Divergences on Symmetric Cones and Medians

Sangho Kum*, Yongdo Lim and Sangwoon Yun

Abstract. We are concerned with divergences on the Cartan–Hadamard Riemannian manifold of symmetric cones, self-dual homogeneous cones in Euclidean spaces, and related optimization problems. We introduce a parameterized version of fidelity on symmetric cones, namely sandwiched quasi-relative entropies, and construct a one-parameter family of divergences based on these entropies. We consider the median minimization problem of finite points over these divergences and establish existence and uniqueness of minimizer. The global linear rate convergence of a gradient projection algorithm for solving the median minimization problem is analyzed based on the derived upper bound of the condition number of the Hessian function.

1. Introduction

Divergence functions on Riemannian manifolds (see Section 2), non-symmetric measurements of proximity, play a central role in statistical inference, machine learning, optimization, and many other fields. They can measure the dissimilarity of two points in a space. 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 function. Important examples are *Bregman divergence* derived from twice differentiable and strictly convex functions $f \colon \Omega \to \mathbb{R}$ on an open convex set Ω of a Euclidean space. A divergence Φ naturally gives rise to an important optimization problem on the Riemannian manifold M, called *median optimization problem* (see Section 2). Like the least squares averaging, its geometric meaning is evident.

In this paper we are mainly concerned with this median optimization problem on symmetric cones, self-dual homogeneous convex cones in Euclidean spaces. Moreover, a brief scheme of gradient-based optimization method with linear convergence for finding the unique minimizer is described. By the Koecher–Vinberg theorem symmetric cones correspond to the open convex cone of invertible squares in finite-dimensional real Euclidean Jordan algebras (formally real Jordan algebras). Each symmetric cone is a Cartan–Hadamard Riemannian manifold. The convex cone of $n \times n$ positive definite Hermitian

matrices and the second order cones (or Lorentz cones) are standard examples of symmetric cones.

This work is organized as follows. In Section 2, we give a description of the median optimization problem and a main theorem. In Section 3, we take a brief look at basic facts regarding Euclidean Jordan algebras and symmetric cones. In Section 4, the Löwner operator induced by a real valued function defined on the positive real numbers is introduced and some known useful properties of its derivative, gradient and Hessian are presented. According to these properties, the gradient of the sandwiched quasi-relative entropy is computed. In Section 5, a proof of the unique existence of solution of the optimization problem (2.1) based on the strict convexity of the objective function and Brouwer's fixed point theorem is provided. In Section 6, we provide an upper bound of the condition number of the Hessian of the objective function, which plays a key role in a gradient-based method for finding the unique minimizer. A complete proof of being divergence of Φ_t (see p. 869) is also given. Section 7 deals with a sketch of the gradient-based optimization method with linear convergence for finding the unique minimizer. In Section 8, an explicit formula of the unique minimizer for a special case ($m = 2$) is derived by exploiting the Riccati Lemma (3.2). In this case, the unique minimizer is nothing but the Wasserstein barycenter in the symmetric cone Ω. Section 9 is a conclusion of this work which contains a summary of results and an open problem for the future research.

2. Problem setting

A formal definition of divergence is the following (see e.g., [1–3].) A divergence on the Riemannian manifold M is a real valued function $\Phi\colon M \times M \to \mathbb{R}$ 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

$$D\Phi(a,x)\big|_{x=a} = 0;$$

(D3) its Hessian is positive definite on the diagonal

$$D^2\Phi(a,x)\big|_{x=a}(y,y) \geq 0 \quad \text{for all } a \in \Omega,\, y \in V.$$

A divergence Φ gives rise to the following optimization problem on the Riemannian manifold M:

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

where $a_1, \ldots, a_m \in M$ and $\omega = (w_1, \ldots, w_m) \in \mathbb{R}^m$ is a positive probability vector. The unique minimizer whenever it exists provides alternatively a barycenter or averaging on the manifold M, called the ω-*weighted* Φ-*median of* $a_1, \ldots, a_m$. Several medians such as Log-Euclidean and Bhattacharyya, have been derived from the divergence metrics [4, 10, 14]. Recently, Bhatia, Gaubert and Jain [6] considered this problem on the Cartan–Hadamard Riemannian manifold of positive definite matrices and obtained an explicit form of the unique minimizer for Bregman divergences of Legendre type. Divergences are not uniquely determined in a given manifold from which we have more general but new optimization problems: for divergences $\Phi_1, \Phi_2, \ldots, \Phi_m$ on the manifold M,

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

In the present work, we deal with this kind of median optimization problem on symmetric cones as mentioned in the introduction.

Let V be a Euclidean Jordan algebra and let Ω be the symmetric cone [12]. We consider the function $\Phi_t \colon \Omega \times \Omega \to \mathbb{R}$ defined by

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

where tr is the trace functional and P is the quadratic representation of the Jordan algebra V. This real valued function involves the t-weighted arithmetic mean of a and b, and the *sandwiched quasi-relative entropy*

$$F_t(a, b) := \operatorname{tr}\left(P\big(a^{\frac{1-t}{2t}}\big)b\right)^t,$$

which is well-known in the theory of quantum information; for positive semidefinite matrices A and B,

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

This is a parameterized version of the *fidelity* $F_{1/2}(A, B) = \operatorname{tr}\left(A^{1/2} B A^{1/2}\right)^{1/2}$. Fidelity and sandwiched quasi-relative entropies play an important role in quantum information theory and quantum computation, and it has deep connections with quantum entanglement, quantum chaos, and quantum phase transitions. See [13, 20–22]. In addition, the *Bures distance* in the literature on quantum information is defined by

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

which is also known as the *Wasserstein distance* in statistics and the theory of optimal transport [11, 15, 19]. In terms of our divergences,

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

Here is the main result of this paper.

Theorem 2.1. *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*

$$(2.1) \qquad\qquad \arg\min_{x \in \Omega} \sum_{j=1}^{m} w_j \Phi_{t_j}(a_j, x)$$

has a unique minimizer.

3. Euclidean Jordan algebras and symmetric cones

In this section, we briefly describe (following mostly [12]) some Jordan-algebraic concepts pertinent 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 $xy = e$.

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 $c_i^2 = c_i$, $c_i c_j = 0$, $i \neq j$, $c_1 + \cdots + c_k = e$. An idempotent is 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 3.1. (Spectral theorem, first version [12, Theorem III.1.1]) *Let V be a Euclidean Jordan algebra. Given $x \in V$, 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*

$$(3.1) \qquad\qquad x = \sum_{i=1}^{k} \lambda_i c_i.$$

The numbers λ_i are called the eigenvalues and (3.1) is called the spectral decomposition of x.

Theorem 3.2. (Spectral theorem, second version [12, 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 $\mathrm{rank}(V)$). 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 3.3. Let V be a Euclidean Jordan algebra of $\mathrm{rank}(V) = r$. The spectral mapping $\lambda \colon V \to \mathbb{R}^r$ 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 3.2 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. It turns out that Q has non-empty interior $\Omega := \operatorname{int}(Q)$, and Ω is a symmetric cone, that is, the group $G(\Omega) = \{g \in \operatorname{GL}(V) \mid g(\Omega) = \Omega\}$ acts transitively on it and Ω is a self-dual cone with respect to the inner product $\langle \cdot, \cdot \rangle$, where $\operatorname{GL}(V)$ is the Lie group of the invertible linear operators on V (see [12]). Furthermore, for any a in Ω, $P(a) \in G(\Omega)$ and is positive definite.

Note that $\overline{\Omega} = \{x \in V \mid \lambda_i(x) \geq 0, i = 1, \ldots, r\}$. For $x, y \in V$, we define

$$x \leq y \quad \text{if } y - x \in \overline{\Omega}$$

and $x < y$ if $y - x \in \Omega$. Clearly $\overline{\Omega} = \{x \in V \mid x \geq 0\}$ and $\Omega = \{x \in V \mid x > 0\}$.

On the other hand, the symmetric cone Ω in a Euclidean Jordan algebra V has an important geometric feature. That is, it admits a $G(\Omega)$-invariant Riemannian metric defined by

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

For this, refer to [12]. So Ω is a symmetric Riemannian space of non-compact type with respect to its distance metric. In this case, it is shown in [17, Proposition 2.6] that the unique geodesic curve (up to parameterization) joining a and b is

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

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 3.2. Moreover, $a \# b := a \#_{1/2} b$ is a unique geodesic middle between a and b and coincides with the unique solution in Ω of the Riccati equation

(3.2) $$P(x)a^{-1} = b.$$

Basically the trace is an inner product on V, and the Jordan algebra V endowed with the trace inner product $\langle x, y \rangle = \operatorname{tr}(xy)$ is still Euclidean [12]. Every Euclidean Jordan algebra admits a unique direct sum decompositions with irreducible (simple) Euclidean Jordan algebras. Since the trace of a product of Euclidean Jordan algebras is the sum of their trace functionals, *we may assume that V is a simple Euclidean Jordan algebra of rank r equipped with the trace inner product.*

Let $\{c_1, c_2, \ldots, c_r\}$ be a Jordan frame of V. For $i, j \in \{1, 2, \ldots, r\}$, we consider the following subspaces of V:

$$V_{ii} = V(c_i, 1) = \mathbb{R}c_i, \quad V_{ij} = V(c_i, 1/2) \cap V(c_j, 1/2) \quad \text{for } i \neq j$$

where $V(c_i, \alpha) = \{x \in V \mid L(c_i)x = \alpha x\}$. The following, called the Peirce decomposition (see [12, Theorem IV.2.1]), plays an important role for our purpose. V is the orthogonal direct sum: $V = \bigoplus_{i \leq j} V_{ij}$. Moreover,

$$V_{ij} \cdot V_{ij} \subseteq V_{ii} + V_{jj}, \quad V_{ij} \cdot V_{jk} \subseteq V_{ik} \quad \text{if } i \neq k,$$
$$V_{ij} \cdot V_{kl} = \{0\} \quad \text{if } \{i, j\} \cap \{k, l\} = \emptyset.$$

If we denote by P_{ij} the orthogonal projection onto V_{ij}, then we can write, for all $x \in V$,

$$x = \sum_{i=1}^{r} P_{ii}(x) + \sum_{i<j} P_{ij}(x) = \sum_{i=1}^{r} x_i + \sum_{i<j} x_{ij}.$$

4. Löwner operators

Let f be a differentiable map from Ω into $\mathbb{R}$. We denote by $Df(x)$ the (Fréchet) derivative of f at x, and by $\nabla f(x)$ the gradient of f at x. $Df(x)$ is a linear map from the space V into $\mathbb{R}$, and its action is given by

$$Df(x)(y) = \frac{d}{dt}\Big|_{t=0} f(x + ty).$$

$\nabla f(x)$ is an element of V and is related to $Df(x)$ by the equation

$$Df(x)(y) = \langle \nabla f(x), y \rangle = \text{tr}(\nabla f(x)y).$$

Let $\{c_1, \ldots, c_r\}$ be a Jordan frame of the simple Euclidean Jordan algebra V. For a map $\phi \colon (0, \infty) \to \mathbb{R}$, define the Löwner operator ϕ_V on Ω by

$$\phi_V \colon \Omega \to V, \quad \phi_V(x) = \sum_{j=1}^{r} \phi(\lambda_j(x)) c_j$$

where x has the spectral decomposition $x = \sum_{j=1}^{r} \lambda_j(x) c_j$. Denote

$$\phi^{\text{tr}} \colon \Omega \to \mathbb{R}, \quad \phi^{\text{tr}}(x) = \text{tr}\, \phi_V(x).$$

That is,

$$\phi^{\text{tr}}(x) = \sum_{j=1}^{r} \phi(\lambda_j(x)).$$

By applying Korányi's result [16] we have that for a continuously differentiable ϕ on an open interval $(0, \infty)$, the Löwner operator ϕ_V is differentiable at $x \in \Omega$, and for any $y \in V$,

$$(4.1) \qquad D\phi_V(x)y = \sum_{j=1}^{r} \phi'(\lambda_j(x))y_j + \sum_{i<j} \left[\phi_V^{[1]}(\lambda(x))\right]_{ij} y_{ij}$$

where $y = \sum_{i \le j} y_{ij}$ is the Peirce decomposition of y respect to the Jordan frame $\{c_1, \ldots, c_r\}$ and

$$\left[\phi_V^{[1]}(\lambda(x))\right]_{ij} = \begin{cases} \frac{\phi(\lambda_i(x)) - \phi(\lambda_j(x))}{\lambda_i(x) - \lambda_j(x)} & \text{if } \lambda_i(x) \ne \lambda_j(x), \\ \phi'(\lambda_j(x)) & \text{if } \lambda_i(x) = \lambda_j(x). \end{cases}$$

We remark that for the Euclidean Jordan algebra of Hermitian matrices, Korányi's formula is an extension of the Daleckii–Krein formula (see [5, Theorem V.3.3]).

The following derivative computations of ϕ^{tr} appears in [9, Lemma 3.1 and Theorem 3.2]. For a (continuously) differentiable ϕ on $(0, \infty)$,

(C1) ϕ^{tr} is (continuously) differentiable with

$$D\phi^{\mathrm{tr}}(x)(y) = \langle (\phi')_V(x), y \rangle.$$

That is,

$$(4.2) \qquad \nabla\phi^{\mathrm{tr}}(x) = (\phi')_V(x) = \sum_{j=1}^{r} \phi'(\lambda_j(x))c_j.$$

(C2) If ϕ is twice differentiable with $\phi''(t) > 0$ for all $t > 0$, then ϕ^{tr} is strictly convex and

$$\nabla^2\phi^{\mathrm{tr}}(x) = D(\phi')_V(x).$$

That is,

$$D^2\phi^{\mathrm{tr}}(x)(y, z) = \langle D(\phi')_V(x)(y), z \rangle, \quad x \in \Omega, \; y, z \in V.$$

Let $g \in G(\Omega)$ and let $\phi \colon (0, \infty) \to \mathbb{R}$ be a differentiable map. By chain rule,

$$(4.3) \qquad \nabla\phi^{\mathrm{tr}}(g(x)) = g^*\left(\nabla\phi^{\mathrm{tr}}(g(x))\right),$$

where g^* is the adjoint of the linear map g on V. Indeed,

$$D(\phi^{\mathrm{tr}} \circ g)(x)(y) = D\phi^{\mathrm{tr}}(g(x))(Dg(x)(y)) = \langle \nabla\phi^{\mathrm{tr}}(g(x)), Dg(x)(y) \rangle$$
$$= \langle \nabla\phi^{\mathrm{tr}}(g(x)), g(y) \rangle = \langle g^*\left(\nabla\phi^{\mathrm{tr}}(g(x))\right), y \rangle.$$

Proposition 4.1. *For every $a, x \in \Omega$ and $t \in \mathbb{R}$,*

$$\nabla\phi^{\mathrm{tr}}(P(a)x)^t = t \cdot (a^2 \#_{1-t} x^{-1}), \quad x \in \Omega.$$

Proof. Let $\phi(s) = s^t$ on $(0, \infty)$. Then by (4.2),

$$\nabla\phi^{\mathrm{tr}}(x^t) = (\phi')_V(x^t) = \sum_{j=1}^{r} t(\lambda_j(x))^{t-1}c_j = tx^{t-1}, \quad x \in \Omega.$$

By (4.3),

$$\nabla\phi^{\mathrm{tr}}(P(a)x)^t = P(a)(t(P(a)x)^{t-1}) = tP(a)(P(a^{-1})x^{-1})^{1-t}$$
$$= t(a^2 \#_{1-t} x^{-1}).$$

For the second equality, we used the fact that $(P(a)b)^{-1} = P(a^{-1})b^{-1}$ for every invertible a and b.   □

5. Medians

We first show that the objection function of (2.1) is strictly convex. It suffices to show that for each $0 < t < 1$ and $a \in \Omega$, the map $\varphi_{a,t} \colon \Omega \to \mathbb{R}$

$$\varphi_{a,t}(x) := \Phi_t(a, x) = \mathrm{tr}((1-t)a + tx) - \mathrm{tr}\left(P\left(a^{\frac{1-t}{2t}}\right)x\right)^t$$

is strictly convex. By the linearity of $x \mapsto \mathrm{tr}(x)$, it is enough to show the strict convexity of

$$x \mapsto -F_t(a, x) = -\mathrm{tr}\left(P\left(a^{\frac{1-t}{2t}}\right)x\right)^t.$$

Since the term $P(a^{\frac{1-t}{2t}})$ is independent of the variable x, it reduces to that of

$$x \mapsto -\mathrm{tr}(P(a)x)^t, \quad a \in \Omega.$$

This follows directly from chain rule and (C2) with $\phi(s) = -s^t$ for $s > 0$ and $P(a) \in G(\Omega)$. In addition, the following formula of the gradient of $\varphi_{a,t}$ is direct from Proposition 4.1

$$\nabla\varphi_{a,t}(x) = t\left(e - \left(a^{\frac{1-t}{t}} \#_{1-t} x^{-1}\right)\right).$$

Next, we shall show that the minimization problem (2.1) has a unique minimizer. Let $a_1, \ldots, a_m \in \Omega$ and let $\mathbf{t} = (t_1, \ldots, t_m) \in (0, 1)^m$. The objective function of (2.1)

$$\varphi(x) := \sum_{j=1}^m w_j \varphi_{a_j, t_j}(x)$$

is a sum of strictly convex functions and hence is strictly convex with its gradient

$$\nabla\varphi(x) = \sum_{j=1}^m w_j \nabla\varphi_{a_j, t_j}(x) = \sum_{j=1}^m w_j t_j \left[e - \left(a_j^{\frac{1-t_j}{t_j}} \#_{1-t_j} x^{-1}\right)\right].$$

Setting

(5.1) $$u := \sum_{j=1}^m w_j t_j \quad \text{and} \quad u_j = \frac{w_j t_j}{u}, \quad j = 1, \ldots, m$$

yields

$$\sum_{j=1}^{m} u_j = 1$$

and leads to

(5.2) $$\nabla\varphi(x) = 0 \quad\Longleftrightarrow\quad e = \sum_{j=1}^{m} u_j \left(a_j^{\frac{1-t_j}{t_j}} \#_{1-t_j} x^{-1} \right).$$

From

$$a_j^{\frac{1-t_j}{t_j}} \#_{1-t_j} x^{-1} = x^{-1}\#_{t_j} a_j^{\frac{1-t_j}{t_j}} = P(x^{-1/2}) \left(P(x^{1/2}) a_j^{\frac{1-t_j}{t_j}} \right)^{t_j},$$

(5.3) $$\nabla\varphi(x) = 0 \quad\Longleftrightarrow\quad x = \sum_{j=1}^{m} u_j \left(P(x^{1/2}) a_j^{\frac{1-t_j}{t_j}} \right)^{t_j}.$$

Theorem 5.1. *The optimization problem* (2.1) *has a unique solution.*

Proof. It is enough to show that the gradient vanishing equation has a solution in Ω, because of its strict convexity. Let

$$\alpha := \min\left\{\lambda_{\min}(a_j) \mid j = 1,\dots,m\right\}, \quad \beta := \max\left\{\lambda_{\max}(a_j) \mid j = 1,\dots,m\right\}$$

where $\lambda_{\min}(a_j)$ and $\lambda_{\max}(a_j)$ denote the minimum and maximum eigenvalue of a_j, respectively. We note that

$$a_j \in [\alpha e, \beta e] := \{x \mid \alpha e \le x \le \beta e\}, \quad j = 1,\dots,m.$$

Define a mapping $F\colon [\alpha e, \beta e] \to [\alpha e, \beta e]$ by

$$F(x) = \sum_{i=1}^{m} u_j \left(P(x^{1/2}) a_j^{\frac{1-t_j}{t_j}} \right)^{t_j}$$

where $u_j = \frac{w_j t_j}{\sum_{i=1}^{m} w_i t_i}$, $j = 1,\dots,m$. To see that F is a self-map, let $x \in [\alpha e, \beta e]$. Then from $\alpha e \le a_j \le \beta e$,

$$\alpha^{\frac{1-t_j}{t_j}} e \le a_j^{\frac{1-t_j}{t_j}} \le \beta^{\frac{1-t_j}{t_j}} e$$

and hence

$$\alpha^{\frac{1}{t_j}} e = \alpha^{\frac{1-t_j}{t_j}}(\alpha e) \le \alpha^{\frac{1-t_j}{t_j}} x = \alpha^{\frac{1-t_j}{t_j}} P(x^{1/2})e = P(x^{1/2})\left(\alpha^{\frac{1-t_j}{t_j}} e\right)$$
$$\le P(x^{1/2}) a_j^{\frac{1-t_j}{t_j}} \le P(x^{1/2})\left(\beta^{\frac{1-t_j}{t_j}} e\right) = \beta^{\frac{1-t_j}{t_j}} P(x^{1/2})e = \beta^{\frac{1-t_j}{t_j}} x$$
$$\le \beta^{\frac{1-t_j}{t_j}}(\beta e) = \beta^{\frac{1}{t_j}} e.$$

Therefore

$$\alpha e = \sum_{j=1}^{m} u_j \alpha e = \sum_{j=1}^{m} u_j \left(\alpha^{\frac{1}{t_j}} \right)^{t_j} e$$

$$\leq \sum_{j=1}^{m} u_j \left(P(x^{1/2}) a_j^{\frac{1-t_j}{t_j}} \right)^{t_j} (= F(x))$$

$$\leq \sum_{j=1}^{m} u_j \left(\beta^{\frac{1}{t_j}} \right)^{t_j} e = \sum_{j=1}^{m} u_j \beta e = \beta e.$$

By Brouwer's fixed point theorem, there exists a point $x \in [\alpha e, \beta e]$ such that $x = F(x)$, that is, from (5.3),

$$\nabla \varphi(x) = 0.$$

This completes the proof. $\qquad \square$

From now on, we denote by $W_{\mathbf{t}}(\omega; a_1, \ldots, a_m)$ the unique minimizer of (2.1). For the uniform cases $t := t_1 = t_2 = \cdots = t_m$ and $w_j = 1/m$, $j = 1, \ldots, m$, we simply write

$$W_t(\omega; a_1, \ldots, a_m) \quad \text{and} \quad W_{\mathbf{t}}(a_1, \ldots, a_m),$$

respectively.

We note that for every $a \in \Omega$ and $0 < t < 1$,

$$\varphi_{a,t}(a) = \text{tr}(a) - \text{tr}\left(P\left(a^{\frac{1-t}{2t}}\right)a \right)^t = \text{tr}(a) - \text{tr}\left(a^{\frac{1-t}{t}} a \right)^t = \text{tr}(a) - \text{tr}(a) = 0$$

and

(5.4) $$\nabla \varphi_{a,t}(a) = t \left(e - \left(a^{\frac{1-t}{t}} \#_{1-t} a^{-1} \right) \right) = t(e - e) = 0.$$

Therefore for $a := a_1 = a_2 = \cdots = a_m$,

$$\varphi(a) = 0, \quad \nabla \varphi(a) = 0.$$

This shows that for all $a \in \Omega$ and $\mathbf{t} = (t_1, \ldots, t_m) \in (0,1)^m$,

$$W_{\mathbf{t}}(\omega; a, \ldots, a) = a,$$

and

$$\min_{x \in \Omega} \sum_{j=1}^{m} w_j \Phi_{t_j}(a, x) = 0.$$

As a direct consequence, we get the following corollary.

Corollary 5.2. *The map $x \mapsto \Phi_t(a, x)$ has a unique minimizer at $x = a$ and its minimum*

$$\min_{x \in \Omega} \Phi_t(a, x) = 0.$$

In particular, $\Phi_t(a, b) \geq 0$ for all $a, b \in \Omega$ and $\Phi_t(a, b) = 0$ if and only if $a = b$.

This together with (5.4) shows that the map Φ_t satisfies (D1) and (D2). Moreover, by (5.2),

$$x = W_{\mathbf{t}}(\omega; a_1, \ldots, a_m) \quad \Longleftrightarrow \quad e = \sum_{j=1}^{m} u_j \left(x^{-1} \#_{t_j} a_j^{\frac{1-t_j}{t_j}} \right),$$

where u_j's are in (5.1). Then we have

$$e = \sum_{j=1}^{m} u_j \left(x^{-1} \#_{t_j} a_j \right) \quad \Longleftrightarrow \quad x = W_{\mathbf{t}} \left(\omega; a_1^{\frac{t_j}{1-t_j}}, \ldots, a_m^{\frac{t_j}{1-t_j}} \right)$$

or

$$e = \sum_{j=1}^{m} u_j \left(x \#_{t_j} a_j \right) \quad \Longleftrightarrow \quad x = W_{\mathbf{t}} \left(\omega; a_1^{\frac{t_j}{1-t_j}}, \ldots, a_m^{\frac{t_j}{1-t_j}} \right)^{-1}.$$

If $t = t_1 = t_2 = \cdots = t_m$ for all j, then $u = t$ and $u_j = w_j$. If $w_j = 1/m$, $j = 1, \ldots, m$, then

$$u = \frac{1}{m} \sum_{j=1}^{m} t_j, \quad u_j = \frac{t_j}{\sum_{i=1}^{m} t_i}.$$

Therefore

$$e = \sum_{j=1}^{m} w_j \left(x \#_t a_j \right) \quad \Longleftrightarrow \quad x = W_t \left(\omega; a_1^{\frac{t}{1-t}}, \ldots, a_m^{\frac{t}{1-t}} \right)^{-1}$$

and

$$e = \sum_{j=1}^{m} \frac{t_j}{\sum_{i=1}^{m} t_i} \left(x \#_{t_j} a_j \right) \quad \Longleftrightarrow \quad x = W_{\mathbf{t}} \left(a_1^{\frac{t_j}{1-t_j}}, \ldots, a_m^{\frac{t_j}{1-t_j}} \right)^{-1}.$$

Furthermore, we have the corollary below.

Corollary 5.3. *The following equations have unique solutions in Ω;*

$$e = \sum_{j=1}^{m} w_j (x \#_t a_j), \quad e = \sum_{j=1}^{m} \frac{t_j}{\sum_{i=1}^{m} t_i} (x \#_{t_j} a_j)$$

for every positive weight $\omega = (w_1, \ldots, w_m)$ and $0 < t_j, t < 1$, $j = 1, \ldots, m$.

Remark 5.4. The equations considered in above are special cases of the following

$$e = \sum_{j=1}^{m} w_j (x \#_{t_j} a_j).$$

It is not easy to see that this equation has a unique solution.

For the symmetric cone of positive definite matrices, Bhatia–Jain–Lim [7] proved the preceding results for $t = t_1 = t_2 = \cdots = t_m$.

6. Divergences

In this section we shall show that Φ_t is a divergence for all $0 < t < 1$. In what follows, we assume $0 < \alpha \le \beta$ and $0 < t < 1$.

Proposition 6.1. *Let* $g \colon \Omega \to V$ *be defined by* $g(x) = x^{t-1}$. *Then for* $x \in [\alpha e, \beta e]$ *and* $y \in V$,

$$(t-1)\alpha^{t-2}\|y\|^2 \le \langle Dg(x)y, y \rangle \le (t-1)\beta^{t-2}\|y\|^2.$$

Proof. Let $\{c_1, c_2, \ldots, c_r\}$ be a Jordan frame of V such that $x = \sum_{i=1}^{r} \lambda_i c_i$. According to (4.1), for any $y = \sum_{i \le j} y_{ij} \in V$ with the Peirce decomposition with respect to the Jordan frame $\{c_1, \ldots, c_r\}$

$$Dg(x)y = \sum_{i=1}^{r}(t-1)\lambda_i^{t-2}y_i + \sum_{i<j}[g^1(\lambda(x))]_{ij}y_{ij},$$

where

$$[g^1(\lambda(x))]_{ij} = \begin{cases} \frac{\lambda_i^{t-1} - \lambda_j^{t-1}}{\lambda_i - \lambda_j} & \text{if } \lambda_i \ne \lambda_j, \\ (t-1)\lambda_i^{t-2} & \text{if } \lambda_i = \lambda_j. \end{cases}$$

By the orthogonality of the Peirce decomposition and the mean value theorem,

$$\langle Dg(x)y, y \rangle = \left\langle \sum_{i=1}^{r}(t-1)\lambda_i^{t-2}y_i + \sum_{i<j}[g^1(\lambda(x))]_{ij}y_{ij}, \sum_{i=1}^{r}y_i + \sum_{i<j}y_{ij} \right\rangle$$

$$= \sum_{i=1}^{r}(t-1)\lambda_i^{t-2}\|y_i\|^2 + \sum_{i<j}[g^1(\lambda(x))]_{ij}\|y_{ij}\|^2$$

$$\ge \sum_{i=1}^{r}(t-1)\alpha^{t-2}\|y_i\|^2 + \sum_{i<j}(t-1)c^{t-2}\|y_{ij}\|^2$$

for some $\alpha \le c \le \beta$. From $t < 1$,

$$(t-1)c^{t-2} \ge (t-1)\alpha^{t-2}$$

and hence

$$\sum_{i=1}^{r}(t-1)\alpha^{t-2}\|y_i\|^2 + \sum_{i<j}(t-1)c^{t-2}\|y_{ij}\|^2 \ge (t-1)\alpha^{t-2}\sum_{i\le j}\|y_{ij}\|^2 = (t-1)\alpha^{t-2}\|y\|^2.$$

Similarly, we get

$$\langle Dg(x)y, y \rangle \le (t-1)\beta^{t-2}\|y\|^2.$$

This completes the proof. $\square$

Theorem 6.2. *Let $a \in \Omega$ and let $0 < t < 1$. Let $h \colon \Omega \to \mathbb{R}$ be the function defined by $h(x) = \operatorname{tr}(P(a)x)^t$. Then we have for $x \in [\alpha e, \beta e]$,*

$$t(t-1)\alpha^{t-2}(\lambda_{\min}(a))^{2(t-2)}(\lambda_{\max}(a))^4 I$$
$$\leq \nabla^2 h(x) \leq t(t-1)\beta^{t-2}(\lambda_{\max}(a))^{2(t-2)}(\lambda_{\min}(a))^4 I,$$

where I denotes the identity operator on V.

Proof. We note that in the notation of gradients

$$D^2 h(x)(y, z) = \langle \nabla^2 h(x)(y), z \rangle,$$

and that for $x \in [\alpha e, \beta e]$,

$$\alpha a^2 = \alpha P(a)e = P(a)(\alpha e) \leq P(a)x \leq P(a)(\beta e) = \beta a^2.$$

Hence

(6.1) $\qquad \alpha(\lambda_{\min}(a))^2 e = \alpha \lambda_{\min}(a^2)e \leq P(a)x \leq \beta \lambda_{\max}(a^2)e = \beta(\lambda_{\max}(a))^2 e.$

By Proposition 4.1,

$$\nabla h(x) = t(a^2 \#_{1-t} x^{-1}) = tP(a)\big(P(a^{-1})x^{-1}\big)^{1-t} = tP(a)(P(a)x)^{t-1}$$

from which

$$\nabla h(x) = t(P(a) \circ g \circ P(a))(x)$$

where $g(x) = x^{t-1}$. That is,

$$Dh(x)(y) = t\langle (P(a) \circ g \circ P(a))(x), y \rangle.$$

Then

$$D^2 h(x)(y, z) = t\langle \big[P(a) \circ Dg(P(a)x) \circ P(a)\big](y), z \rangle$$

and hence

$$\nabla^2 h(x)(y) = t\big[P(a) \circ Dg(P(a)x) \circ P(a)\big](y).$$

By Proposition 6.1 and (6.1), we then have that for $y \in V$,

$$\langle \nabla^2 h(x)y, y \rangle = t\langle P(a)Dg(P(a)x)P(a)y, y \rangle = t\langle Dg(P(a)x)P(a)y, P(a)y \rangle$$
$$\geq t(t-1)(\alpha(\lambda_{\min}(a))^2)^{t-2}\langle P(a)y, P(a)y \rangle$$
$$= t(t-1)\alpha^{t-2}(\lambda_{\min}(a))^{2(t-2)}\langle P(a)^2 y, y \rangle$$
$$\geq t(t-1)\alpha^{t-2}(\lambda_{\min}(a))^{2(t-2)}\lambda_{\max}(P(a^2))\|y\|^2$$
$$\geq t(t-1)\alpha^{t-2}(\lambda_{\min}(a))^{2(t-2)}(\lambda_{\max}(a))^4\|y\|^2,$$

using the fact that the eigenvalues of $P(a)$ are $\{\lambda_i(a)\lambda_j(a) \mid i, j = 1, \ldots, r\}$.

Similarly, we obtain

$$\langle \nabla^2 h(x)y, y \rangle \leq t(t-1)\beta^{t-2}(\lambda_{\max}(a))^{2(t-2)}(\lambda_{\min}(a))^4\|y\|^2. \qquad \square$$

Note that

$$\alpha e \leq a \leq \beta e \quad \Longrightarrow \quad \alpha^{\frac{1-t}{t}} e \leq a^{\frac{1-t}{t}} \leq \beta^{\frac{1-t}{t}} e.$$

Replacing a by $a^{\frac{1-t}{2t}}$ and changing the sign in Theorem 6.2 yields the following corollary.

Corollary 6.3. *Let $a \in \Omega$ and $\alpha e \leq a \leq \beta e$. Then for every $x \in [\alpha e, \beta e]$,*

$$t(1-t)\beta^{t-2}\lambda_{\max}(a)^{\frac{(1-t)(t-2)}{t}}\lambda_{\min}(a)^{\frac{2(1-t)}{t}}I$$

$$\leq \nabla^2 \varphi_{a,t}(x) \leq t(1-t)\alpha^{t-2}\lambda_{\min}(a)^{\frac{(1-t)(t-2)}{t}}\lambda_{\max}(a)^{\frac{2(1-t)}{t}}I.$$

It is direct to see that (D3) holds for Φ_t by applying the preceding corollary with sufficiently small α and large β because $D^2\Phi_t(a,x)\big|_{x=a}(y,y) = \nabla^2\varphi_{a,t}(x)\big|_{x=a}(y,y) \geq 0$. We conclude that for $0 < t < 1$, $\Phi_t \colon \Omega \times \Omega \to \mathbb{R}$ is a divergence on Ω.

The condition number of an operator A is defined as

$$\mathrm{cond}(A) = \|A\|\|A^{-1}\|.$$

We then have

$$\mathrm{cond}\left(\nabla^2\varphi_{a,t}(x)\right) \leq \left[\frac{\beta}{\alpha}\right]^{2-t}\left[\frac{\lambda_{\max}(a)}{\lambda_{\min}(a)}\right]^{\frac{(1-t)(4-t)}{t}}, \quad x \in [\alpha e, \beta e].$$

Applying with $\alpha := \lambda_{\min}(a)$ and $\beta := \lambda_{\max}(a)$ leads to

Corollary 6.4. *Let $a \in \Omega$. Then for every $x \in [\lambda_{\min}(a)e, \lambda_{\max}(a)e]$,*

$$t(1-t)\lambda_{\max}(a)^{\frac{t-2}{t}}\lambda_{\min}(a)^{\frac{2-2t}{t}}I \leq \nabla^2\varphi_{a,t}(x) \leq t(1-t)\lambda_{\min}(a)^{\frac{t-2}{t}}\lambda_{\max}(a)^{\frac{2-2t}{t}}I$$

and

$$\mathrm{cond}\left(\nabla^2\varphi_{a,t}(x)\right) \leq \left[\frac{\lambda_{\max}(a)}{\lambda_{\min}(a)}\right]^{\frac{4-3t}{t}}.$$

7. A gradient projection algorithm

In this section we present an optimization algorithm to find the unique minimizer of φ:

$$\varphi(x) := \sum_{j=1}^{m} w_j \varphi_{a_j,t_j}(x) = \sum_{j=1}^{m} w_j \left[\mathrm{tr}((1-t_j)a_j + t_j x) - \mathrm{tr}\left(P\big(a^{\frac{1-t_j}{2t_j}}\big)x\right)^{t_j}\right],$$

where $a_1, \ldots, a_m \in \Omega$ and $0 < t_j < 1$, $j = 1, \ldots, m$, are given. We have seen that

$$\nabla\varphi(x) = \sum_{j=1}^{m} w_j t_j \left[e - \big(a^{\frac{1-t_j}{t_j}} \#_{1-t_j} x^{-1}\big)\right].$$

Pick $\alpha, \beta > 0$ so that

$$a_j \in [\alpha e, \beta e], \quad j = 1, \ldots, m.$$

By Corollary 6.3,

$$t_{\min}(1 - t_{\max})\beta^{\frac{t_{\min}-2}{t_{\min}}}\alpha^{\frac{2-2t_{\min}}{t_{\min}}}I \leq \nabla^2\varphi(x) \leq t_{\max}(1 - t_{\min})\alpha^{\frac{t_{\min}-2}{t_{\min}}}\beta^{\frac{2-2t_{\min}}{t_{\min}}}I$$

for all $x \in [\alpha e, \beta e]$, where $t_{\min}$ and $t_{\max}$ denote respectively

$$t_{\min} = \min\{t_j \mid j = 1, \ldots, m\}, \quad t_{\max} = \max\{t_j \mid j = 1, \ldots, m\}.$$

This, in particular, implies that φ is

$$\alpha_* := t_{\min}(1 - t_{\max})\beta^{\frac{t_{\min}-2}{t_{\min}}}\alpha^{\frac{2-2t_{\min}}{t_{\min}}}$$

-strongly convex and

$$\beta_* := t_{\max}(1 - t_{\min})\alpha^{\frac{t_{\min}-2}{t_{\min}}}\beta^{\frac{2-2t_{\min}}{t_{\min}}}$$

-smooth on $[\alpha e, \beta e]$. Note that $\varphi(x)$ is strongly convex with parameter $\alpha_* > 0$ if and only if $\nabla^2\varphi(x) \succeq \alpha_* I$ for all $x \in \operatorname{dom}\varphi$ and $\varphi(x)$ is called β_*-smooth when the gradient of $\varphi(x)$ is Lipschitz continuous with parameter $\beta_* > 0$, i.e., $\|\nabla\varphi(x) - \nabla\varphi(y)\| \leq \beta_*\|x - y\|$ for all $x, y \in \operatorname{dom}\varphi$.

Now we apply the classical gradient projection method for solving the median minimization problem. Based on the strong convexity and smoothness of the objective function derived above, we describe the global linear rate convergence of the algorithm below. Let

$$x_{k+1} = \left[x_k - \eta\nabla\varphi(x_k)\right]_+$$
$$= \left[x_k - \left(\eta\sum_{j=1}^{m}w_jt_j\right)e + \eta\sum_{j=1}^{n}w_jt_j\left(a_j^{\frac{1-t_j}{t_j}}\#_{1-t_j}x_k^{-1}\right)\right]_+$$

where the initial iterate $x_0 \in [\alpha e, \beta e]$, $[\cdot]_+$ denotes the projection onto $[\alpha e, \beta e]$, and η is the stepsize satisfying $0 < \eta < 2/\beta_*$. Then the iterates converge to the unique minimizer $x_* := W_{\mathbf{t}}(\omega; a_1, \ldots, a_m)$ with linear convergence rate as follows:

$$\begin{aligned}
\|x_{k+1} - x_*\|_2 &= \left\|\left[x_k - \eta\nabla\varphi(x_k)\right]_+ - \left[x_* - \eta\nabla\varphi(x_*)\right]_+\right\|_2 \\
&\leq \left\|x_k - \eta\nabla\varphi(x_k) - x_* + \eta\nabla\varphi(x_*)\right\|_2 \\
&= \left\|x_k - x_* - \eta\left(\int_0^1\nabla^2\varphi(x_k^\tau)\,d\tau\right)(x_k - x_*)\right\|_2 \\
&= \left\|\left(I - \eta\int_0^1\nabla^2\varphi(x_k^\tau)\,d\tau\right)(x_k - x_*)\right\|_2 \\
&\leq \left\|I - \eta\nabla^2\varphi(x_k^\tau)\right\|_2\|x_k - x_*\|_2 \\
&\leq q^k\|x_k - x_*\|_2
\end{aligned}$$

where the first inequality uses the non-expensiveness of the projection, $x_k^\tau = x_k + \tau(x_* - x_k)$, and

$$q = \max\{|1 - \eta\alpha_*|, |1 - \eta\beta_*|\}.$$

If $\eta = 1/\beta_*$, then we can obtain

$$\|x_{k+1} - x_*\|_2^2 \le e^{-\frac{k\alpha_*}{\beta_*}} \|x_1 - x_*\|_2^2 = e^{-k\frac{t_{\min}(1-t_{\max})}{t_{\max}(1-t_{\min})}\left(\frac{\alpha}{\beta}\right)^{\frac{4-3t_{\min}}{t_{\min}}}} \|x_1 - x_*\|_2^2.$$

See [8, Theorem 3.10] for detailed proof.

8. A formula for $m = 2$

In this section we provide an explicit formula for $m = 2$ and $t_1 = t_2 = 1/2$. In this case the gradient vanishing equation of φ is

$$e = (1 - s)\left(x^{-1}\#_t a^{\frac{1-t}{t}}\right) + s\left(x^{-1}\#_t b^{\frac{1-t}{t}}\right), \quad 0 < s, t < 1.$$

Theorem 8.1. *Let $a, b \in \Omega$ and $0 < s < 1$. Then*

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

Furthermore,

(8.1) $$W_{1/2}(1 - s, s; a, b) = W_{1/2}(s, 1 - s; b, a).$$

Proof. Let $0 < s < 1$. We first consider the equation

(8.2) $$e = a\#x^{-1} + b\#x^{-1}.$$

We note that (8.2) is equivalent to

(8.3) $$e = (1 - s)(a'\#x^{-1}) + s(b'\#x^{-1})$$

where $a' := a/(1 - s)^2$ and $b' := b/s^2$. That is, for $0 < s < 1$, x is a solution of (8.2) if and only if

(8.4) $$x = W_{1/2}(1 - s, s; a', b').$$

In particular, (8.2) has a unique solution in Ω. Changing the role of a' and b' in (8.3) leads to

$$W_{1/2}(1 - s, s; a, b) = W_{1/2}(s, 1 - s; b, a)$$

for all $a, b \in \Omega$ and $0 < s < 1$.

Suppose that $x \in \Omega$ is a positive definite solution of (8.2). Setting $u = a\#x^{-1}$ and $v = b\#x^{-1}$ yields by the Riccati Lemma (3.2) that

$$x^{-1} = P(u)a^{-1} = P(v)b^{-1} \quad \text{and} \quad u + v = e.$$

That is,

$$P(u)a^{-1} = P(v)b^{-1} = P(e - u)b^{-1}.$$

Since

$$
\begin{aligned}
P(u^{-1})P(e - u) &= P(u^{-1/2})\big[P(u^{-1/2})P(e - u)P(u^{-1/2})\big]P(u^{1/2}) \\
&= P(u^{-1/2})P(P(u^{-1/2})(e - u))P(u^{1/2}) \\
&= P(u^{-1/2})P(u^{-1} - e)P(u^{1/2}) \\
&= P(u^{-1/2})P(u^{1/2})P(u^{-1} - e) \\
&= P(u^{-1} - e),
\end{aligned}
$$

we have

$$a^{-1} = P(u^{-1})P(e - u)b^{-1} = P(u^{-1} - e)b^{-1}.$$

By $u < u + v = e$ and Riccati Lemma, $u^{-1} - e = a^{-1}\#b$. That is, $u^{-1} = a^{-1}\#b + e$. From $x = P(u^{-1})a$, we have

$$x = P(a^{-1}\#b + e)a.$$

By (8.4), we have

$$P(a^{-1}\#b + e)a = W_{1/2}\big(1 - s, s; a/(1 - s)^2, b/s^2\big), \quad 0 < s < 1.$$

This implies that for $0 < s < 1$,

$$W_{1/2}(1 - s, s; a, b) = P\big(s(a^{-1}\#b) + (1 - s)e\big)a.$$

By a direct computation, the later part is equal to

$$P\big(s(a^{-1}\#b) + (1 - s)e\big)a = P(a^{-1/2})\big((1 - s)a + s(P(a^{1/2})b)^{1/2}\big)^2. \qquad \square$$

Remark 8.2. By (8.1) and the fact that $a\#b = b\#a$ and $(a\#b)^{-1} = a^{-1}\#b^{-1}$ (e.g., [17,18]),

$$P(s(a^{-1}\#b) + (1 - s)e)a = P((1 - s)(a^{-1}\#b)^{-1} + se)b$$

for all $a, b \in \Omega$ and $0 < s < 1$. This is new and is not easy to prove directly even for $s = 1/2$;

$$P(a^{-1}\#b + e)a = P(b^{-1}\#a + e)b, \quad a, b \in \Omega.$$

9. Conclusion

In this paper, we provide a general framework of a median optimization problem (2.1) on symmetric cones together with a theoretical analysis of the gradient projection method regarding this optimization problem. This problem may be regarded as a generalization of the well-known problem of finding *Wasserstein barycenter* in statistics and the theory of optimal transport when the symmetric cone Ω is the positive definite Hermitian matrices of fixed size. The Wasserstein barycenter problem is based on the *Wasserstein distance.* In terms of our divergences, the square of the Wasserstein distance is $\Phi_{1/2}(a, b)$ (see p. 869). Then we close with the following question:

Is the square root of $\Phi_{1/2}(a, b)$ a distance on the symmetric cone Ω?

This is a challenging problem for future research.

Acknowledgments

References

[1] S. Amari, *Divergence function, information monotonicity and information geometry*, Workshop on Information Theoretic Methods in Science and Engineering (WITMSE), 2007.

[2] _____, *Information Geometry and its Applications*, Applied Mathematical Sciences **194**, Springer, 2016.

[3] S. Amari and A. Cichocki, *Information geometry of divergence functions*, Bull. Pol. Acad. Sci.: Tech. Sci. **58** (2010), no. 1, 183–195.

[4] V. Arsigny, P. Fillard, X. Pennec and N. Ayache, *Log-Euclidean metrics for fast and simple calculus on diffusion tensors*, Magn. Reson. Med. **56** (2006), no. 2, 411–421.

[5] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics **169**, Springer-Verlag, New York, 1997.

[6] R. Bhatia, S. Gaubert and T. Jain, *Matrix versions of the Hellinger distance*, Lett. Math. Phys. **109** (2019), no. 8, 1777–1804.

[7] R. Bhatia, T. Jain and Y. Lim, *Strong convexity of sandwiched entropies and related optimization problems*, Rev. Math. Phys. **30** (2018), no. 9, 1850014, 18 pp.

[8] S. Bubeck, *Convex optimization: Algorithms and complexity*, Found. Trends Mach. Learn. **8** (2015), no. 3-4, 231–357.

[9] Y.-L. Chang and J.-S. Chen, *Convexity of symmetric cone trace functions in Euclidean Jordan algebras*, J. Nonlinear Convex Anal. **14** (2013), no. 1, 53–61.

[10] M. Charfi, Z. Chebbi, M. Moakher and B. Vemuri, *Bhattacharyya median of symmetric positive-definite matrices and application to the denoising of diffusion-tensor fields*, Proc. IEEE Int. Symp. Biomed. Imaging **2013**, 1227–1230.

[11] D. C. Dowson and B. V. Landau, *The Fréchet distance between multivariate normal distributions*, J. Multivariate Anal. **12** (1982), no. 3, 450–455.

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

[13] R. L. Frank and E. H. Lieb, *Monotonicity of a relative Rényi entropy*, J. Math. Phys. **54** (2013), no. 12, 122201, 5 pp.

[14] X. Hua, Y. Cheng, Y. Li, H. Wang and Y. Qin, *A divergence median-based geometric detector with a weighted averaging filter*, J. Phys.: Conf. Ser. **960** (2018), 012045, 6 pp.

[15] M. Knott and C. S. Smith, *On the optimal mapping of distributions*, J. Optim. Theory Appl. **43** (1984), no. 1, 39–49.

[16] A. Korányi, *Monotone functions on formally real Jordan algebras*, Math. Ann. **269** (1984), no. 1, 73–76.

[17] Y. Lim, *Geometric means on symmetric cones*, Arch. Math. (Basel) **75** (2000), no. 1, 39–45.

[18] _____, *Applications of geometric means on symmetric cones*, Math. Ann. **319** (2001), no. 3, 457–468.

[19] I. Olkin and F. Pukelsheim, *The distance between two random vectors with given dispersion matrices*, Linear Algebra Appl. **48** (1982), 257–263.

[20] A. Uhlmann, *Density operators as an arena for differential geometry*, Rep. Math. Phys. **33** (1993), no. 1-2, 253–263.

[21] ———, *Transition probability (fidelity) and its relatives*, Found. Phys. **41** (2011), no. 3, 288–298.

[22] M. M. Wilde, A. Winter and D. Yang, *Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy*, Comm. Math. Phys. **331** (2014), no. 2, 593–622.

Sangho Kum

Department of Mathematics Education, Chungbuk National University, Cheongju 28644, South Korea

E-mail address: `shkum@cbnu.ac.kr`

Yongdo Lim

Department of Mathematics, Sungkyunkwan University, Suwon 440-746, South Korea

E-mail address: `ylim@skku.edu`

Sangwoon Yun

Department of Mathematics Education, Sungkyunkwan University, Jongro-gu, Seoul 03722, South Korea

E-mail address: `yswmathedu@skku.edu`