\mathrm{d}(x, y)$, i.e.,

$$\mathrm{d}(x,y) = 2(x_0 + y_0) - 2\sqrt{2(x_0y_0 + \bar{x}^T\bar{y}) + 2\sqrt{(x_0^2 - \|\bar{x}\|^2)(y_0^2 - \|\bar{y}\|^2)}},$$

it is obvious that $\mathrm{d}(x, y) = \mathrm{d}(y, x)$. □

Theorem 2.6. *Let the function* $\mathrm{d}(\cdot,\cdot)$ *be defined by (2.1). Then, for any a fixed* x *or* y*,* $\mathrm{d}(x, y)$ *is convex.*

Proof. The arguments are similar to those in Theorem 2.4. We omit them here. □

In Theorem 2.5, we have shown that the function $\mathrm{d}(\cdot,\cdot)$ on $\mathcal{K}^n$ satisfies (D1)-(D3), and hence it is a semi-distance on the second-order cone $\mathcal{K}^n$. However, in light of the convexity of d (Theorem 2.4), we can verify that the triangle inequality fails. To see this, given any x, y in symmetric cone $\mathcal{K}$, taking $z = \lambda x + (1 - \lambda)y$, $0 < \lambda < 1$, we have

$$\mathrm{d}(x,z) = \mathrm{d}(x, \lambda x + (1-\lambda)y) \le \lambda \mathrm{d}(x,x) + (1-\lambda)\mathrm{d}(x,y) = (1-\lambda)\mathrm{d}(x,y), \tag{2.4}$$

$$\mathrm{d}(z,y) = \mathrm{d}(\lambda x + (1-\lambda)y, y) \le \lambda \mathrm{d}(x,y) + (1-\lambda)\mathrm{d}(y,y) = \lambda \mathrm{d}(x,y). \tag{2.5}$$

Then, adding (2.4) and (2.5) together yields

$$\mathrm{d}(x,z) + \mathrm{d}(z,y) \le \mathrm{d}(x,y).$$

Moreover, in the setting of second-order cone, we hereby provide an example to explain that the inequality holds strictly, and consequently the triangle inequality fails. Consider $x = (1, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, $y = (1, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$, and $z = \frac{1}{2}(x+y) = (1,0,0)$. It is clear that $x, y, z \in \mathcal{K}^3$, and

$$\mathrm{d}(x,z) + \mathrm{d}(z,y) = (4 - 2\sqrt{2}) + (4 - 2\sqrt{2}) < 4 = \mathrm{d}(x,y).$$

3. A new proximal distance with respect to $\mathrm{int}(\mathcal{K}^n)$

In this section, we construct a new proximal distance associated with second-order cone based on the aforementioned semi-distance. To proceed, we first present the definition of proximal distance, which was introduced by Auslender and Teboulle [1]. For more details of its properties, please refer to [1].

Definition 3.1. A function $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ is called proximal distance with respect to an open nonempty convex set $C \subset \mathbb{R}^n$ if for each $y \in C$ it satisfies the following properties:

(P1) $d(\cdot, y)$ is proper, l.s.c., convex, continuously differentiable on C;

(P2) $\operatorname{dom} d(\cdot, y) \subset \bar{C}$ and $\operatorname{dom}\partial_1 d(\cdot, y) = C$, where $\partial_1 d(\cdot, y)$ denotes the classical subgradient map of the function $d(\cdot, y)$ with respect to the first variable;
(P3) $d(\cdot, y)$ is level bounded on $\mathbb{R}^n$ i.e., $\lim_{\|u\|\to+\infty} d(u, y) = +\infty$;
(P4) $d(y, y) = 0$.

We denote by $\mathcal{D}(C)$ the family of functions d satisfying Definition 3.1. Property (P1) is needed to preserve convexity of $d(\cdot, y)$, (P2) will force the iterate x^k to stay in C, and (P3) is used to guarantee the existence of such an iterate. For each $y \in C$, let $\nabla_1 d(\cdot, y)$ denote the gradient map of the function $d(\cdot, y)$ with respect to the first variable. Note that by definition $d(\cdot, \cdot) \geq 0$, and from (P4) the global minimum of $d(\cdot, y)$ is obtained at y, which shows that $\nabla_1 d(y, y) = 0$.

Definition 3.2. Given $C \subset \mathbb{R}^n$, open and convex, and $d \in \mathcal{D}(C)$, a function $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ is called the induced proximal distance to d if H is finite valued on $C \times C$ and for each $a, b \in C$ satisfies

(H1) $H(a, a) = 0$;
(H2) $\langle c - b, \nabla_1 d(b, a)\rangle \leq H(c, a) - H(c, b)$ for each $c \in C$.

We write $(d, H) \in \mathcal{F}(C)$ to quantify the triple $[C, d, H]$ that satisfies the premises of Definition 3.2. We also denote $(d, H) \in \mathcal{F}(\bar{C})$ if there exists H such that

(H3) H is finite valued on $\bar{C} \times C$ satisfying (H_1) and (H_2), for each $c \in \bar{C}$;
(H4) For each $c \in \bar{C}$, $H(c, \cdot)$ has bounded level set on C.

Finally, we write $(d, H) \in \mathcal{F}_+(\bar{C})$ if $(d, H) \in \mathcal{F}(\bar{C})$ and

(H5) for all $y \in \bar{C}$ and any $\{y_k\} \subset C$ bounded with $\lim_{k\to+\infty} H(y, y_k) = 0$, then $\lim_{k\to+\infty} y_k = y$;
(H6) for all $y \in \bar{C}$ and any $\{y_k\} \subset C$ converges to y, we have $\lim_{k\to+\infty} H(y, y_k) = 0$.

Clearly, we have $\mathcal{F}_+(\bar{C}) \subset \mathcal{F}(\bar{C}) \subset \mathcal{F}(C)$. According to the proximal distances, Auslender and Teboulle [1] also proposed some algorithms and derived global convergence. In particular, they demonstrated several examples of proximal distances, including Bregman distances, proximal distances based on φ-divergence, self-proximal distances, and distances based on second order homogeneous proximal distances, for more details, please see [1, Section 3].

In view of the distance function d defined as in (2.1), we now construct a new type of proximal distance which is different from the ones given in [1]. To our best knowledge, it may be the only proximal distance which is not induced from Bregman distance, φ-divergence, or distance-like entropy function. For any $x, y \in \mathbb{R}^n$, we define $\mathbf{d} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ by

$$\mathbf{d}(x, y) := \begin{cases} \operatorname{tr}(x + y) - 2\operatorname{tr}\left(P(x^{\frac{1}{2}})y\right)^{\frac{1}{2}} & \forall x \in \operatorname{int}(\mathcal{K}^n),\ y \in \mathcal{K}^n, \\ +\infty & \text{otherwise}. \end{cases} \tag{3.1}$$

This function, as will be shown below, is a proximal distance with respect to $\mathrm{int}(\mathcal{K}^n)$. First, we note that the function $\mathbf{d}$ satisfies Property (P4) by Theorem 2.5. Next, we discuss the differentiability of $\mathbf{d}$ with respect to the first variable.

Proposition 3.3. *Let* $\mathbf{d}$ *be defined as in (3.1). Then, for any* $y \in \mathrm{int}(\mathcal{K}^n)$, $\mathbf{d}(\cdot, y)$ *is continuously differentiable on* $\mathrm{int}(\mathcal{K}^n)$.

Proof. Given any $x \in \mathrm{int}(\mathcal{K}^n)$, let $x := (x_0, \bar{x}), y := (y_0, \bar{y}) \in \mathbb{R} \times \mathbb{R}^{n-1}$, we know that

$$\mathbf{d}(x, y) = 2(x_0 + y_0) - 2\sqrt{2(x_0 y_0 + \bar{x}^T \bar{y}) + 2\sqrt{(x_0^2 - \|\bar{x}\|^2)(y_0^2 - \|\bar{y}\|^2)}}.$$

Here we use expression (2.2), and $x_0 y_0 + \bar{x}^T \bar{y} > 0$, $x_0 > \|\bar{x}\|$, $y_0 > \|\bar{y}\|$ so that the terms inside the square root is positive. Besides, note that the following three functions

$$\begin{aligned} x &\mapsto x_0, \\ x &\mapsto x_0 y_0 + \bar{x}^T \bar{y}, \\ x &\mapsto x_0^2 - \|\bar{x}\|^2, \end{aligned}$$

are continuously differentiable on $\mathbb{R}^n$, and the function $t \mapsto \sqrt{t}$ is continuously differentiable on $\mathbb{R}_{++}$. Thus, we conclude that the distance function $\mathbf{d}(\cdot, y)$ is also continuously differentiable on $\mathrm{int}(\mathcal{K}^n)$ because it is a composition of continuously differentiable functions.

Moreover, by the direct calculation, we have

$$\nabla_1 \mathbf{d}(x, y) = 2e - \frac{2}{\sqrt{2\langle x, y\rangle + 2\sqrt{\det(x)\det(y)}}} \left(y + \frac{\det(y)}{\sqrt{\det(x)\det(y)}} \begin{bmatrix} x_0 \\ -\bar{x} \end{bmatrix} \right)$$

where $e = (1, 0) \in \mathbb{R} \times \mathbb{R}^{n-1}$, $\det(x) = x_0^2 - \|\bar{x}\|^2$, and $\det(y) = y_0^2 - \|\bar{y}\|^2$. Then, the proof is complete. $\square$

For any $y \in \mathrm{int}(\mathcal{K}^n)$, it is clear that $\mathbf{d}(\cdot, y)$ is proper, lower semicontinuous on $\mathrm{int}(\mathcal{K}^n)$ by the definition of $\mathbf{d}$. In addition, it is convex and continuously differentiable on $\mathrm{int}(\mathcal{K}^n)$ by Theorem 2.6 and Proposition 3.3, respectively. Thus, the distance function $\mathbf{d}$ satisfies the Property (P1).

Remark 3.4. Proposition 3.3 still holds for any symmetric cone without assuming operator commute since

$$\mathrm{d}(x, y) = \mathrm{d}(y, x) = \mathrm{tr}(y + x) - 2\,\mathrm{tr}\left(P(y^{\frac{1}{2}})x\right)^{\frac{1}{2}},$$

the map $x \mapsto P(y^{\frac{1}{2}})x$ is smooth on the cone, the map $\mathrm{tr}(\cdot)^{\frac{1}{2}}$ is smooth on the interior of the cone by [7, Lemma 3.1], and trace functional is the inner product with the Jordan identity.

Proposition 3.5. *Let* $\mathbf{d}$ *be defined as in (3.1). Then, the distance function* $\mathbf{d}$ *satisfies the Property (P2).*

Proof. For any $y \in \text{int}(\mathcal{K}^n)$, it is clear that $\text{dom}\,\mathbf{d}(\cdot, y) = \text{int}(\mathcal{K}^n) \subset \mathcal{K}^n$ from the definition of $\mathbf{d}$. In addition, since $\mathbf{d}(\cdot, y)$ is convex and continuously differentiable on $\text{int}(\mathcal{K}^n)$, applying [24, Theorem 25.1] gives $\partial_1 \mathbf{d}(x, y) = \{\nabla_1 \mathbf{d}(x, y)\} \neq \emptyset$ for any $x \in \text{int}(\mathcal{K}^n)$. Thus, it remains to show that $\partial_1 \mathbf{d}(x, y) = \emptyset$ for any $x \in \text{bd}(\mathcal{K}^n)$. Indeed, there is no $\xi \in \mathbb{R}^n$ satisfying

$$\mathbf{d}(z, y) \geq \mathbf{d}(x, y) + \langle \xi, z - x \rangle \quad \forall z \in \mathbb{R}^n$$

because $\mathbf{d}(x, y) = +\infty$. Therefore, we conclude $\text{dom}\,\partial_1 \mathbf{d}(\cdot, y) = \text{int}(\mathcal{K}^n)$. $\square$

Proposition 3.6. *Let* $\mathbf{d}$ *be defined as in (3.1). Then, the distance function* $\mathbf{d}$ *satisfies the Property (P3).*

Proof. Suppose $y \in \text{int}(\mathcal{K}^n)$, we first note that for $u \in \text{int}(\mathcal{K}^n)$, $u_0 + \|\bar{u}\| \geq \|u\|$, and from inequality (2.3),

$$\begin{aligned} \mathbf{d}(u, y) &\geq \left(\sqrt{u_0 + \|\bar{u}\|} - \sqrt{y_0 + \|\bar{y}\|}\right)^2 + \left(\sqrt{u_0 - \|\bar{u}\|} - \sqrt{y_0 - \|\bar{y}\|}\right)^2 \\ &\geq \left(\sqrt{u_0 + \|\bar{u}\|} - \sqrt{y_0 + \|\bar{y}\|}\right)^2, \end{aligned}$$

we have

$$\mathbf{d}(u, y) \geq \left(\sqrt{\|u\|} - \sqrt{y_0 + \|\bar{y}\|}\right)^2 \tag{3.2}$$

whenever $\sqrt{\|u\|} \geq \sqrt{y_0 + \|\bar{y}\|}$. In fact, the inequality (3.2) holds for all $u \in \mathbb{R}^n$ since $\mathbf{d}(u, y) = +\infty$ if $u \notin \text{int}(\mathcal{K}^n)$. Then, we conclude that $\mathbf{d}(u, y) \to +\infty$ as $\|u\| \to +\infty$, that is, $\mathbf{d}(\cdot, y)$ is level bounded on $\mathbb{R}^n$. $\square$

Theorem 3.7. *Let* $\mathbf{d}$ *be defined as in (3.1). Then, the function* $\mathbf{d}$ *is a proximal distance with respect to* $\text{int}(\mathcal{K}^n)$*.*

Proof. This is an immediate consequence of Props. 3.3-3.6. $\square$

4. Concluding remarks

In this paper, we propose a semi-distance associated with symmetric cone, which is further proved a proximal distance with respect to $\text{int}(\mathcal{K}^n)$. However, we do not know whether or not the distance function $\mathbf{d}$ can become a proximal distance with respect to symmetric cone. The main difficulty is that the differentiability of eigenvalues $\lambda_i(x)$ of x associated with symmetric cone is unknown yet if we regard $\lambda_i(x)$ as a function of x.

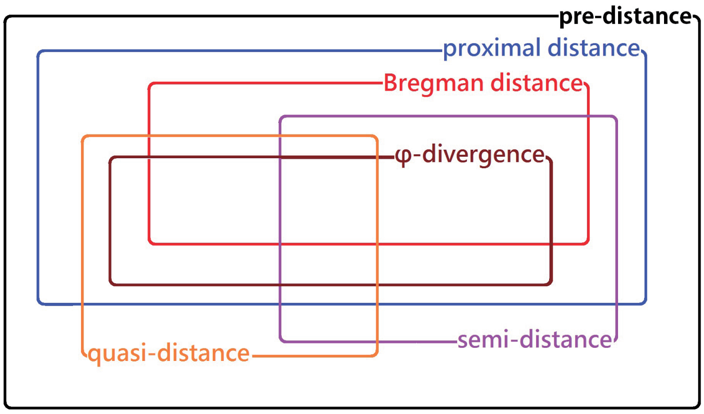

FIGURE 1. Relationship between these distance-like function.

There are a few future research directions based on the new discovery of the proximal distance $\mathbf{d}$ in this paper. For example, consider the following convex second-order cone programming (CSOCP):

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in \mathcal{K}^n, \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper, l.s.c., convex function. Pan and Chen [22] proposed a class of interior proximal-like algorithms for the (CSOCP) via distance-like functions. One direction is employing this proximal distance $\mathbf{d}$ to the proximal-like algorithms and doing the numerical comparison. Of course, analyzing the convergence rate is also desirable. There are some other questions to be answered in the future:

- Is the distance function $\mathbf{d}$ also a proximal distance in the setting of symmetric cone (without assuming operator commute)?
- Can the distance function $\mathbf{d}$ further become a Bregman distance or φ-divergence? In other words, does there exist a ϕ or φ so that $\mathbf{d}$ corresponds to $\mathbf{d}_\phi$ or $\mathbf{d}_\varphi$?
- Can the distance function $\mathbf{d}$ be extended to nonsymmetric cone setting? In particular, for circular cone $\mathcal{L}_\theta$, we have already known one type of spectral decomposition of x and some differentiabilities of $\lambda_i(x)$, see [28]. By using these facts, we may consider to construct an analogous distance function $\mathbf{d}$ in the setting of circular cone.

We leave all the above as our future works.

REFERENCES

[1] A. Auslender and M. Teboulle, *Interior gradient and proximal methods for convex and conic optimization*, SIAM J. Optim. **16** (2006), 697–725.

[2] R. Bhatia, T. Jain, and Y. Lim, *On the Bures-Wasserstein distance between positive definite matrices*, Expo. Math. **37** (2019), 165–191.

[3] A. Banerjee, I. Dhillon, and J. Ghosh, *Clustering with Bregman divergences*, J. Mach. Learn. Res. **6** (2005), 1705–1749.

[4] A. Banerjee, I. Dhillon, J. Ghosh, S. Merugu, and D. Modha, *A generalized maximum Entropy approach to Bregman co-clustering and matrix approximations*, J. Mach. Learn. Res. **6** (2007), 1919–1986.

[5] L. M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Comput. Math. Math. Phys. **7** (1967), 200–217.

[6] Y. Censor and A. Zenios, *Proximal minimization algorithms with D-functions*, J. Optim. Theory Appl. **73** (1992), 451–464.

[7] Y.-L. Chang and J-S. Chen, *Convexity of symmetric cone trace functions in Euclidean Jordan algebras*, J. Nonlinear Convex Anal. **14** (2013), 53–61.

[8] J.-S. Chen, *SOC Functions and Their Applications*, Springer Optimization and Its Applications 143, Springer, Singapore, 2019.

[9] J.-S. Chen and S. Pan, *Proximal-like algorithm using the quasi D-function for convex second-order cone programming*, J. Optim. Theory Appl. **138** (2008), 95–113.

[10] G. Chen and M. Teboulle, *Convergence analysis of proximal-like minimization algorithm using Bregman functions*, SIAM J. Optim. **3** (1993), 538–543.

[11] J.-S. Chen, X. Chen, P. Tseng, *Analysis of nonsmooth vector-valued functions associated with second-order cones*, Math. Program. **101** (2004), 95–117.

[12] I. Dhillon and S. Sra, *Generalized nonnegative matrix approximations with Bregman divergences*, Advanced in Neural Information Processing Systems (2006), 283–290.

[13] S. Dhillon and J. A. Tropp, *Matrix nearness problems with Bregman divergences*, SIAM J. Matrix Anal. Appl. **29** (2007), 1120–1146.

[14] M. Fukushima, Z.-Q. Luo, and P. Tseng, *Smoothing functions for second-order-cone complimentarity problems*, SIAM J. Optim. **12** (2002), 436–460.

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

[16] C. R. Givens, R. M. Shortt, *A class of Wasserstein metrics for probability distributions*, Mich. Math. J. **31** (2014), 231–240.

[17] G. J. Gordon, *Generalized linear models*, Advances in Neural Information Processing Systems (2002), 577–584.

[18] K. C. Kiwiel, *Proximal minimization methods with generalized Bregman functions*, SIAM J. Control Optim. **35** (1997), 1142–1168.

[19] M. Knott and C. Smith, *On the optimal mapping of distributions*, J. Optim. Theory Appl. **43** (1984), 39–49.

[20] B. Kulis, M. Sustik, and I. Dhillon, *Low-Rank Kernel Learning with Bregman Matrix Divergences*, J. Mach. Learn. Res. **10** (2009), 341–376.

[21] M. Malik, S. R. Mohan, *On Q and R_0 properties of a quadratic representation in linear complementarity prolems over the second-order cone*, Linear Algebra Appl. **397** (2005), 85–97.

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

[23] E. A. Papa Quiroz, *An extension of the proximal point algorithm with Bregman distances on Hadamard manifolds*, J. Global Optim. **56** (2013), 43–59.

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

[25] W. Stummer and I. Vajda, *On Bregman distances and divergences of probability measures*, IEEE T. Inform. Theory **58** (2012), 1277–1288.

[26] M. Teboulle, *Entropic proximal mappings with applications to nonlinear programming*, Math. Oper. Res. **17** (1992), 670–690.

[27] K. Tsuda, G. Ratsch, and M. K. Warmuth, *Matrix exponentiated gradient updates for on-line learning and Bregman projection*, J. Mach. Learn. Res. **6** (2005), 995–1018.

[28] J. C. Zhou, J. Y. Tang, and J.-S. Chen, *Parabolic second-order directional differentiability in the Hadamard sense of the vector-valued functions associated with circular cones*, J. Optim. Theory Appl. **172** (2017), 802–823.

Manuscript received 2 October 2019

X.-H. Miao
School of Mathematics, Tianjin University, Tianjin 300072, China
E-mail address: `xinhemiao@tju.edu.cn.`

C.-H. Huang
General Education Center, Wenzao Ursuline University of Languages, Kaohsiung 80793, Taiwan
E-mail address: `qqnick0719@yahoo.com.tw`

Y. Lim
Department of Mathematics, Sungkyunkwan University, Suwon, Korea
E-mail address: `ylim@skku.edu`

J.-S. Chen
Department of Mathematics, National Taiwan Normal University, Taipei 11677, Taiwan
E-mail address: `jschen@math.ntnu.edu.tw`