Nonlinear Functional Analysis and Applications
Vol. 26, No. 4 (2021), pp. 649-662
ISSN: 1229-1595(print), 2466-0973(online)

https://doi.org/10.22771/nfaa.2021.26.04.01
http://nfaa.kyungnam.ac.kr/journal-nfaa

COMPUTATION OF DIVERGENCES AND MEDIANS IN SECOND ORDER CONES

Sangho Kum

Department of Mathematics Education, College of Education
Chungbuk National University, Cheongju 28644, Korea
e-mail: shkum@cbnu.ac.kr

Abstract. Recently the author studied a one-parameter family of divergences and considered the related median minimization problem of finite points over these divergences in general symmetric cones. In this article, to utilize the results practically, we deal with a special symmetric cone called second order cone, which is important in optimization fields. To be more specific, concrete computations of divergences with its gradients and the unique minimizer of the median minimization problem of two points are provided skillfully.

1. Divergence and Median

A divergence, which measures discrepancy between two points, plays a crucial role in many problems such as information theory, statistics, optimization, computational vision, and neural networks [1, 2, 3]. A divergence Φ on the Riemannian manifold M is a real valued function $\Phi : M \times M \to \mathbb{R}$ which satisfies

(D1) $\Phi(a, b) \geq 0$ for all $a, b \in M$ with equality if and only if $a = b$;
(D2) the first derivative $D\Phi$ with respect to the second variable vanishes on the diagonal;
$$D\Phi(a, x)|_{x=a} = 0,$$
(D3) its Hessian is positive definite on the diagonal;
$$D^2\Phi(a, x)|_{x=a}(y, y) \geq 0 \quad \text{for all } a \in \Omega,\ y \in V. \tag{1.1}$$

[0]Received December 29, 2020. Revised June 16, 2021. Accepted July 2, 2021.
[0]2010 Mathematics Subject Classification: 17C20, 15B48, 90C25.
[0]Keywords: Symmetric cone, Euclidean Jordan algebra, divergence, median, second order cone, Wasserstein barycenter.

The most familiar examples include Kullback-Leibler divergence defined between two probability distributions, Bregman divergence in optimization and signal processing, etc. Indeed, a divergence is almost a distance function except the symmetry with respect to its arguments and the triangle inequality. For instance, the square of a distance function is a (symmetric) divergence. Thus a divergence Φ gives rise to an important optimization problem on the Riemannian manifold M :

$$\arg\min_{x\in M} \sum_{j=1}^{m} w_j \Phi(a_j, x), \tag{1.2}$$

where $a_1, \ldots, a_m \in M$ and $\omega = (w_1, \ldots, w_m) \in \mathbb{R}^m$ is a positive probability vector. This minimization problem looks like a lease square problem in some sense. So a minimizer whenever it exists provides alternatively a barycenter or averaging on M, which is called the *ω-weighted Φ-median of* $a_1, \ldots, a_m$.

In a recent work [10], this median optimization problem on a special Riemannian manifold called symmetric cones is studied. The main result in [10] is briefly summarized as follow: Let V be a Euclidean Jordan algebra and let Ω be the symmetric cone (see section 2 for basic facts regarding Euclidean Jordan algebras and symmetric cones). Consider the function $\Phi_t : \Omega \times \Omega \to \mathbb{R}$ is defined by

$$\Phi_t(a, b) = \operatorname{tr}((1-t)a + tb) - \operatorname{tr}\left(P(a^{\frac{1-t}{2t}})b\right)^t, \qquad 0 < t < 1 \tag{1.3}$$

where tr is the trace functional and P is the quadratic representation of V. Then we have:

Theorem 1.1. ([10]) *For every $0 < t < 1$, Φ_t is a divergence on Ω. Moreover, for every m-tuple $\mathbf{t} = (t_1, \ldots, t_m) \in (0,1)^m$, the minimization problem:*

$$\arg\min_{x\in \Omega} \sum_{j=1}^{m} w_j \Phi_{t_j}(a_j, x) \tag{1.4}$$

has a unique minimizer. For the case $t := t_1 = t_2 = \cdots = t_m$, (1.4) *reduces to* (1.2)*.*

Besides, the following gradient formula is also derived:

$$\nabla_x \Phi_t(a, x) = t\left(e - \left(a^{\frac{1-t}{t}} \#_{1-t} x^{-1}\right)\right). \tag{1.5}$$

(See (2.3) for the definition of (1.5).) A meaningful reason to take the divergence (1.3) into account comes from the followings: The term $F_t(a, b) = \operatorname{tr}\left(P(a^{\frac{1-t}{2t}})b\right)^t$ in (1.3) is known as *sandwiched quasi-relative entropy* in the theory of quantum information; for positive (semi)definite matrices A and B,

$$F_t(A,B) = \operatorname{tr}\left(A^{\frac{1-t}{2t}} B A^{\frac{1-t}{2t}}\right)^t, \quad t \in (0,1). \tag{1.6}$$

This is a parameterized version of the *fidelity* $F_{\frac{1}{2}}(A,B) = \operatorname{tr}\left(A^{\frac{1}{2}} B A^{\frac{1}{2}}\right)^{\frac{1}{2}}$. Fidelity and sandwiched quasi-relative entropies play an essential role in quantum information theory and quantum computation [6, 14, 15, 16]. In addition, the *Bures distance* in quantum information is defined by

$$d_W(A,B) = \left[\frac{\operatorname{tr}(A+B)}{2} - \operatorname{tr}\left(A^{\frac{1}{2}} B A^{\frac{1}{2}}\right)^{\frac{1}{2}}\right]^{\frac{1}{2}},$$

which is also known as the *Wasserstein distance* in statistics and the theory of optimal transport [5, 8, 12]. Clearly, $d_W^2(A,B) = \Phi_{\frac{1}{2}}(A,B)$. This implies that the divergence (1.3) may have a rich background in various areas mentioned above.

2. Euclidean Jordan algebras and symmetric cones

Before stating the motivation of this work, first we briefly describe (following mostly [7], [10]) some Jordan-algebraic concepts relevant to our purpose. A Jordan algebra V over $\mathbb{R}$ is a commutative algebra satisfying $x^2(xy) = x(x^2y)$ for all $x, y \in V$. For $x \in V$, let $L(x)$ be the linear operator defined by $L(x)y = xy$, and let $P(x) = 2L(x)^2 - L(x^2)$. The map P is called the quadratic representation of V. An element $x \in V$ is said to be invertible if there exists an element y (denoted by $y = x^{-1}$) in the subalgebra generated by x and e (the Jordan identity) such that $xy = e$. In this case, $P(x)^{-1} = P(x^{-1})$ [7, Proposition II.3.1].

An element $c \in V$ is called an idempotent if $c^2 = c \neq 0$. We say that $c_1, \ldots, c_k$ is a complete system of orthogonal idempotents if $c_i^2 = c_i, c_ic_j = 0, i \neq j, c_1 + \cdots + c_k = e$. An idempotent is said to be primitive if it is non-zero and cannot be written as the sum of two non-zero idempotents. A Jordan frame is a complete system of orthogonal primitive idempotents.

A finite-dimensional Jordan algebra V with an identity element e is said to be *Euclidean* if there exists an inner product $\langle \cdot, \cdot \rangle$ such that $\langle xy, z \rangle = \langle y, xz \rangle$ for all $x, y, z \in V$.

Theorem 2.1. (Spectral theorem, first version [7, Theorem III.1.1]) *Let V be a Euclidean Jordan algebra. Then for $x \in V$, there exist real numbers $\lambda_1, \ldots, \lambda_k$ all distinct and a unique complete system of orthogonal idempotents $c_1, \ldots, c_k$ such that*

$$x = \sum_{i=1}^{k} \lambda_i c_i. \tag{2.1}$$

The numbers λ_i are called the eigenvalues and (2.1) *is called the spectral decomposition of x.*

Theorem 2.2. (Spectral theorem, second version [7, Theorem III.1.2]) *Any two Jordan frames in a Euclidean Jordan algebra V have the same number of elements (called the rank of V, denoted by $\operatorname{rank}(V)$). Given $x \in V$, there exists a Jordan frame $c_1, \ldots, c_r$ and real numbers $\lambda_1, \ldots, \lambda_r$ such that*

$$x = \sum_{i=1}^{r} \lambda_i c_i.$$

The numbers λ_i (with their multiplicities) are uniquely determined by x.

Definition 2.3. Let V be a Euclidean Jordan algebra of $\operatorname{rank}(V) = r$. The spectral mapping $\lambda : V \to \mathbb{R}^r$ is defined by $\lambda(x) = (\lambda_1(x), \ldots, \lambda_r(x))$, where $\lambda_i(x)$'s are eigenvalues of x (with multiplicities) as in Theorem 2.2 in non-increasing order $\lambda_{\max}(x) = \lambda_1(x) \geq \lambda_2(x) \geq \cdots \geq \lambda_r(x) = \lambda_{\min}(x)$. Furthermore, $\det(x) = \prod_{i=1}^{r} \lambda_i(x)$ and $\operatorname{tr}(x) = \sum_{i=1}^{r} \lambda_i(x)$.

Let Q be the set of all square elements of V. Then Q is a closed convex cone of V with $Q \cap -Q = \{0\}$, and is the set of elements $x \in V$ such that $L(x)$ is positive semi-definite. It turns out that Q has non-empty interior $\Omega := \operatorname{int}(Q)$, and Ω is a symmetric cone, that is, the group $G(\Omega) = \{g \in \mathrm{GL}(V) \mid g(\Omega) = \Omega\}$ acts transitively on it and Ω is a self-dual cone with respect to the inner product $\langle \cdot, \cdot \rangle$ (see [7]). Note that $\overline{\Omega} = \{x \in V \mid \lambda_i(x) \geq 0, i = 1, \ldots, r\}$.

On the other hand, the symmetric cone Ω in V admits a $G(\Omega)$-invariant Riemannian metric defined by

$$\langle u, v \rangle_x = \langle P(x)^{-1} u, v \rangle, \ x \in \Omega, \ u, v \in V. \tag{2.2}$$

So Ω is a Riemannian manifold [7]. It is shown in [11, Proposition 2.6] that the unique geodesic joining a and b is

$$t \mapsto a \#_t b := P(a^{1/2})(P(a^{-1/2})b)^t, \tag{2.3}$$

where $a^t = \sum_{j=1}^{r} \lambda_j(a)^t c_j$ for the spectral decomposition $a = \sum_{j=1}^{r} \lambda_j(a) c_j$ in Theorem 2.2. *The geometric mean of a and b* is defined to be $a \# b := a \#_{1/2} b$, which is a unique geodesic middle between a and b.

3. Motivation

To utilize Theorem 1.1 practically, it is necessary to consider more tangible settings. To this end, we pay attention to the typical symmetric cones in optimization fields.

• The Euclidean space of all symmetric matrices of fixed size with the Jordan product $X \circ Y := (1/2)(XY + YX)$ and the trace inner product is a standard example of Euclidean Jordan algebra. In this case, the corresponding symmetric cone Ω is the cone of positive definite matrices $\mathbb{P}_n$.

• The Euclidean space $\mathbb{R}^n$ (*as column vectors*) with the Jordan product defined by

$$x \circ y = (\langle x, y\rangle,\ x_1 y_2 + y_1 x_2)$$

is a Euclidean Jordan algebra equipped with the standard inner product $\langle \cdot, \cdot \rangle$ where $x = (x_1, x_2),\ y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$. In this case, the corresponding symmetric cone is the second order cone (simply, SOC)

$$\mathcal{K} := \left\{(x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid \|x_2\| < x_1\right\}.$$

After getting Theorem 1.1, it is natural to raise the question: What is the unique minimizer? In other words, can we obtain an explicit formula of the minimizer? In fact, any closed form of it is not provided yet except the special case for $m = 2$ and $t_1 = t_2 = 1/2$. The formula on the general symmetric cone Ω is derived in [10] as follow:

Theorem 3.1. ([10]) *Let $a, b \in \Omega$ and $0 < s < 1$. Then the unique minimizer of* (1.4) *is*

$$\begin{aligned} W_{\frac{1}{2}}(1-s, s; a, b) &= P\left(s(a^{-1} \# b) + (1-s)e\right) a \\ &= P(a^{-1/2})\left((1-s)a + s(P(a^{1/2})b)^{1/2}\right)^2. \end{aligned}$$

Indeed, in the case of $\mathbb{P}_n$, the formula is already computed in [4] as

$$W_{\frac{1}{2}}(1-s, s; A, B) = (1-s)^2 A + s^2 B + s(1-s)[(AB)^{1/2} + (BA)^{1/2}]. \quad (3.1)$$

Especially, when $s = 1/2$, we call it the *Wasserstein mean (or barycenter) of A and B*.

Then *how about the second order cone case? What is the Wasserstein-type barycenter (or mean) of a and b, that is, the $\Phi_{1/2}$-median of a and b in $\mathcal{K}$? In addition, what are the divergence* (1.3) *and its gradient?* (Gradients are crucial for numerical implementations.) These questions are direct motivations of this article. In general, a concrete computation of the minimizer is not so simple even in the case of SOC.

In the next section, we deal with those questions and provide explicit formulas skillfully.

4. Second order cones

Note that for each $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, the eigenvalues of x are given by

$$\lambda_i(x) = x_1 + (-1)^i \|x_2\|, \quad i = 1,\, 2.$$

Hence the *determinant* and the *trace* of x are written as

$$\det(x) = x_1^2 - \|x_2\|^2, \quad \mathrm{tr}(x) = 2x_1.$$

In addition, x is said to be *invertible* if $\det(x) \neq 0$. In this case, x has a unique inverse y in the sense that $x \circ y = y \circ x = e$ where $e = (1, 0, \cdots, 0)$ is the unit element of the Euclidean Jordan algebra $V = (\mathbb{R}^n, \circ)$. In fact, we have

$$y = x^{-1} = \frac{1}{x_1^2 - \|x_2\|^2}(x_1, -x_2) = \frac{1}{\det(x)}(x_1, -x_2). \tag{4.1}$$

Moreover, for $x \in \mathcal{K}$, there exists a unique $x^{1/2} \in \mathcal{K}$ represented by

$$x^{1/2} = \left(\kappa,\, \frac{x_2}{2\kappa}\right) \ \text{ where } \ \kappa = \sqrt{\left(x_1 + \sqrt{\det(x)}\right)/2}. \tag{4.2}$$

From the definition of Jordan product, it is seen [13] that

$$L(x) = \begin{pmatrix} x_1 & x_2^T \\ x_2 & x_1 I \end{pmatrix} \quad \text{and} \quad P(x) = \begin{pmatrix} \|x\|^2 & 2x_1 x_2^T \\ 2x_1 x_2 & \det(x) I + 2x_2 x_2^T \end{pmatrix}. \tag{4.3}$$

Lemma 4.1. *For $a = (a_1, a_2)$ $(a_2 \neq 0) \in \mathcal{K}$,*

$$P(a^{\frac{1}{2}}) = \begin{pmatrix} a_1 & a_2^T \\ a_2 & \sqrt{\det(a)} I + \left(a_1 - \sqrt{\det(a)}\right) \frac{a_2 a_2^T}{\|a_2\|^2} \end{pmatrix}. \tag{4.4}$$

Thus, for $x \in \mathcal{K}$,

$$P(a^{\frac{1}{2}})x = \begin{pmatrix} \langle a,\, x\rangle \\ \sqrt{\det(a)} x_2 + \left(x_1 + \left(a_1 - \sqrt{\det(a)}\right) \frac{\langle a_2, x_2\rangle}{\|a_2\|^2}\right) a_2 \end{pmatrix}. \tag{4.5}$$

Proof. By (4.2) and (4.3), we get

$$\begin{aligned} P(a^{\frac{1}{2}}) &= \begin{pmatrix} \|a^{\frac{1}{2}}\|^2 & a_2^T \\ a_2 & \sqrt{\det(a)} I + \frac{a_2 a_2^T}{2\kappa^2} \end{pmatrix} \\ &= \begin{pmatrix} \|a^{\frac{1}{2}}\|^2 & a_2^T \\ a_2 & \sqrt{\det(a)} I + \frac{a_2 a_2^T}{a_1 + \sqrt{\det(a)}} \end{pmatrix} \\ &= \begin{pmatrix} \|a^{\frac{1}{2}}\|^2 & a_2^T \\ a_2 & \sqrt{\det(a)} I + \left(a_1 - \sqrt{\det(a)}\right) \frac{a_2 a_2^T}{\|a_2\|^2} \end{pmatrix}. \end{aligned}$$

Also

$$\begin{aligned}\|a^{\frac{1}{2}}\|^2 &= s^2 + \frac{\|a_2\|^2}{4\kappa^2}\\ &= \frac{a_1+\sqrt{\det(a)}}{2} + \frac{\|a_2\|^2}{2(a_1+\sqrt{\det(a)})}\\ &= \frac{(a_1+\sqrt{\det(a)})^2+\|a_2\|^2}{2(a_1+\sqrt{\det(a)})}\\ &= a_1.\end{aligned}$$

This completes the proof. □

In the following lemma, we introduce a simple but clever way to get the formula of $\lambda_i\left(P(a^{\frac{1}{2}})x\right)$ without a direct computation.

Lemma 4.2. *For $a = (a_1, a_2)(a_2 \neq 0) \in \mathcal{K}$ and $x \in \mathcal{K}$,*

$$\lambda_i\left(P(a^{\frac{1}{2}})x\right) = \langle a,\, x\rangle + (-1)^i\sqrt{\langle a, x\rangle^2 - \det(a)\det(x)}, \quad i = 1,\, 2. \tag{4.6}$$

Proof. By the definition of eigenvalues, (4.5) and [7, Proposition III.4.2], we have

$$\begin{aligned}\lambda_1\left(P(a^{\frac{1}{2}})x\right) + \lambda_2\left(P(a^{\frac{1}{2}})x\right) &= 2\,\langle a, x\rangle,\\ \lambda_1\left(P(a^{\frac{1}{2}})x\right)\cdot\lambda_2\left(P(a^{\frac{1}{2}})x\right) &= \det\left(P(a^{\frac{1}{2}})x\right)\\ &= \left(\det(a^{\frac{1}{2}})\right)^2\det(x)\\ &= \det(a)\det(x).\end{aligned}$$

This implies that λ_1 and λ_2 are solutions of the elementary quadratic equation:

$$t^2 - 2\,\langle a,\, x\rangle t + \det(a)\det(x) = 0.$$

This yields the conclusion. □

Remark 4.3. Since for $i = 1,\ 2,$

$$\begin{aligned}\lambda_i\left(P(a^{\frac{1}{2}})x\right) &= \langle a,\, x\rangle\\ &+ (-1)^i\left\|\sqrt{\det(a)}x_2 + \left(x_1 + \left(a_1 - \sqrt{\det(a)}\right)\frac{\langle a_2,\, x_2\rangle}{\|a_2\|^2}\right)a_2\right\|,\end{aligned}$$

it follows from Lemma 4.2 that

$$\left\|\sqrt{\det(a)}x_2 + \left(x_1 + \left(a_1 - \sqrt{\det(a)}\right)\frac{\langle a_2,\, x_2\rangle}{\|a_2\|^2}\right)a_2\right\| = \sqrt{\langle a, x\rangle^2 - \det(a)\det(x)}.$$

Actually, this can be checked by a rather long calculation.

Theorem 4.4. *For $a \in \mathcal{K}$ and $x \in \mathcal{K}$,*

$$\lambda_i\left(P(a^{\frac{1}{2}})x\right) = \langle a,\, x\rangle + (-1)^i\sqrt{\langle a, x\rangle^2 - \det(a)\det(x)}, \quad i = 1,\, 2.$$

Proof. By Lemma 4.2, we have only to check the case $a_2 = 0$. Indeed, by (4.2) and (4.3) with the proof of Lemma 4.1, we get

$$P(a^{\frac{1}{2}}) = a_1 I, \quad P(a^{\frac{1}{2}})x = a_1 x = \begin{pmatrix} \langle a,\, x\rangle \\ a_1 x_2 \end{pmatrix}. \tag{4.7}$$

So, for $i = 1,\, 2$,

$$\begin{aligned} \lambda_i\left(P(a^{\frac{1}{2}})x\right) &= \langle a,\, x\rangle + (-1)^i\|a_1 x_2\| \\ &= \langle a,\, x\rangle + (-1)^i\sqrt{a_1^2x_1^2 - (a_1^2 - 0)(x_1^2 - \|x_2\|^2)} \\ &= \langle a,\, x\rangle + (-1)^i\sqrt{\langle a, x\rangle^2 - \det(a)\det(x)}. \end{aligned}$$

This completes the proof. □

As a direct consequence, we obtain

Corollary 4.5. *For $t = 1/2$ and $a \in \mathcal{K}$, the map $\Phi_{\frac{1}{2}}(a, \cdot) : \mathcal{K} \to \mathbb{R}$*

$$\begin{aligned} \Phi_{\frac{1}{2}}(a, x) &= \frac{1}{2}\operatorname{tr}(a + x) - \operatorname{tr}\left(P(a^{\frac{1}{2}})x\right)^{\frac{1}{2}} \\ &= a_1 + x_1 - \left(\langle a,\, x\rangle + \sqrt{\langle a, x\rangle^2 - \det(a)\det(x)}\right)^{\frac{1}{2}} \\ &\quad - \left(\langle a,\, x\rangle - \sqrt{\langle a, x\rangle^2 - \det(a)\det(x)}\right)^{\frac{1}{2}} \\ &= a_1 + x_1 - \sqrt{2}\left(\langle a,\, x\rangle + \sqrt{\det(a)\det(x)}\right)^{\frac{1}{2}}. \end{aligned}$$

Proof. The second equality immediately comes from Theorem 4.4 and the third one follows from (4.2) and the fact that the first coordinate of $P(a^{\frac{1}{2}})x$ is $\langle a,\, x\rangle$ (see (4.5) and (4.7)) and $\det\left(P(a^{\frac{1}{2}})x\right) = \det(a)\det(x)$ by [7, Proposition III.4.2]. □

For an estimation of the gradient $\nabla_x\Phi_{\frac{1}{2}}(a, x)$ as a particular case of (1.5), we adopt (6.7) in [9]:

$$a\#b = a\#_{\frac{1}{2}}b = \frac{1}{\sqrt{2}\sqrt{\alpha\beta + a_1b_1 - \langle a_2, b_2\rangle}}\begin{pmatrix} \beta a_1 + \alpha b_1 \\ \alpha b_2 + \beta a_2 \end{pmatrix}, \tag{4.8}$$

where $a = (a_1, a_2),\ b = (b_1, b_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ and $\alpha = \sqrt{\det(a)},\ \beta = \sqrt{\det(b)}$. Then we obtain the following from this and (1.5).

Theorem 4.6. *For $a,\ x \in \mathcal{K}$, we have*

$$\begin{aligned}\nabla_x \Phi_{\frac{1}{2}}(a, x) &= \frac{1}{2}(e - a\#x^{-1}) \\ &= \frac{1}{2}\left(e - \frac{1}{\sqrt{2}\sqrt{\langle a, x\rangle + \sqrt{\det(a)\det(x)}}}\left(a + \sqrt{\det(a)\det(x)}x^{-1}\right)\right).\end{aligned}$$

Proof. By (4.8)and (4.1), we have

$$\begin{aligned} a\#x^{-1} &= \frac{1}{\sqrt{2}\sqrt{\sqrt{\frac{\det(a)}{\det(x)}} + a_1 \cdot \frac{x_1}{\det(x)} - \langle a_2, \frac{-x_2}{\det(x)}\rangle}} \begin{pmatrix} \frac{a_1}{\sqrt{\det(x)}} + \frac{\sqrt{\det(a)}}{\det(x)}x_1 \\ \frac{a_2}{\sqrt{\det(x)}} - \frac{\sqrt{\det(a)}}{\det(x)}x_2 \end{pmatrix} \\ &= \frac{1}{\sqrt{2}\sqrt{\sqrt{\frac{\det(a)}{\det(x)}} + \frac{\langle a,x\rangle}{\det(x)}}} \cdot \frac{1}{\det(x)} \begin{pmatrix} \sqrt{\det(x)}a_1 + \sqrt{\det(a)}x_1 \\ \sqrt{\det(x)}a_2 - \sqrt{\det(a)}x_2 \end{pmatrix} \\ &= \frac{1}{\sqrt{2}\sqrt{\sqrt{\det(a)\det(x)} + \langle a, x\rangle}} \cdot (a + \sqrt{\det(a)\det(x)}\, x^{-1}). \end{aligned}$$

This completes the proof. □

Remark 4.7. By the third equality of Corollary 4.5, Theorem 4.6 is rephrased as

$$\nabla_x \Phi_{\frac{1}{2}}(a, x) = \frac{1}{2}\left(e - \frac{1}{\operatorname{tr}\left(P(a^{\frac{1}{2}})x\right)^{\frac{1}{2}}}\left[a + \det\left(P(a^{\frac{1}{2}})x\right)^{\frac{1}{2}} x^{-1}\right]\right).$$

Let's go back to Theorem 3.1 and compute the unique minimizer for the case of SOC in comparison to (3.1).

Theorem 4.8. *Let $a, b \in \mathcal{K}$ and $0 < s < 1$. Then the unique minimizer of* (1.4) *is*

$$\begin{aligned} W_{\frac{1}{2}}(1-s, s; a, b) &= P(a^{-1/2})\left((1-s)a + s(P(a^{1/2})b)^{1/2}\right)^2 \\ &= (1-s)^2 a + s^2 b + \frac{s(1-s)}{\kappa}(ab + \sqrt{\det(a)\det(b)}\, e). \end{aligned}$$

Proof. We first expand the term

$$\Big((1-s)a+s(P(a^{\frac12})b)^{\frac12}\Big)^2=(1-s)^2a^2+s^2P(a^{\frac12})b+2(1-s)s\,[(P(a^{\frac12})b)^{\frac12}\,a].$$

Hence, letting $\alpha=2(1-s)s$,

$$\begin{aligned}&P(a^{-1/2})\Big((1-s)a+s(P(a^{1/2})b)^{1/2}\Big)^2\\&=(1-s)^2a+s^2b+\alpha\,P(a^{-1/2})[(P(a^{1/2})b)^{1/2}\,a].\end{aligned}\tag{4.9}$$

Case 1. $a_2\neq 0$.

By (4.2) and (4.5), we have

$$(P(a^{1/2})b)^{1/2}=\Big(\kappa,\ \frac{x_2}{2\kappa}\Big)\tag{4.10}$$

where

$$\kappa=\left(\frac{\langle a,b\rangle+\sqrt{\det(a)\det(b)}}{2}\right)^{1/2}$$

and

$$x_2=\sqrt{\det(a)}b_2+\left(b_1+\Big(a_1-\sqrt{\det(a)}\Big)\frac{\langle a_2,\,b_2\rangle}{\|a_2\|^2}\right)a_2.\tag{4.11}$$

Thus

$$(P(a^{1/2})b)^{1/2}a=\begin{pmatrix}\langle a,\,(P(a^{1/2})b)^{1/2}\rangle\\ \frac{a_1}{2\kappa}x_2+\kappa a_2\end{pmatrix}.$$

Appealing to (4.1) and (4.4) yields that

$$P(a^{-1/2})=\frac{1}{\det(a)}\begin{pmatrix}a_1 & -a_2^T\\ -a_2 & \sqrt{\det(a)}I+\Big(a_1-\sqrt{\det(a)}\Big)\frac{a_2a_2^T}{\|a_2\|^2}\end{pmatrix}.\tag{4.12}$$

So

$$\begin{aligned}P(a^{-\frac12})[(P(a^{\frac12})b)^{\frac12}a]\;&=\;\frac{1}{\det(a)}\begin{pmatrix}a_1 & -a_2^T\\ -a_2 & \sqrt{\det(a)}I+\Big(a_1-\sqrt{\det(a)}\Big)\frac{a_2a_2^T}{\|a_2\|^2}\end{pmatrix}\\&\quad\times\begin{pmatrix}\langle a,(P(a^{\frac12})b)^{\frac12}\rangle\\ \frac{a_1}{2\kappa}x_2+\kappa a_2\end{pmatrix}.\end{aligned}$$

The first coordinate of the above vector is

$$\begin{aligned}\frac{1}{\det(a)}\left[a_1\left(a_1\kappa+\frac{\langle a_2,x_2\rangle}{2\kappa}\right)-\frac{a_1}{2\kappa}\langle a_2,x_2\rangle-\kappa\|a_2\|^2\right]\;&=\;\frac{\kappa}{\det(a)}(a_1^2-\|a_2\|^2)\\&=\;\kappa.\end{aligned}$$

The other is

$$\frac{1}{\det(a)}\left[-\left(a_1\kappa+\frac{\langle a_2,x_2\rangle}{2\kappa}\right)a_2+\frac{a_1\sqrt{\det(a)}}{2\kappa}\,x_2\right]$$
$$+\frac{1}{\det(a)}\left[\frac{(a_1-\sqrt{\det(a)})}{\|a_2\|^2}\cdot\frac{a_1\langle a_2,x_2\rangle}{2\kappa}\,a_2+\kappa a_1 a_2\right]$$
$$=\frac{1}{\det(a)}\left[\frac{a_1\sqrt{\det(a)}}{2\kappa}\,x_2+\frac{\langle a_2,x_2\rangle}{2\kappa}\left(\frac{a_1(a_1-\sqrt{\det(a)})}{\|a_2\|^2}-1\right)a_2\right]$$
$$=\frac{1}{\det(a)}\left[\frac{a_1\sqrt{\det(a)}}{2\kappa}\,x_2+\frac{\langle a_2,x_2\rangle}{2\kappa}\cdot\frac{\det(a)-a_1\sqrt{\det(a)}}{\|a_2\|^2}\,a_2\right]$$
$$=\frac{1}{2\kappa\sqrt{\det(a)}}\left[a_1x_2-\langle a_2,x_2\rangle\cdot\frac{a_1-\sqrt{\det(a)}}{\|a_2\|^2}\,a_2\right].$$

Therefore

$$P(a^{-\frac{1}{2}})[(P(a^{\frac{1}{2}})b)^{\frac{1}{2}}a]=\begin{pmatrix}\kappa\\ \frac{1}{2\kappa\sqrt{\det(a)}}\left[a_1x_2-\langle a_2,x_2\rangle\cdot\frac{a_1-\sqrt{\det(a)}}{\|a_2\|^2}\,a_2\right]\end{pmatrix}.\tag{4.13}$$

Moreover, by (4.11),

$$\langle a_2,x_2\rangle=b_1\|a_2\|^2+a_1\langle a_2,b_2\rangle.$$

Hence

$$\langle a_2,x_2\rangle\cdot\frac{a_1-\sqrt{\det(a)}}{\|a_2\|^2}=(a_1-\sqrt{\det(a)})\left(b_1+\frac{\langle a_2b_2\rangle}{\|a_2\|^2}a_1\right),$$

and

$$\begin{aligned}a_1x_2 \quad &- \quad \langle a_2,x_2\rangle\cdot\frac{a_1-\sqrt{\det(a)}}{\|a_2\|^2}\,a_2\\
&= \quad \left[a_1b_1+\frac{a_1(a_1-\sqrt{\det(a)})}{\|a_2\|^2}\langle a_2,x_2\rangle\right.\\
&\qquad \left.-(a_1-\sqrt{\det(a)})\left(b_1+\frac{\langle a_2,b_2\rangle}{\|a_2\|^2}a_1\right)\right]a_2+a_1\sqrt{\det(a)}b_2\\
&= \quad b_1\sqrt{\det(a)}a_2+a_1\sqrt{\det(a)}b_2\\
&= \quad \sqrt{\det(a)}\,(b_1a_2+a_1b_2).\end{aligned}$$

Substituting this into (4.13) yields that

$$\begin{aligned} P(a^{-1/2})[(P(a^{1/2})b)^{1/2}a] &= \begin{pmatrix} \kappa \\ \frac{1}{2\kappa\sqrt{\det(a)}} \cdot \sqrt{\det(a)}\,(b_1a_2 + a_1b_2) \end{pmatrix} \\ &= \frac{1}{2\kappa}\begin{pmatrix} 2\kappa^2 \\ b_1a_2 + a_1b_2 \end{pmatrix} \\ &= \frac{1}{2\kappa}\begin{pmatrix} \langle a, b\rangle + \sqrt{\det(a)\det(b)} \\ b_1a_2 + a_1b_2 \end{pmatrix} \\ &= \frac{1}{2\kappa}\left[\begin{pmatrix} \langle a, b\rangle \\ b_1a_2 + a_1b_2 \end{pmatrix} + \begin{pmatrix} \sqrt{\det(a)\det(b)} \\ 0 \end{pmatrix}\right] \\ &= \frac{1}{2\kappa}(ab + \sqrt{\det(a)\det(b)}\,e). \end{aligned}$$

Plugging this into (4.9) shows that the conclusion of the theorem holds true.

Case 2. $a_2 = 0$.

By (4.7) and (4.2), we get

$$\begin{aligned} P(a^{-1/2}) &= P(a^{1/2})^{-1} = \frac{1}{a_1}I, \\ (P(a^{1/2})b)^{1/2} &= \begin{pmatrix} \kappa \\ \frac{a_1}{2\kappa}b_2 \end{pmatrix} \end{aligned}$$

where κ is the same as (4.10). Hence as above

$$\begin{aligned} P(a^{-1/2})[(P(a^{1/2})b)^{1/2}a] &= \frac{1}{a_1}\begin{pmatrix} a_1\kappa \\ \frac{a_1^2}{2\kappa}b_2 \end{pmatrix} = \frac{1}{2\kappa}\begin{pmatrix} 2\kappa^2 \\ a_1b_2 \end{pmatrix} \\ &= \frac{1}{2\kappa}\begin{pmatrix} 2\kappa^2 \\ a_1b_2 + b_1a_2 \end{pmatrix} \\ &= \frac{1}{2\kappa}(ab + \sqrt{\det(a)\det(b)}\,e). \end{aligned}$$

Again plugging this into (4.9) finishes the proof. □

When $s = 1/2$, we get a Wasserstein-type barycenter (or mean) in $\mathcal{K}$.

Corollary 4.9.

$$W_{\frac{1}{2}}(1/2, 1/2; a, b) = \frac{1}{4}\left(a + b + \frac{1}{\kappa}(ab + \sqrt{\det(a)\det(b)}\,e)\right).$$

Remark 4.10. The above formula appears to be different from the corresponding one in the case of $\mathbb{P}_n$. In that case, the Wasserstein barycenter (or mean) of A and B is

$$W_{\frac{1}{2}}(1/2, 1/2; A, B) = \frac{1}{4}\left(A + B + (AB)^{1/2} + (BA)^{1/2}\right).$$

5. Final remark

As mentioned in the introduction, for $A,\ B \in \mathbb{P}_n$,

$$d_W(A, B) = \Phi_{\frac{1}{2}}(A, B)^{\frac{1}{2}}$$

is a metric on $\mathbb{P}_n$. In the case of second order cone $\mathcal{K}$, the symmetry, that is, $\Phi_{\frac{1}{2}}(a, b)^{\frac{1}{2}} = \Phi_{\frac{1}{2}}(b, a)^{\frac{1}{2}}$ is obvious from Corollary 4.5. However, the triangle inequality is not proved yet. That is why the $\Phi_{1/2}$-median of a and b in Corollary 4.9 is just named by *a Wasserstein-type barycenter*. So we close this section with the challenging problem:

$$\textit{Is } \Phi_{\frac{1}{2}}(a, b)^{\frac{1}{2}} \textit{ a metric on } \mathcal{K}?$$

Acknowledgments: This work was supported by Basic Science Research Program through NRF Grant No.NRF-2017R1A2B1002008.

References

[1] S. Amari, *Information geometry and its applications,* Springer (Tokyo), 2016.
[2] S. Amari, *Divergence function, information monotonicity and information geometry,* Workshop on Information Theoretic Methods in Science and Engineering (WITMSE), 2009.
[3] S. Amari and A. Cichocki, *Information geometry of divergence functions,* Bull. of the Polish Acad. of Sci. Tech. Sci., **58** (2010), 183-195.
[4] R. Bhatia, T. Jain and Y. Lim, *On the BuresWasserstein distance between positive definite matrices,* Expo. Math., **37** (2019), 165-191.
[5] D. Dowson and B. Landau, *The Fréchet distance between multivariate normal distributions,* J. Math. Anal., **12** (1982), 450-455.
[6] R. Frank and E. Lieb, *Monotonicity of a relative Rényi entropy,* J. Math. Phys. **54** (2013), 122201.
[7] J. Faraut and A. Korányi, *Analysis on symmetric cones*, Clarendon Press, Oxford, 1994.
[8] M. Knott and C. Smith, *On the optimal mapping of distributions,* J. Optim. Theory Appl., **43** (1984), 39-49.
[9] J. Kim and Y. Lim, *Jordan automorphic generators of Euclidean Jordan algebras,* J. Korean Math. Soc., **43** (2006), 507-528.
[10] S. Kum, Y. Lim and S. Yun, *Divergences on symmetric cones and medians,* submitted for publication.
[11] Y. Lim, *Geometric means on symmetric cones,* Arch. der Math., **75** (2000), 39-45.
[12] I. Olkin and F. Pukelsheim, *The distance between two random vectors with given dispersion matrices,* Linear Algebra Appl., **48** (1982), 257-263.

[13] S.H. Schmieta and F. Alizadeh,*Extension of primal-dual interior point algorithms to symmetric cones,* Math. Program. **96** (2003), 409-438.

[14] A. Uhlmann, *Density operators as an arena for differential geometry,* Rep. Math. Phys., **33** (1993), 255-263.

[15] A. Uhlmann, *Transition probability (fidelity) and its relatives,* Found. Phys., **41** (2011), 288-298.

[16] M. Wilde, A. Winter and D. Yang, *Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Renyi relative entropy,* Commu. Math. Phys., **331** (2014), 593-622