J. Appl. Math. & Informatics Vol. 42(2024), No. 4, pp. 985 - 995
https://doi.org/10.14317/jami.2024.985

THE SYMMETRIZED LOG-DETERMINANT DIVERGENCE†

SEJONG KIM* AND VATSALKUMAR N. MER

Abstract. We see fundamental properties of the log-determinant α-divergence including the convexity of weighted geometric mean and the reversed sub-additivity under tensor product. We introduce a symmetrized divergence and show its properties including the boundedness and monotonicity on parameters. Finally, we discuss the barycenter minimizing the weighted sum of symmetrized divergences.

AMS Mathematics Subject Classification : 94A17, 15B48.
Key words and phrases : Log-determinant divergence, tensor product, Hadamard product, barycenter, weighted geometric mean.

1. Introduction

The notion of a divergence is a kind of statistical distance in information theory. In other words, it is a binary function which separates one probability distribution to another on a statistical manifold. A divergence over a set X is an almost distance function except that it needs not to be symmetric and not to satisfy the triangle inequality. In some literature, a divergence can be considered as a generalization of squared distance.

One of the important divergences is the Kullback-Leibler divergence [7]: for two positive definite (Hermitian) matrices A and B, which represents covariance matrices of two zero-mean Gaussian distributions,

$$D_{\mathrm{KL}}(A, B) = \mathrm{tr}(AB^{-1} - I) - \log\det(AB^{-1}).$$

See the reference [4] for the derivation of Kullback-Leibler divergence between two Gaussian distributions. Throughout the paper, log is the natural logarithmic map. Various versions of the Kullback-Leibler divergence are used in several

Received February 23, 2024. Revised July 1, 2024. Accepted July 7, 2024. *Corresponding author.
†This research was supported by Chungbuk National University Korea National University Development Project (2023).

research areas, including applied statistics, machine learning, neuroscience, and signal processing [10, 11, 12, 14]. There are numerous types of divergences and classes of divergences, for instance, f-divergence and Bregman divergence.

We study in this paper the *log-determinant α-divergence*, introduced by Chebbi and Moakher [3]: for any $\alpha \in (-1, 1)$

$$D_\alpha(A|B) := \frac{4}{1-\alpha^2} \log \frac{\det\left(\frac{1-\alpha}{2}A + \frac{1+\alpha}{2}B\right)}{(\det A)^{(1-\alpha)/2}(\det B)^{(1+\alpha)/2}}.$$

This is a one-parameter family of Kullback-Leibler divergences which is related with the Stein's loss. We prove its fundamental properties including the convexity of weighted geometric mean and see effects of tensor product on the log-determinant α-divergence in Section 2.

A symmetrized Kullback-Leibler divergence with parameter $\mu (\geq 0)$ has been introduced [6]:

$$D_s^\mu(A, B) = \frac{1}{2}[D_{\mathrm{KL}}^\mu(A, B) + D_{\mathrm{KL}}^\mu(B, A)],$$

where $D_{\mathrm{KL}}^\mu(A, B) = \mathrm{tr}((A - B)(B + \mu I)^{-1}) - \log\det(A + \mu I) + \log\det(B + \mu I)$. Furthermore, the authors have defined a new multivariable mean by solving the optimization problem aimed at minimizing a weighted sum of the symmetrized Kullback-Leibler divergences. In Section 3, we also introduce the symmetrized log-determinant α-divergence and prove its properties including the monotonicity on parameters. Finally in Section 4, we discuss some open question on the (symmetrized) log-determinant α-divergence and its minimization problem.

2. Log-determinant divergence

Let $A, B \in \mathbb{P}_m$, the open convex cone of all $m \times m$ positive definite (Hermitian) matrices. The *log-determinant α-divergence* $D_\alpha(A|B)$ for $\alpha \in (-1, 1)$ is defined by

$$D_\alpha(A|B) := \frac{4}{1-\alpha^2} \log \frac{\det\left(\frac{1-\alpha}{2}A + \frac{1+\alpha}{2}B\right)}{(\det A)^{(1-\alpha)/2}(\det B)^{(1+\alpha)/2}}. \tag{1}$$

We can rewrite it as

$$D_\alpha(A|B) = \frac{4}{1-\alpha^2}\mathrm{tr}\left[\log\left(\frac{1-\alpha}{2}A + \frac{1+\alpha}{2}B\right) - \frac{1-\alpha}{2}\log A - \frac{1+\alpha}{2}\log B\right]. \tag{2}$$

One can see that

$$D_{-1}(A|B) := \lim_{\alpha \to -1} D_\alpha(A|B) = \mathrm{tr}(A^{-1}B - I) - \log\det(A^{-1}B),$$
$$D_1(A|B) := \lim_{\alpha \to 1} D_\alpha(A|B) = \mathrm{tr}(B^{-1}A - I) - \log\det(B^{-1}A).$$

We have

$$D_{-1}(A|B) = \|X - \log X - I\|_1 \quad \text{and} \quad D_1(A|B) = \|X^{-1} + \log X - I\|_1,$$

where $X = A^{-1/2}BA^{-1/2}$ and $\|Z\|_1 := \mathrm{tr}((Z^*Z)^{1/2})$ denotes the Schatten 1-norm of $Z \in M_m$, the set of all $m \times m$ complex matrices. Indeed, since $X - \log X - I$ is positive semi-definite,

$$\|X - \log X - I\|_1 = \mathrm{tr}(X - \log X - I).$$

Similarly, $\|X^{-1} + \log X - I\|_1 = \mathrm{tr}(X^{-1} + \log X - I)$.

We provide fundamental properties of the log-determinant α-divergence D_α. We denote as GL_m the general linear group of all $m \times m$ invertible matrices.

Proposition 2.1. *Let $A, B \in \mathbb{P}_m$ and $\alpha \in [-1, 1]$. Then*

(i) $D_\alpha(PAQ|PBQ) = D_\alpha(A|B)$ *for any* $P, Q \in \mathrm{GL}_m$,
(ii) $D_\alpha(A^{-1}|B^{-1}) = D_\alpha(B|A)$, *and*
(iii) $D_\alpha(A^t|B^t) \leq t\, D_\alpha(A|B)$ *for any* $t \in [0, 1]$.
(iv) $D_\alpha(A^s|A^t) \leq (t-s)D_\alpha(I|A)$ *for* $0 \leq s \leq t \leq 1$, *and* $D_\alpha(A^s|A^t) \leq (s-t)D_\alpha(A|I)$ *for* $0 \leq t \leq s \leq 1$.

Proof. All properties (i)-(iii) have been proved in [9, Lemma 5.3]. It remains to show (iv). We first assume $0 \leq s \leq t \leq 1$. Since

$$\frac{\det\left(\frac{1-\alpha}{2}A^s + \frac{1+\alpha}{2}A^t\right)}{(\det A^s)^{(1-\alpha)/2}(\det A^t)^{(1+\alpha)/2}} = \frac{\det\left(\frac{1-\alpha}{2}I + \frac{1+\alpha}{2}A^{t-s}\right)}{(\det A^{t-s})^{(1+\alpha)/2}},$$

we have $D_\alpha(A^s|A^t) = D_\alpha(I|A^{t-s}) \leq (t-s)D_\alpha(I|A)$ from (iii). The second assertion follows similarly. □

Remark 2.1. Note from [13] that

$$d_S(A, B) := \frac{1}{2}\sqrt{D_0(A|B)} = \sqrt{\log\det\left(\frac{A+B}{2}\right) - \frac{1}{2}\log\det(AB)}$$

is a distance on $\mathbb{P}_m$. Since $\log\det A = \mathrm{tr}\log A$ for any $A \in \mathbb{P}_m$, we have an alternative expression of $d_S(A, B)$ such as

$$d_S(A, B) = \sqrt{\mathrm{tr}\left[\log\left(\frac{A+B}{2}\right) - \frac{\log A + \log B}{2}\right]}.$$

By Proposition 2.1 we obtain the following properties for d_S;

(1) $d_S(MAM^*, MBM^*) = d_S(A, B)$ for any $M \in \mathrm{GL}_m$,
(2) $d_S(A^{-1}, B^{-1}) = d_S(A, B)$,
(3) $d_S(A^t, B^t) \leq \sqrt{t}d_S(A, B)$ for any $t \in [0, 1]$,
(4) $d_S(A^s, A^t) \leq \sqrt{|s-t|}d_S(A, I)$ for any $s, t \in [0, 1]$.

The weighted geometric mean of $A, B \in \mathbb{P}_m$ is given by

$$A\#_t B = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2},\ t \in [0, 1],$$

which is the unique geodesic for Riemannian trace metric: see [2, Chapter 6] for more information. We see certain convexity of the weighted geometric mean for d_S.

Theorem 2.2. *Let $A, B, C, D \in \mathbb{P}_m$ and $s, t \in [0, 1]$. Then*

$$d_S(A\#_s B, C\#_t D) \leq \sqrt{1-s}\, d_S(A, C) + \sqrt{s}\, d_S(B, D) + \sqrt{|s-t|}\, d_S(C, D).$$

Proof. By (1) and (3) in Remark 2.1, we have the following inequality of the weighted geometric mean for d_S:

$$d_S(A\#_t B, A\#_t C) \leq \sqrt{t}\, d_S(B, C)$$

for $A, B, C \in \mathbb{P}_m$ and $t \in [0, 1]$. Thus, by the triangle inequality for d_S and $A\#_t B = B\#_{1-t} A$

$$\begin{aligned} d_S(A\#_s B, C\#_t D) &\leq d_S(A\#_s B, A\#_s D) + d_S(A\#_s D, C\#_s D) + d_S(C\#_s D, C\#_t D) \\ &\leq \sqrt{s}\, d_S(B, D) + \sqrt{1-s}\, d_S(A, C) + \sqrt{|s-t|}\, d_S(C, D). \end{aligned}$$

Indeed, by Remark 2.1 (1), (4)

$$\begin{aligned} d_S(C\#_s D, C\#_t D) &= d_S((C^{-1/2} D C^{-1/2})^s, (C^{-1/2} D C^{-1/2})^t) \\ &\leq \sqrt{|s-t|}\, d_S(I, C^{-1/2} D C^{-1/2}) = \sqrt{|s-t|}\, d_S(C, D). \end{aligned}$$

□

Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be arbitrary matrices with certain sizes. The *tensor product* (or *Kronecker product*) $A \otimes B$ of A and B is the matrix given by

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1nB} \\ a_{21}B & a_{22}B & \cdots & a_{2nB} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mnB} \end{bmatrix}.$$

One can see easily that the tensor product is bilinear and associative, but not commutative. Moreover, it preserves the positivity: the tensor product of two positive semi-definite (positive definite) matrices is positive semi-definite (positive definite, respectively).

The following provides a useful formula of the logarithmic map $\log : \mathbb{P}_m \to \mathbb{H}_m$ with tensor product, where $\mathbb{H}_m$ denotes the real vector space of all $m \times m$ Hermitian matrices.

Lemma 2.3. *Let $A, B \in \mathbb{P}_m$. Then* $\log(A \otimes B) = (\log A) \otimes I + I \otimes (\log B)$.

Proof. Note that $(A \otimes B)^t = A^t \otimes B^t$ for any $t \in \mathbb{R}$. Taking derivative on both sides yield

$$(A \otimes B)^t \log(A \otimes B) = (A^t \log A) \otimes B^t + A^t \otimes (B^t \log B).$$

Putting $t = 0$ we obtain the desired property. □

We see the effect of tensor product on the log-determinant α-divergence. In the following, the relation $\leq$ means the Loewner partial order on $\mathbb{H}_m$.

Theorem 2.4. *Let $A, B, C, D \in \mathbb{P}_m$ and $\alpha \in [-1, 1]$. Then*

(i) *if either* $A \geq B, C \geq D$ *or* $A \leq B, C \leq D$ *then*

$$D_\alpha(A \otimes C|B \otimes D) \geq m\left[D_\alpha(A|B) + D_\alpha(C|D)\right],$$

(ii) *if either* $A \geq B, C \leq D$ *or* $A \leq B, C \geq D$ *then*

$$D_\alpha(A \otimes C|B \otimes D) \leq m\left[D_\alpha(A|B) + D_\alpha(C|D)\right].$$

Proof. We first assume either $A \geq B, C \geq D$ or $A \leq B, C \leq D$. Then

$$\left(\frac{1-\alpha}{2}A + \frac{1+\alpha}{2}B\right) \otimes \left(\frac{1-\alpha}{2}C + \frac{1+\alpha}{2}D\right) \leq \frac{1-\alpha}{2}A \otimes C + \frac{1+\alpha}{2}B \otimes D. \quad (3)$$

Indeed, by the linearity of tensor product

$$\frac{1-\alpha}{2}A \otimes C + \frac{1+\alpha}{2}B \otimes D - \left(\frac{1-\alpha}{2}A + \frac{1+\alpha}{2}B\right) \otimes \left(\frac{1-\alpha}{2}C + \frac{1+\alpha}{2}D\right)$$
$$= \frac{1-\alpha^2}{4}(A-B) \otimes (C-D).$$

So the assumption implies that the right-hand side is positive semi-definite.

Since the logarithmic map is monotone increasing and the trace map is linear, we obtain from (2) that

$$\begin{aligned} &D_\alpha(A \otimes C|B \otimes D) \\ &\geq \frac{4}{1-\alpha^2}\operatorname{tr}\log\left[\left(\frac{1-\alpha}{2}A + \frac{1+\alpha}{2}B\right) \otimes \left(\frac{1-\alpha}{2}C + \frac{1+\alpha}{2}D\right)\right] \\ &\quad - \frac{1+\alpha}{2}\operatorname{tr}\log(A \otimes C) - \frac{1-\alpha}{2}\operatorname{tr}\log(B \otimes D) \\ &= \frac{4m}{1-\alpha^2}\left[\operatorname{tr}\log\left(\frac{1-\alpha}{2}A + \frac{1+\alpha}{2}B\right) + \operatorname{tr}\log\left(\frac{1-\alpha}{2}C + \frac{1+\alpha}{2}D\right)\right] \\ &\quad - \frac{(1+\alpha)m}{2}\left[\operatorname{tr}\log A + \operatorname{tr}\log C\right] - \frac{(1-\alpha)m}{2}\left[\operatorname{tr}\log B + \operatorname{tr}\log D\right] \\ &= m\left[D_\alpha(A|B) + D_\alpha(C|D)\right] \end{aligned}$$

for $\alpha \in (-1,1)$. The second equality follows from Lemma 2.3 and the fact that $\operatorname{tr}(A \otimes B) = (\operatorname{tr}A)(\operatorname{tr}B)$. Taking limit as $\alpha \to \pm 1$ on the above inequality, we obtain (i) for $\alpha \in [-1,1]$.

With the assumption that either $A \geq B, C \leq D$ or $A \leq B, C \geq D$, the inequality (3) is reversed. So we can obtain (ii) by the similar process as above. □

Remark 2.2. Since $D_\alpha(A|B) \geq 0$ for any $A, B \in \mathbb{P}_m$ by [3, Proposition 3.5], Theorem 2.4 (i) implies

$$D_\alpha(A \otimes C|B \otimes D) \geq D_\alpha(A|B) + D_\alpha(C|D)$$

when either $A \geq B, C \geq D$ or $A \leq B, C \leq D$. This can be considered as a reversed sub-additivity of the log-determinant α-divergence under tensor product. On the other hand, it is an interesting question what happens between $D_\alpha(A \otimes C|B \otimes D)$ and $D_\alpha(A|B), D_\alpha(C|D)$ for general $A, B, C, D \in \mathbb{P}_m$.

3. Symmetrized log-determinant divergence

We naturally define a *symmetrized log-determinant α-divergence* by

$$S_\alpha(A,B) := \frac{1}{2}\left[D_\alpha(A|B) + D_\alpha(B|A)\right]. \tag{4}$$

It is obvious from Remark 2.1 that $S_0(A,B) = D_0(A|B)$ and $d_S(A,B) = \frac{1}{2}\sqrt{S_0(A,B)}$.

Proposition 3.1. *For any $\alpha \in (-1,1)$ and $A, B \in \mathbb{P}_m$,*

$$S_\alpha(A,B) = \frac{2}{1-\alpha^2}\operatorname{tr}\log\left[\frac{1+\alpha^2}{2}I + \frac{1-\alpha^2}{2}\left(\frac{X+X^{-1}}{2}\right)\right],$$

where $X = A^{-1/2}BA^{-1/2}$.

Proof. Note that

$$D_\alpha(A|B) = \frac{4}{1-\alpha^2}\operatorname{tr}\left[\log\left(\frac{1-\alpha}{2}I + \frac{1+\alpha}{2}X\right) - \frac{1+\alpha}{2}\log X\right]$$

and

$$D_\alpha(B|A) = \frac{4}{1-\alpha^2}\operatorname{tr}\left[\log\left(\frac{1-\alpha}{2}X + \frac{1+\alpha}{2}I\right) - \frac{1-\alpha}{2}\log X\right],$$

where $X = A^{-1/2}BA^{-1/2}$. So by direct calculation

$$\begin{aligned}
&S_\alpha(A,B)\\
&= \frac{2}{1-\alpha^2}\operatorname{tr}\left[\log\left(\frac{1-\alpha}{2}I + \frac{1+\alpha}{2}X\right) + \log\left(\frac{1-\alpha}{2}X + \frac{1+\alpha}{2}I\right) - \log X\right]\\
&= \frac{2}{1-\alpha^2}\operatorname{tr}\log\left[\left(\frac{1-\alpha}{2}I + \frac{1+\alpha}{2}X\right)X^{-1}\left(\frac{1-\alpha}{2}X + \frac{1+\alpha}{2}I\right)\right]\\
&= \frac{2}{1-\alpha^2}\operatorname{tr}\log\left[\left(\frac{1-\alpha}{2}\right)^2 I + \left(\frac{1-\alpha^2}{2}\right)\left(\frac{X+X^{-1}}{2}\right) + \left(\frac{1+\alpha}{2}\right)^2 I\right]\\
&= \frac{2}{1-\alpha^2}\operatorname{tr}\log\left[\frac{1+\alpha^2}{2}I + \frac{1-\alpha^2}{2}\left(\frac{X+X^{-1}}{2}\right)\right].
\end{aligned}$$

□

Remark 3.1. Alternatively, we have

$$S_\alpha(A,B) = \frac{2}{1-\alpha^2}\operatorname{tr}\log\left[\frac{1+\alpha^2}{2}I + \frac{1-\alpha^2}{2}\left(\frac{A^{-1}B+B^{-1}A}{2}\right)\right]. \tag{5}$$

Indeed,

$$\begin{aligned}
&\operatorname{tr}\log\left[\frac{1+\alpha^2}{2}I+\frac{1-\alpha^2}{2}\left(\frac{X+X^{-1}}{2}\right)\right]\\
&=\operatorname{tr}\log\left[\frac{1+\alpha^2}{2}I+\frac{1-\alpha^2}{2}A^{1/2}\left(\frac{A^{-1}B+B^{-1}A}{2}\right)A^{-1/2}\right]\\
&=\operatorname{tr}\left[A^{1/2}\log\left[\frac{1+\alpha^2}{2}I+\frac{1-\alpha^2}{2}\left(\frac{A^{-1}B+B^{-1}A}{2}\right)\right]A^{-1/2}\right]\\
&=\operatorname{tr}\log\left[\frac{1+\alpha^2}{2}I+\frac{1-\alpha^2}{2}\left(\frac{A^{-1}B+B^{-1}A}{2}\right)\right].
\end{aligned}$$

Remark 3.2. By equation (5), we have $S_\alpha(A,B) = S_{-\alpha}(A,B)$ for any $\alpha \in (-1,1)$. So the symmetrized log-determinant α-divergence S_α can be defined only for $\alpha \in [0,1)$.

The following provides the lower and upper bounds for the symmetrized log-determinant α-divergence. The (weighted) arithmetic-geometric-harmonic mean inequalities are useful: for $A,B \in \mathbb{P}_m$ and $t \in [0,1]$

$$[(1-t)A^{-1}+tB^{-1}]^{-1} \le A\#_t B \le (1-t)A+tB. \tag{6}$$

We denote as $\lambda_i(X)$ eigenvalues of $X \in \mathbb{H}_m$ in decreasing order: $\lambda_1(X) \ge \lambda_2(X) \ge \cdots \ge \lambda_m(X)$.

Lemma 3.2. *Let $A,B \in \mathbb{P}_m$ and $\alpha \in [0,1)$. Then*

$$\operatorname{tr}\log\left(\frac{A^{-1}B+B^{-1}A}{2}\right) \le S_\alpha(A,B) \le \frac{2}{1-\alpha^2}\operatorname{tr}\log\left(\frac{A^{-1}B+B^{-1}A}{2}\right). \tag{7}$$

Moreover,

$$\frac{2}{1-\alpha^2}\operatorname{tr}\log\left(\frac{A^{-1}B+B^{-1}A}{2}\right) \le \frac{2m}{1-\alpha^2}\log\left(\frac{R+r}{2\sqrt{Rr}}\right),$$

where $R = \lambda_1(A^{-1}B)$ and $r = \lambda_m(A^{-1}B)$.

Proof. Let $X = A^{-1/2}BA^{-1/2} \in \mathbb{P}_m$. Since the logarithmic map $\log : \mathbb{P}_m \to \mathbb{H}_m$ is operator concave and $\log I = O$, which is a zero matrix,

$$\log\left[\frac{1+\alpha^2}{2}I+\frac{1-\alpha^2}{2}\left(\frac{X+X^{-1}}{2}\right)\right] \ge \frac{1-\alpha^2}{2}\log\left(\frac{X+X^{-1}}{2}\right).$$

Taking the trace on both sides and applying Proposition 3.1 yield the first inequality of (7).

Since

$$\frac{X+X^{-1}}{2} \ge X\#X^{-1} = I$$

by (6) and the logarithmic map $\log : \mathbb{P}_m \to \mathbb{H}_m$ is operator monotone,

$$\log\left[\frac{1+\alpha^2}{2}I+\frac{1-\alpha^2}{2}\left(\frac{X+X^{-1}}{2}\right)\right] \le \log\left(\frac{X+X^{-1}}{2}\right).$$

So we obtain the second inequality of (7).

Since the map $\Phi(X) = \dfrac{X+X^{-1}}{2}$ is strictly positive and unital, we have the following by applying [2, Proposition 2.7.8] with $rI \leq X = A^{-1/2}BA^{-1/2} \leq RI$ and (6)

$$\frac{X+X^{-1}}{2} \leq \frac{(R+r)^2}{4Rr}\left(\frac{X+X^{-1}}{2}\right)^{-1} \leq \frac{(R+r)^2}{4Rr}X\#X^{-1} = \frac{(R+r)^2}{4Rr}I.$$

Since $\dfrac{(R+r)^2}{4Rr} \geq 1$,

$$\frac{1+\alpha^2}{2}I + \frac{1-\alpha^2}{2}\left(\frac{X+X^{-1}}{2}\right) \leq \frac{(R+r)^2}{4Rr}I.$$

Since the logarithmic map and trace map are monotone increasing, we obtain the last assertion. □

Corollary 3.3. *For any $\alpha \in [0,1)$, $S_\alpha : \mathbb{P}_m \times \mathbb{P}_m \to \mathbb{R}$ is a symmetric divergence. That is, for $A, B \in \mathbb{P}_m$*

$$S_\alpha(A,B) \geq 0,$$

and the equality holds if and only if $A = B$.

Proof. Note from (6) that $\dfrac{X+X^{-1}}{2} \geq X\#X^{-1} = I$ for any $X \in \mathbb{P}_m$. Applying the first inequality of (7) in Lemma 3.2 we obtain

$$S_\alpha(A,B) \geq \frac{2}{1-\alpha^2}\operatorname{tr}\log\left(\frac{X+X^{-1}}{2}\right) \geq \operatorname{tr}\log I = 0.$$

If $A = B$, then $X = I$, so it is easy to see $S_\alpha(A,B) = 0$. Conversely, assume $S_\alpha(A,B) = 0$. Then $\dfrac{X+X^{-1}}{2} = I$ from the above. It is satisfied only when $X = I$, that is, $A = B$. □

Remark 3.3. The symmetrized log-determinant α-divergence S_α is a semi-metric on $\mathbb{P}_m$. In other words, it satisfies all axioms of metric but not necessarily the triangle inequality. The following is such an example: let $\alpha = \frac{1}{2}$,

$$A = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 3 \end{bmatrix}, \ B = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } C = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}.$$

Then $S_\alpha(A,B) \approx 0.67888 > 0.49853 \approx S_\alpha(A,C) + S_\alpha(C,B)$. On the other hand, it is an interesting question to find the condition of A, B, C to fulfill the triangle inequality or the subset of $\mathbb{P}_m$ such that S_α is a metric.

Theorem 3.4. *Let $A, B \in \mathbb{P}_m$. For any α, β such that $0 \leq \alpha \leq \beta < 1$,*

$$(1-\alpha^2)S_\alpha(A,B) \geq (1-\beta^2)S_\beta(A,B).$$

Proof. Let $Y, Z \in \mathbb{P}_m$ such that $Y \leq Z$. For $0 < p \leq q \leq 1$ and any monotone increasing function f,

$$f(Y) \leq f((1-p)Y + pZ) \leq f((1-q)Y + qZ) \leq f(Z).$$

Let us replace Y, Z, p, and q by I, $\frac{X + X^{-1}}{2}$, $\frac{1-\beta^2}{2}$, and $\frac{1-\alpha^2}{2}$. Then $0 < p \leq q \leq \frac{1}{2}$, and moreover, from (6)

$$Z = \frac{X + X^{-1}}{2} \geq X \# X^{-1} = I = Y.$$

So taking a logarithmic map which is monotone increasing yields

$$\log\left[\frac{1+\alpha^2}{2} I + \frac{1-\alpha^2}{2}\left(\frac{Y+Y^{-1}}{2}\right)\right] \geq \log\left[\frac{1+\beta^2}{2} I + \frac{1-\beta^2}{2}\left(\frac{Y+Y^{-1}}{2}\right)\right] \geq O.$$

Since $A \geq B \geq O$ implies $\mathrm{tr}A \geq \mathrm{tr}B \geq 0$, we obtain the desired inequality. □

Remark 3.4. Theorem 3.4 with Remark 2.1 implies

$$2d_S(A, B) \geq \sqrt{(1-\alpha^2)S_\alpha(A, B)}$$

for $\alpha \in (0, 1)$.

4. Discussion on divergence and barycenter

The *Hadamard* (or *Schur*) *product* $A \circ B$ of $A = [a_{ij}]$ and $B = [b_{ij}]$ in $M_{m,k}$ is the $m \times k$ matrix, which is defined by the entrywise product:

$$A \circ B := [a_{ij} b_{ij}].$$

Note that Hadamard product is bilinear, commutative, and associative. Moreover, the Hadamard product also preserves positivity; the Hadamard product of two positive definite (positive semidefinite, respectively) matrices is positive definite (positive semidefinite, respectively) matrices. This is known as the Schur product theorem [5, 15]. There is a canonical relationship between the tensor product and Hadamard product via a positive unital linear map.

Lemma 4.1. [1, Lemma 4] *There exists a strictly positive and unital linear map* Ψ *such that for any* $A, B \in M_m$

$$\Psi(A \otimes B) = A \circ B.$$

Theorem 2.4 tells us how tensor product effects to the log-determinant α-divergence D_α. We can naturally ask the relationship between $D_\alpha(A \circ C | B \circ D)$ and $D_\alpha(A|B), D_\alpha(C|D)$ for $A, B, C, D \in \mathbb{P}_m$.

Note that

$$S(A, B) := \lim_{\alpha \to 1} S_\alpha(A, B) = \mathrm{tr}\left(\frac{X + X^{-1}}{2} - I\right) = \mathrm{tr}\left(\frac{A^{-1}B + B^{-1}A}{2} - I\right),$$

which is known as the *symmetrized Kullback-Leibler divergence*. Let $A_1, \dots, A_n \in \mathbb{P}_m$ and $(w_1, \dots, w_n)$ be a positive probability vector. By [6, Theorem 5.3]

$$\underset{X\in\mathbb{P}_m}{\arg\min}\sum_{i=1}^{n} w_i S(X, A_i) = \left(\sum_{i=1}^{n} w_i A_i\right) \# \left(\sum_{i=1}^{n} w_i A_i^{-1}\right)^{-1}, \tag{8}$$

where the right-hand side of (8) is called the weighted $\mathcal{A}\#\mathcal{H}$-mean (See [8]). Here, $\mathcal{A}$ and $\mathcal{H}$ denote the weighted arithmetic and harmonic means, respectively, and

$$A\#B = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$$

is the unique midpoint of $A, B \in \mathbb{P}_m$ for the Riemannian trace metric, as well as the unique solution $X \in \mathbb{P}_m$ of Riccati equation $XA^{-1}X = B$. In particular, for $n = 2$ and $\omega = (1/2, 1/2)$

$$A\#B = \underset{X\in\mathbb{P}_m}{\arg\min}\, S(X, A) + S(X, B)$$

since the following holds from the Riccati equation:

$$\left(\frac{A+B}{2}\right) \# \left(\frac{A^{-1}+B^{-1}}{2}\right)^{-1} = A\#B.$$

In this point of view, it is an interesting question whether the following minimization problem

$$\min \sum_{i=1}^{n} w_i S_\alpha(X, A_i)$$

for $\alpha \in (0, 1)$ has a unique minimizer in $\mathbb{P}_m$.

Conflicts of interest : The authors declare no conflict of interest.

Data availability : Not applicable

Acknowledgments : We thank to anonymous reviewers for valuable comments.

References

1. T. Ando, *Concavity of certain maps on positive definite matrices and applications to Hadamard products*, Linear Algebra Appl. **26** (1979), 203-241.
2. R. Bhatia, *Positive Definite Matrices*, Princeton Series in Applied Mathematics, 2007.
3. Z. Chebbi and M. Moakher, *Means of hermitian positive-definite matrices based on the log-determinant α-divergence function*, Linear Algebra Appl. **436** (2012), 1872-1889.
4. John Duchi, *Derivations for Linear Algebra and Optimization*, Berkeley, California 3.1 (2007): 2325-5870.
5. R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, 2009.
6. S. Kim, J. Lawson and Y. Lim, *The matrix geometric mean of parameterized, weighted arithmetic and harmonic means*, Linear Algebra Appl. **435** (2011), 2114-2131.
7. S. Kullback and R. A. Leibler, *On information and sufficiency*, Ann. Math. Statistics **22** (1951), 79-86.

8. S. Kum and Y. Lim, *A geometric mean of parameterized arithmetic and harmonic means of convex functions*, Abstr. Appl. Anal. **15** (2012), Art. ID 836804.
9. V.N. Mer and S. Kim, *New multivariable mean from nonlinear matrix equation associated to the harmonic mean*, Acta Sci. Math. (Szeged) (2024). https://doi.org/10.1007/s44146-024-00132-y
10. Ignacio Montes, *Neighbourhood models induced by the euclidean distance and the Kullback-Leibler divergence*, Proc. Mach. Learn. Res. **215** (2023), 367-378.
11. Dunbiao Niu, Enbin Song, Zhi Li, Linxia Zhang, Ting Ma, Juping Gu and Qingjiang Shi, *A marginal distributionally robust MMSE estimation for a multisensor system with Kullback-Leibler divergence constraints*, IEEE Trans. Signal Process **71** (2023), 3772-3787.
12. F.J. Pinski, G. Simpson, A.M. Stuart and H. Weber, *Kullback-Leibler approximation for probability measures on infinite dimensional spaces*, SIAM J. Math. Anal. **47** (2015), 4091-4122.
13. S. Sra, *A new metric on the manifold of kernel matrices with application to matrix geometric means*, NIPS (2012), 144-152.
14. J. Watson, L. Nieto-Barajas and C. Holmes, *Characterizing variation of nonparametric random probability measures using the Kullback-Leibler divergence*, Statistics **51** (2017), 558-571.
15. F. Zhang, *Matrix Theory: Basic Results and Techniques*, 2nd edition, Springer, 2011.

Sejong Kim received Master of Science from Kyungpook National University, and Ph.D. from Louisiana State University, USA. He is currently a professor at Chungbuk National University since 2013. His research interests are operator mean and quantum information.

Department of Mathematics, Chungbuk National University, Cheongju 28644, Korea.
e-mail: skim@chungbuk.ac.kr

Vatsalkumar N. Mer received Master of Science from Sardar Patel University and Ph.D. from IISER Thiruvananthapuram, India. He is currently a postdoc at Institute for Industrial and Applied Mathematics, Chungbuk National University. His research interests include matrix analysis and preserver problem.

Institute for Industrial and Applied Mathematics, Chungbuk National University, Cheongju 28644, Korea.
e-mail: vnm232657@gmail.com