Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

SOME NOTES ON QUANTUM HELLINGER DIVERGENCES WITH HEINZ MEANS*

NGUYEN LAM† AND REBECCA MILLEY†

Abstract. The information geometry, convexity, in-betweenness property and the barycenter problem of quantum Hellinger divergences with Heinz means is studied. The limiting cases are also considered.

Key words. Positive definite matrix, Hellinger divergence, Convexity, In-betweenness property, Barycenter.

AMS subject classifications. 47A63, 15A60, 81P45, 94A17.

1. Introduction. A recurring problem in many fields of science is to find a structured "point" that satisfies certain properties and conditions and that best approximates a given "point" on a manifold. Hence, the notions of distance measure play key roles in many areas. Typically, one wants to measure the dissimilarity between "points" with a metric distance, a measure that is non-negative; symmetric; zero if and only if the distributions are identical; and obeys the triangle inequality. However, metrics are not always justifiable in applications, and in many settings, it may be better to adjust the distance measure to the context, even when it is not even a metric.

An information-theoretic divergence was introduced in the literature as a distance-like function to represent a degree of separation of two points in a manifold, but it or its square root is not a distance in general. It does not necessarily satisfy the symmetry condition nor triangle inequality. We recall (see [18]) that a dissimilarity $\Phi_\alpha(X,Y)$ between two "points" is called a (asymmetric) divergence if

(D0) $\Phi_\alpha(X,Y) \geq 0$ *and* $\Phi_\alpha(X,Y) = 0$ *if and only if* $X = Y$.

If $\Phi_\alpha(X,Y)$ also satisfies the symmetry property $\Phi_\alpha(X,Y) = \Phi_\alpha(Y,X)$ and the triangle inequality $\Phi_\alpha(X,Y) \leq \Phi_\alpha(X,Z) + \Phi_\alpha(Z,Y)$, then $\Phi_\alpha(X,Y)$ is called a metric. In this paper, we will focus on dissimilarities that are not symmetric in general. Of course, one can easily symmetrize a divergence. However, the asymmetry of divergence plays a meaningful part in information geometry.

The main subject of this note is the divergences on manifold of positive definite matrices. It is worth mentioning that quantum divergences for positive definite matrices have been studied intensively in the literature. We also refer the interested reader to [4, 8, 12, 13, 17, 19, 20, 25, 28, 29, 32, 34, 35, 37, 44, 45, 46], to mention just a few, for various types of quantum divergences, their properties and their applications.

In this paper, we will study some quantum α-divergences that have close connection with the Hellinger distance. We recall that the classical Hellinger distance between two discrete probability distributions $\overrightarrow{p} = (p_1,\ldots,p_n)$ and $\overrightarrow{q} = (q_1,\ldots,q_n)$ is

$$d_H(\overrightarrow{p},\overrightarrow{q}) = \left[\sum \frac{(p_i+q_i)}{2} - \sum \sqrt{p_i q_i}\right]^{\frac{1}{2}}.$$

*Received by the editors on June 6, 2020. Accepted for publication on August 11, 2020. Handling Editor: Fuad Kittaneh. Corresponding Author: Nguyen Lam.

†School of Science and the Environment, Grenfell Campus, Memorial University of Newfoundland, Corner Brook, NL A2H5G4, Canada (nlam@grenfell.mun.ca, rmilley@grenfell.mun.ca).

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

We can observe that

$$d_H\left(\overrightarrow{p},\overrightarrow{q}\right)=\left[\mathrm{Tr}\mathcal{A}\left(\overrightarrow{p},\overrightarrow{q}\right)-\mathrm{Tr}\mathcal{G}\left(\overrightarrow{p},\overrightarrow{q}\right)\right]^{\frac{1}{2}},$$

where $\mathcal{A}\left(\overrightarrow{p},\overrightarrow{q}\right)$ is the arithmetic mean and $\mathcal{G}\left(\overrightarrow{p},\overrightarrow{q}\right)$ is the geometric mean of $\overrightarrow{p}$ and $\overrightarrow{q}$. In the setting of quantum mechanics, we could replace the discrete probability distributions by positive semidefinite density matrices. Also, for positive semidefinite matrices, the matrix arithmetic mean is obvious: $\mathcal{A}(X,Y)=\frac{X+Y}{2}$. However, the geometric mean $\mathcal{G}(X,Y)$ could have different forms and meanings. Therefore, the classical Hellinger divergence has various counterparts in quantum information theory. For instance, if we consider the geometric mean $\mathcal{G}(X,Y)=\left(X^{\frac{1}{2}}YX^{\frac{1}{2}}\right)^{\frac{1}{2}}$, then once again $d(X,Y)=\left[\mathrm{Tr}\left(\frac{X+Y}{2}\right)-\mathrm{Tr}\left(X^{\frac{1}{2}}YX^{\frac{1}{2}}\right)^{\frac{1}{2}}\right]$ is a metric. This is known in the literature as the Bures distance in the quantum information and the Wasserstein metric in optimal transport. It has played an important role in the literature and has been studied intensively and extensively. Note that when X and Y commute, $\left(X^{\frac{1}{2}}YX^{\frac{1}{2}}\right)^{\frac{1}{2}}=X^{\frac{1}{2}}Y^{\frac{1}{2}}$. Hence, it is also natural to consider $\mathcal{G}(X,Y)=X^{\frac{1}{2}}Y^{\frac{1}{2}}$. In this situation, it turns out that

$$d_H(X,Y)=\left[\mathrm{Tr}\left(\frac{X+Y}{2}\right)-\mathrm{Tr}\left(X^{\frac{1}{2}}Y^{\frac{1}{2}}\right)\right]^{\frac{1}{2}}=\left\|X^{\frac{1}{2}}-Y^{\frac{1}{2}}\right\|_2 \tag{1.1}$$

is also a metric.

Probably, the most well-known geometric mean in matrix analysis is

$$X\#Y=X^{\frac{1}{2}}\left(X^{-\frac{1}{2}}YX^{-\frac{1}{2}}\right)^{\frac{1}{2}}X^{\frac{1}{2}}.$$

It was introduced by Pusz and Woronowicz in [39] and has been investigated intensively and extensively in the literature. See, for instance, the monographs [9, 10, 30, 43]. In this case, the authors in [11] showed that the corresponding quantum Hellinger distance $d(X,Y)=\left[\mathrm{Tr}\left(\frac{X+Y}{2}\right)-\mathrm{Tr}\left(X\#Y\right)\right]^{\frac{1}{2}}$ is not a metric because it does not satisfy the triangle inequality. Nevertheless, it was proved that $d^2(X,Y)=\left[\mathrm{Tr}\left(\frac{X+Y}{2}\right)-\mathrm{Tr}\left(X\#Y\right)\right]$ is actually a divergence, satisfies several nice properties, and therefore, it can be used as a good distance measure. Moreover, this divergence has been used by the authors in [11] to study the Karcher mean problem, that is, the minimization problem

$$\min_{X>0}\sum_{j=1}^{m}w_jd^2\left(X,A_j\right).$$

It has to be emphasized that $\#$ is the most well-known geometric mean in matrix analysis and is a very important example of the class of Kubo-Ando operator means. It is also worth noting that the class of Kubo-Ando operator means is one of the most studied objects in the area of operator means in the last thirty years. Hence, in [38], Pitrik and Virosztek introduced and investigated a family of generalized quantum Hellinger divergences of the form

$$\phi(X,Y)=\mathrm{Tr}\left(\left(1-\alpha\right)X+\alpha Y-X\sigma Y\right),$$

where σ is an arbitrary Kubo-Ando mean, and $\alpha\in(0,1)$ is the weight of σ. It was showed that these divergences belong to the family of maximal quantum f -divergences, and hence are jointly convex, and satisfy the data processing inequality. Moreover, there is an intimate relation between generalized quantum Hellinger divergences and operator valued Bregman divergences.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

We should mention here that not all operator means in current use are Kubo–Ando means. One example is the class of Heinz means: $H_\alpha(X,Y) = X^{\frac{1-\alpha}{2}} Y^{\frac{1+\alpha}{2}}$. Obviously, the Heinz mean is not symmetric in general. Hence, sometimes we also want to use the following version of the Heinz mean which is slightly more symmetrical: $H'_\alpha(X,Y) = \frac{1}{2}\left(X^{\frac{1-\alpha}{2}} Y^{\frac{1+\alpha}{2}} + X^{\frac{1+\alpha}{2}} Y^{\frac{1-\alpha}{2}}\right)$. We can consider H_α and H'_α as weighted geometric means.

Motivated by the results in [11] and the generalized quantum Hellinger divergences in [38], the main goal in this note is to investigate the versions of quantum Hellinger divergence using the Heinz means. More precisely, let $\mathbf{P}_n$ be the set of $n \times n$ complex positive definite matrices. For $\alpha \neq \pm 1$ and $X, Y \in \mathbf{P}_n$, we consider

$$\Phi_\alpha(X,Y) = \frac{4}{1-\alpha^2}\left[\frac{1-\alpha}{2}\mathrm{Tr}X + \frac{1+\alpha}{2}\mathrm{Tr}Y - \mathrm{Tr}\left(X^{\frac{1-\alpha}{2}} Y^{\frac{1+\alpha}{2}}\right)\right]$$

and

$$\Psi_\alpha(X,Y) = \frac{2}{1-\alpha^2}\left[\mathrm{Tr}X + \mathrm{Tr}Y - \mathrm{Tr}\left(X^{\frac{1-\alpha}{2}} Y^{\frac{1+\alpha}{2}}\right) - \mathrm{Tr}\left(X^{\frac{1+\alpha}{2}} Y^{\frac{1-\alpha}{2}}\right)\right].$$

By the matrix Young inequality (see, for example, [5, 31]), it is easy to verify that Φ_α and Ψ_α, $\alpha \neq \pm 1$, are divergences on $\mathbf{P}_n$. Moreover, when $\alpha = 0$, $d_0 = \Phi_0^{\frac{1}{2}} = \Psi_0^{\frac{1}{2}}$ is actually a metric on $\mathbf{P}_n$. However, when $\alpha \neq 0$, it is obvious that $\Phi_\alpha^{\frac{1}{2}}$ is not a metric anymore since it does not satisfy the symmetry property. We mention here that some properties of Φ_α was studied in [27] on the set of density matrices, and in [2] on the manifold of real symmetric positive-definite matrices.

It should be noted that there are many difficulties when generalizing classical divergences to the quantum framework. For instance, because of the non-commutative nature, we often could not adapt the proofs of the classical entropy inequalities to the quantum setting. Also, because of the non-commutativity of the matrix setting, even one classical divergence may admit a multiplicity of distinct quantum analogs. Hence, for an operational meaning, besides (D0), a divergence is also required to have other features. For instance, for possible applications in information geometry, we would usually like to verify whether the divergences satisfy the following property (see [1, 2], for instance):

(D1) When X and Y are sufficiently close, by denoting their coordinates by $\boldsymbol{\xi}_X$ and $\boldsymbol{\xi}_Y = \boldsymbol{\xi}_X + d\boldsymbol{\xi}$, the Taylor expansion of Φ_α is written as

$$\Phi_\alpha(\boldsymbol{\xi}_X, \boldsymbol{\xi}_X + d\boldsymbol{\xi}) = \frac{1}{2}\sum g_{ij}(\boldsymbol{\xi}_X)\, d\xi_i d\xi_j + O\left(|d\boldsymbol{\xi}|^3\right),$$

and matrix (g_{ij}) is positive-definite.

Actually, (D0) and (D1) are used to define a divergence in [2] where Amari investigated the dually flatness, decomposability and invariance under linear transformation on manifold of real positive definite matrices. We also note that on $\mathbf{P}_n$, to prove (D1), it is enough to verify:

(D1.1) The first derivative with respect to the second variable vanished on the diagonal:

$$\frac{\partial \Phi_\alpha}{\partial Y}(X,Y)\,|_{Y=X} = 0.$$

(D1.2) The second derivative is positive on the diagonal:

$$\frac{\partial^2 \Phi_\alpha}{\partial Y^2}(X,Y)\,|_{Y=X}(C,C) \geq 0 \quad \text{for all Hermitian } C.$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

The first main result of our note can be read as follows:

Theorem 1.1. *For $\alpha \in \mathbb{R} \setminus \{\pm 1\}$, Φ_α and Ψ_α are divergences on $\mathbf{P}_n$ satisfying (D1.1) and (D1.2).*

Another characteristic that is often required in the theory of quantum information is the data-processing inequality. This property can be considered as one of the most desirable properties of any divergence-type quantity.

(D2) Data-processing inequality: For completely positive trace preserving maps $\mathcal{E}$, one has

$$\Phi_\alpha(\mathcal{E}X, \mathcal{E}Y) \leq \Phi_\alpha(X, Y).$$

Completely positive trace preserving maps are also known as quantum channels in quantum information theory. Applying any such quantum operation can only make the quantum states harder to distinguish. Therefore, for a divergence to have an operational meaning, it must verify the Data Processing Inequality. Actually, this property can be considered as one of the most desirable properties of any divergence-type quantity and has been investigated intensively and extensively in the literature. The interested reader is referred to [6, 7, 15, 16, 33, 36, 40, 41, 42, 47, 48], to name just a few.

It is known that (D2) has a close connection to the convexity of Φ_α. More precisely, the convexity of Φ_α implies (D2) (see, for instance, [26]):

(D2.1) Joint convexity (concavity): For $\lambda \in [0,1]$,

$$\Phi_\alpha(\lambda X_1 + (1-\lambda) X_2, \lambda Y_1 + (1-\lambda) Y_2) \leq (\geq) \lambda \Phi_\alpha(X_1, Y_1) + (1-\lambda) \Phi_\alpha(X_2, Y_2).$$

Also, for possible applications in the Karcher mean problems, we often concern the strict convexity/concavity of Φ_α in the second variable Y :

(D2.2) $Y \to \Phi_\alpha(X, Y)$ is strictly convex/concave.

Our second aim of this article is to show that

Theorem 1.2. *For $\alpha \in (-1,1)$, Φ_α and Ψ_α are jointly convex. Moreover, $\Phi_\alpha(X,Y)$ is strictly convex in Y when $\alpha \in (-1,0]$ and $\Psi_\alpha(X,Y)$ is strictly convex in Y when $\alpha \in (-1,1)$.*

In [3], Audenaert introduced the notions of in-betweenness and monotonicity with respect to a metric for operator means as a relaxation of the notion of geodesity. They are the operator generalization of their natural counterpart for scalar means. More precisely, an operator mean σ satisfies in-betweenness w.r.t. distance d if and only if for all positive X and Y the distance between X and $\sigma(X,Y)$ does not exceed the distance between X and Y: $d(X, \sigma(X,Y)) \leq d(X,Y)$. Also, a weighted operator mean $\sigma(X,Y,t)$ is distance-monotonic if and only if $d(X, \sigma(X,Y,t))$ decreases monotonically with $t \in [0,1]$.

In [3], Audenaert considered the Euclidean distance

$$d(X,Y) = \sqrt{\operatorname{Tr}\left[(X-Y)^* (X-Y)\right]}$$

and showed that the weighted p-power mean $\mu_p(X,Y,t) = (tX^p + (1-t)Y^p)^{\frac{1}{p}}$, $0 \leq t \leq 1$, is Euclidean distance-monotonic and satisfies in-betweenness w.r.t. Euclidean distance for $1 \leq p \leq 2$. The case $p = \frac{1}{2}$ and $p = \frac{1}{4}$ were also established in [21]. In [22], among other results, the authors proved that the weighted

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

p-power mean $\mu_p(X,Y,t)$ also satisfies the in-betweenness property w.r.t. Hellinger metric $d_H(X,Y) = \left[\mathrm{Tr}\left(\frac{X+Y}{2}\right) - \mathrm{Tr}X^{\frac{1}{2}}Y^{\frac{1}{2}}\right]^{\frac{1}{2}}$ for $\frac{1}{2} \leq p \leq 1$. Several geometric properties of the matrix power mean $\mu_p(X,Y,t)$ with respect to different distance functions have been also investigated in [23, 24].

Motivated by the results in [21, 22], our next goal is to prove that:

THEOREM 1.3. *Let* $\alpha \in (-1,1)$. *Then the weighted p-power mean* $\mu_p(X,Y,t)$, $0 \leq t \leq 1$, *satisfies in-betweenness w.r.t.* Φ_α *for* $\max\left\{\frac{1}{2}, \frac{1+\alpha}{2}\right\} \leq p \leq 1$ *and w.r.t.* Ψ_α *for* $\max\left\{\frac{1+\alpha}{2}, \frac{1-\alpha}{2}\right\} \leq p \leq 1$.

We also investigate in Section 5 the above properties in the limiting cases $\alpha = \pm 1$. Finally, as an application of our results, we will study in Section 6 the barycenter problem using the quantum divergence Φ_α.

2. Information geometry of Φ_α and Ψ_α – Proof of Theorem 1.1.

Proof of Theorem 1.1. We recall that

$$\Phi_\alpha(X,Y) = \frac{4}{1-\alpha^2}\left[\frac{1-\alpha}{2}\mathrm{Tr}X + \frac{1+\alpha}{2}\mathrm{Tr}Y - \mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}\right)\right].$$

Hence, we get

$$\frac{\partial \Phi_\alpha}{\partial Y}(X,Y)(B) = \frac{4}{1-\alpha^2}\left[\frac{1+\alpha}{2}\mathrm{Tr}(B) - \mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}D\left(Y^{\frac{1+\alpha}{2}}\right)(B)\right)\right]$$

and

$$\frac{\partial^2 \Phi_\alpha}{\partial Y^2}(X,Y)(B,B) = -\frac{4}{1-\alpha^2}\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}D^2\left(Y^{\frac{1+\alpha}{2}}\right)(B,B)\right).$$

As a consequence,

$$\frac{\partial \Phi_\alpha}{\partial Y}(X,Y=X)(B) = \frac{4}{1-\alpha^2}\left[\frac{1+\alpha}{2}\mathrm{Tr}(B) - \mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}D\left(X^{\frac{1+\alpha}{2}}\right)(B)\right)\right]$$

and

$$\frac{\partial^2 \Phi_\alpha}{\partial Y^2}(X,Y=X)(B,B) = -\frac{4}{1-\alpha^2}\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}D^2\left(X^{\frac{1+\alpha}{2}}\right)(B,B)\right).$$

Using the fact that

$$\mathrm{Tr}\left(\left(X^{\frac{1+\alpha}{2}}\right)^{\frac{2}{1+\alpha}}\right) = \mathrm{Tr}(X),$$

we have

$$\mathrm{Tr}\left(\frac{2}{1+\alpha}\left(X^{\frac{1+\alpha}{2}}\right)^{\frac{2}{1+\alpha}-1}D\left(X^{\frac{1+\alpha}{2}}\right)(B)\right) = \mathrm{Tr}(B).$$

Equivalently,

$$\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}D\left(X^{\frac{1+\alpha}{2}}\right)(B)\right) = \frac{1+\alpha}{2}\mathrm{Tr}(B).$$

Take the derivative both sides again, we obtain

$$\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}D^2\left(X^{\frac{1+\alpha}{2}}\right)(B,B)\right) = -\mathrm{Tr}\left(D\left(X^{\frac{1-\alpha}{2}}\right)(B)\,D\left(X^{\frac{1+\alpha}{2}}\right)(B)\right).$$

In other words,

$$\frac{\partial^2 \Phi_\alpha}{\partial Y^2}(X,Y=X)(B,B) = \frac{4}{1-\alpha^2}\mathrm{Tr}\left(D\left(X^{\frac{1-\alpha}{2}}\right)(B)\,D\left(X^{\frac{1+\alpha}{2}}\right)(B)\right).$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

We now show that

$$\frac{\partial \Phi_\alpha}{\partial Y}\left(X, Y=X\right)\left(B\right)=0$$

and

$$\frac{\partial^2 \Phi_\alpha}{\partial Y^2}\left(X, Y=X\right)\left(B, B\right)\geq 0$$

for all Hermitian matrix B when $\alpha \in \mathbb{R} \setminus \{\pm 1\}$. The argument here is that we can assume that X is a diagonal matrix. Indeed, let $X = U^* X_D U$ be its spectral decomposition, where $X_D = diag\left(\lambda_1, \ldots, \lambda_n\right)$. Then, we note that

$$\begin{aligned} D\left(X^{\frac{1+\alpha}{2}}\right)\left(B\right) &= \lim_{t\to 0}\frac{\left(X+tB\right)^{\frac{1+\alpha}{2}}-X^{\frac{1+\alpha}{2}}}{t} \\ &= \lim_{t\to 0}\frac{\left(U^*X_DU+tU^*UBU^*U\right)^{\frac{1+\alpha}{2}}-\left(U^*X_DU\right)^{\frac{1+\alpha}{2}}}{t} \\ &= \lim_{t\to 0}U^*\frac{\left(X_D+tUBU^*\right)^{\frac{1+\alpha}{2}}-X_D^{\frac{1+\alpha}{2}}}{t}U \\ &= U^*D\left(X^{\frac{1+\alpha}{2}}\right)|_{X=X_D}\left(UBU^*\right)U. \end{aligned}$$

Hence,

$$\begin{aligned} \frac{\partial \Phi_\alpha}{\partial Y}\left(X, Y=X\right)\left(B\right)=0 &= \frac{1+\alpha}{2}\mathrm{Tr}\left(B\right)-\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}D\left(X^{\frac{1+\alpha}{2}}\right)\left(B\right)\right) \\ &= \frac{1+\alpha}{2}\mathrm{Tr}\left(U^*BU\right)-\mathrm{Tr}\left(U^*X_D^{\frac{1-\alpha}{2}}UU^*D\left(X^{\frac{1+\alpha}{2}}\right)|_{X=X_D}\left(UBU^*\right)U\right) \\ &= \frac{1+\alpha}{2}\mathrm{Tr}\left(U^*BU\right)-\mathrm{Tr}\left(X_D^{\frac{1-\alpha}{2}}D\left(X^{\frac{1+\alpha}{2}}\right)|_{X=X_D}\left(UBU^*\right)\right). \end{aligned}$$

Also,

$$\begin{aligned} D^2\left(X^{\frac{1+\alpha}{2}}\right)\left(B, B\right) &= \frac{d^2}{dtds}|_{s=t=0}\left(X+tB+sB\right)^{\frac{1+\alpha}{2}} \\ &= \frac{d^2}{dtds}|_{s=t=0}\left(U^*X_DU+tB+sB\right)^{\frac{1+\alpha}{2}} \\ &= \frac{d^2}{dtds}|_{s=t=0}U^*\left(X_D+tUBU^*+sUBU^*\right)^{\frac{1+\alpha}{2}}U \\ &= U^*D^2\left(X^{\frac{1+\alpha}{2}}\right)|_{X=X_D}\left(UBU^*, UBU^*\right)U. \end{aligned}$$

Hence,

$$\begin{aligned} &\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}D^2\left(X^{\frac{1+\alpha}{2}}\right)\left(B, B\right)\right) \\ &= \mathrm{Tr}\left(U^*X_D^{\frac{1-\alpha}{2}}UU^*D^2\left(X^{\frac{1+\alpha}{2}}\right)|_{X=X_D}\left(UBU^*, UBU^*\right)U\right) \\ &= \mathrm{Tr}\left(X_D^{\frac{1-\alpha}{2}}D^2\left(X^{\frac{1+\alpha}{2}}\right)|_{X=X_D}\left(UBU^*, UBU^*\right)\right). \end{aligned}$$

Therefore, WLOG, we can assume that $X =$diag$(\lambda_1, \ldots, \lambda_n)$. In this case, we get

$$\left[D\left(X^{\frac{1+\alpha}{2}}\right)\left(B\right)\right]_{ij}=\begin{cases} \frac{\lambda_i^{\frac{\alpha+1}{2}}-\lambda_i^{\frac{\alpha+1}{2}}}{\lambda_i-\lambda_j}B_{ij} & \lambda_i\neq\lambda_j \\ \frac{\alpha+1}{2}\lambda_i^{\frac{\alpha-1}{2}}B_{ii} & \lambda_i=\lambda_j \end{cases}.$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

Hence,

$$\begin{aligned}\operatorname{Tr}\left(X^{\frac{1-\alpha}{2}} D\left(X^{\frac{1+\alpha}{2}}\right)(B)\right) &= \sum_{i=1}^{n} \lambda_i^{\frac{1-\alpha}{2}} \frac{\alpha+1}{2} \lambda_i^{\frac{\alpha-1}{2}} B_{ii} \\ &= \frac{\alpha+1}{2} \operatorname{Tr}(B)\end{aligned}$$

and

$$\frac{\partial \Phi_\alpha}{\partial Y}(X, Y=X)(B) = \frac{4}{1-\alpha^2}\left(\frac{\alpha+1}{2}\operatorname{Tr}(B) - \frac{\alpha+1}{2}\operatorname{Tr}(B)\right) = 0.$$

Next, we recall that

$$\frac{\partial^2 \Phi_\alpha}{\partial Y^2}(X,Y)(B,B) = \frac{4}{1-\alpha^2}\operatorname{Tr}\left(D\left(X^{\frac{1-\alpha}{2}}\right)(B)\, D\left(X^{\frac{1+\alpha}{2}}\right)(B)\right).$$

In this case, we get from the Mean Value Theorem that

$$\left[D\left(X^{\frac{1-\alpha}{2}}\right)(B)\right]_{ij} = \begin{cases} \frac{\lambda_i^{\frac{1-\alpha}{2}} - \lambda_i^{\frac{1-\alpha}{2}}}{\lambda_i - \lambda_j} B_{ij} & \lambda_i \neq \lambda_j \\ \frac{1-\alpha}{2}\lambda_i^{\frac{-1-\alpha}{2}} B_{ii} & \lambda_i = \lambda_j \end{cases} = \frac{1-\alpha}{2} s_{ij}^{\frac{-1-\alpha}{2}} B_{ij},$$

where s_{ij} is between λ_i and λ_j.

Simiarly,

$$\left[D\left(X^{\frac{1+\alpha}{2}}\right)(B)\right]_{ij} = \begin{cases} \frac{\lambda_i^{\frac{\alpha+1}{2}} - \lambda_i^{\frac{\alpha+1}{2}}}{\lambda_i - \lambda_j} B_{ij} & \lambda_i \neq \lambda_j \\ \frac{\alpha+1}{2}\lambda_i^{\frac{\alpha-1}{2}} B_{ii} & \lambda_i = \lambda_j \end{cases} = \frac{\alpha+1}{2} t_{ij}^{\frac{\alpha-1}{2}} B_{ij},$$

where t_{ij} is between λ_i and λ_j. Hence,

$$\begin{aligned}\frac{\partial^2 \Phi_\alpha}{\partial Y^2}(X, Y=X)(B,B) &= \frac{4}{1-\alpha^2}\operatorname{Tr}\left(D\left(X^{\frac{1-\alpha}{2}}\right)(B)\, D\left(X^{\frac{1+\alpha}{2}}\right)(B)\right) \\ &= \frac{4}{1-\alpha^2}\left[\sum_{j,k} \frac{1-\alpha}{2} s_{jk}^{\frac{-1-\alpha}{2}} B_{jk} \frac{\alpha+1}{2} t_{kj}^{\frac{\alpha-1}{2}} B_{kj}\right] \\ &= \sum_{j,k} s_{jk}^{\frac{-1-\alpha}{2}} t_{kj}^{\frac{\alpha-1}{2}} |B_{jk}|^2 \geq 0.\end{aligned}$$

Hence, we have that Φ_α satisfies (D1.1) and (D1.2) for all $\alpha \in \mathbb{R} \setminus \{\pm 1\}$.

Now, note that $\Psi_\alpha(X,Y) = \frac{1}{2}\left[\Phi_\alpha(X,Y) + \Phi_{-\alpha}(X,Y)\right]$, we can also conclude that Ψ_α satisfies (D1.1) and (D1.2) for all $\alpha \in \mathbb{R} \setminus \{\pm 1\}$. ∎

3. Convexity – Proof of Theorem 1.2.

Proof of Theorem 1.2. The fact that Φ_α and Ψ_α are jointly convex for $\alpha \in (-1,1)$ is based on the well-known Lieb concavity theorem: $(X,Y) \to \operatorname{Tr}\left(X^{1-t}Y^t\right)$ is jointly concave for any $t \in [0,1]$. See [26, 49].

Now, we will show that $\Phi_\alpha(X,Y)$ is strictly convex in Y when $\alpha \in (-1,0]$. Indeed, it is enough to show that $Y \to \operatorname{Tr}\left(X^{\frac{1-\alpha}{2}} Y^{\frac{1+\alpha}{2}}\right)$ is strictly concave when $\alpha \in (-1,0]$.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

First, since $Y \to Y^{\frac{1+\alpha}{2}}$ is a matrix concave function for all $\alpha \in (-1,1)$, we have

$$(tY_1 + (1-t)\,Y_2)^{\frac{1+\alpha}{2}} \geq tY_1^{\frac{1+\alpha}{2}} + (1-t)\,Y_2^{\frac{1+\alpha}{2}}. \tag{3.2}$$

Hence,

$$X^{\frac{1-\alpha}{4}}\,(tY_1 + (1-t)\,Y_2)^{\frac{1+\alpha}{2}}\,X^{\frac{1-\alpha}{4}} \geq tX^{\frac{1-\alpha}{4}}Y_1^{\frac{1+\alpha}{2}}X^{\frac{1-\alpha}{4}} + (1-t)\,X^{\frac{1-\alpha}{4}}Y_2^{\frac{1+\alpha}{2}}X^{\frac{1-\alpha}{4}}$$

and we get for all $\alpha \in (-1,1)$

$$\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}\,(tY_1 + (1-t)\,Y_2)^{\frac{1+\alpha}{2}}\right) \geq t\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}Y_1^{\frac{1+\alpha}{2}}\right) + (1-t)\,\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}Y_2^{\frac{1+\alpha}{2}}\right).$$

That is $\Phi_\alpha\,(X,Y)$ is convex in Y when $\alpha \in (-1,1)$.

Now, if

$$\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}\,(tY_1 + (1-t)\,Y_2)^{\frac{1+\alpha}{2}}\right) = t\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}Y_1^{\frac{1+\alpha}{2}}\right) + (1-t)\,\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}Y_2^{\frac{1+\alpha}{2}}\right),$$

then since

$$X^{\frac{1-\alpha}{4}}\,(tY_1 + (1-t)\,Y_2)^{\frac{1+\alpha}{2}}\,X^{\frac{1-\alpha}{4}} - tX^{\frac{1-\alpha}{4}}Y_1^{\frac{1+\alpha}{2}}X^{\frac{1-\alpha}{4}} - (1-t)\,X^{\frac{1-\alpha}{4}}Y_2^{\frac{1+\alpha}{2}}X^{\frac{1-\alpha}{4}} \geq 0,$$

we deduce

$$X^{\frac{1-\alpha}{4}}\,(tY_1 + (1-t)\,Y_2)^{\frac{1+\alpha}{2}}\,X^{\frac{1-\alpha}{4}} = tX^{\frac{1-\alpha}{4}}Y_1^{\frac{1+\alpha}{2}}X^{\frac{1-\alpha}{4}} + (1-t)\,X^{\frac{1-\alpha}{4}}Y_2^{\frac{1+\alpha}{2}}X^{\frac{1-\alpha}{4}}$$

and

$$(tY_1 + (1-t)\,Y_2)^{\frac{1+\alpha}{2}} = tY_1^{\frac{1+\alpha}{2}} + (1-t)\,Y_2^{\frac{1+\alpha}{2}}.$$

If $\alpha = 0$, then

$$(tY_1 + (1-t)\,Y_2)^{\frac{1}{2}} = tY_1^{\frac{1}{2}} + (1-t)\,Y_2^{\frac{1}{2}}.$$

Square both sides to get

$$tY_1 + (1-t)\,Y_2 = t^2Y_1 + (1-t)^2\,Y_2 + t\,(1-t)\left[Y_1^{\frac{1}{2}}Y_2^{\frac{1}{2}} + Y_2^{\frac{1}{2}}Y_1^{\frac{1}{2}}\right].$$

Equivalently,

$$t\,(1-t)\left[Y_1^{\frac{1}{2}} - Y_2^{\frac{1}{2}}\right]^2 = 0.$$

That is $Y_1^{\frac{1}{2}} = Y_2^{\frac{1}{2}}$, and hence, $Y_1 = Y_2$.

If $-1 < \alpha < 0$, then $0 < 1+\alpha < 1$. Squaring both sides of $(tY_1 + (1-t)\,Y_2)^{\frac{1+\alpha}{2}} = tY_1^{\frac{1+\alpha}{2}} + (1-t)\,Y_2^{\frac{1+\alpha}{2}}$ yields

$$\begin{aligned}
&(tY_1 + (1-t)\,Y_2)^{1+\alpha} \\
&= t^2Y_1^{1+\alpha} + (1-t)^2\,Y_2^{1+\alpha} + t\,(1-t)\left[Y_1^{\frac{1+\alpha}{2}}Y_2^{\frac{1+\alpha}{2}} + Y_2^{\frac{1+\alpha}{2}}Y_1^{\frac{1+\alpha}{2}}\right] \\
&\leq \left[t^2 + t\,(1-t)\right]Y_1^{1+\alpha} + \left[(1-t)^2 + t\,(1-t)\right]Y_2^{1+\alpha} \\
&= tY_1^{1+\alpha} + (1-t)\,Y_2^{1+\alpha}.
\end{aligned}$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

Here, we applied the inequality

$$Y_1^{\frac{1+\alpha}{2}} Y_2^{\frac{1+\alpha}{2}} + Y_2^{\frac{1+\alpha}{2}} Y_1^{\frac{1+\alpha}{2}} \leq Y_1^{1+\alpha} + Y_2^{1+\alpha}.$$

Since $1+\alpha < 1$, we get by (3.2) that

$$(tY_1 + (1-t)\,Y_2)^{1+\alpha} \geq tY_1^{1+\alpha} + (1-t)\,Y_2^{1+\alpha}.$$

Hence,

$$(tY_1 + (1-t)\,Y_2)^{1+\alpha} = tY_1^{1+\alpha} + (1-t)\,Y_2^{1+\alpha}.$$

That means,

$$Y_1^{\frac{1+\alpha}{2}} Y_2^{\frac{1+\alpha}{2}} + Y_2^{\frac{1+\alpha}{2}} Y_1^{\frac{1+\alpha}{2}} \leq Y_1^{1+\alpha} + Y_2^{1+\alpha}$$

and

$$\left(Y_1^{\frac{1+\alpha}{2}} - Y_2^{\frac{1+\alpha}{2}}\right)^2 = 0.$$

We deduce $Y_1^{\frac{1+\alpha}{2}} = Y_2^{\frac{1+\alpha}{2}}$ and $Y_1 = Y_2$.

Next, we will prove that $\Psi_\alpha(X,Y)$ is strictly convex in Y when $\alpha \in (-1,1)$. Again, it is enough to show that $Y \to \operatorname{Tr}\left(X^{\frac{1-\alpha}{2}} Y^{\frac{1+\alpha}{2}}\right) + \operatorname{Tr}\left(X^{\frac{1+\alpha}{2}} Y^{\frac{1-\alpha}{2}}\right)$ is strictly concave when $\alpha \in (-1,1)$. The concavity of $Y \to \operatorname{Tr}\left(X^{\frac{1-\alpha}{2}} Y^{\frac{1+\alpha}{2}}\right) + \operatorname{Tr}\left(X^{\frac{1+\alpha}{2}} Y^{\frac{1-\alpha}{2}}\right)$ is obvious since $Y \to \operatorname{Tr}\left(X^{\frac{1-\alpha}{2}} Y^{\frac{1+\alpha}{2}}\right)$ and $Y \to \operatorname{Tr}\left(X^{\frac{1+\alpha}{2}} Y^{\frac{1-\alpha}{2}}\right)$ are concave. Now, noting that $Y \to \operatorname{Tr}\left(X^{\frac{1-\alpha}{2}} Y^{\frac{1+\alpha}{2}}\right)$ is strictly concave if $\alpha \in (-1,0]$, and $Y \to \operatorname{Tr}\left(X^{\frac{1+\alpha}{2}} Y^{\frac{1-\alpha}{2}}\right)$ is strictly concave if $\alpha \in [0,1)$, we deduce that $Y \to \operatorname{Tr}\left(X^{\frac{1-\alpha}{2}} Y^{\frac{1+\alpha}{2}}\right) + \operatorname{Tr}\left(X^{\frac{1+\alpha}{2}} Y^{\frac{1-\alpha}{2}}\right)$ is strictly concave when $\alpha \in (-1,1)$. ∎

4. In-betweenness of weighted power mean – Proof of Theorem 1.3.

Proof of Theorem 1.3. We need to show that $\Phi_\alpha(X, \mu_p(X,Y,t)) \leq \Phi_\alpha(X,Y)$. Equivalently,

$$\begin{aligned}&\frac{1-\alpha}{2}\operatorname{Tr}X + \frac{1+\alpha}{2}\operatorname{Tr}\mu_p(X,Y,t) - \operatorname{Tr}\left(X^{\frac{1-\alpha}{2}} \mu_p(X,Y,t)^{\frac{1+\alpha}{2}}\right)\\ &\leq \frac{1-\alpha}{2}\operatorname{Tr}X + \frac{1+\alpha}{2}\operatorname{Tr}Y - \operatorname{Tr}\left(X^{\frac{1-\alpha}{2}} Y^{\frac{1+\alpha}{2}}\right)\end{aligned}$$

and

$$\begin{aligned}&\frac{1+\alpha}{2}\operatorname{Tr}\mu_p(X,Y,t) - \operatorname{Tr}\left(X^{\frac{1-\alpha}{2}} \mu_p(X,Y,t)^{\frac{1+\alpha}{2}}\right)\\ &\leq \frac{1+\alpha}{2}\operatorname{Tr}Y - \operatorname{Tr}\left(X^{\frac{1-\alpha}{2}} Y^{\frac{1+\alpha}{2}}\right).\end{aligned}$$

Note that since $1 \leq \frac{1}{p} \leq 2$, the function $x^{\frac{1}{p}}$ is operator convex. Hence,

$$\mu_p(X,Y,t) = (tX^p + (1-t)\,Y^p)^{\frac{1}{p}} \leq tX + (1-t)\,Y$$

and

$$\operatorname{Tr}\mu_p(X,Y,t) \leq \operatorname{Tr}\left[tX + (1-t)\,Y\right].$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

Hence, it now is enough to prove

$$\mathrm{Tr}\left[\frac{1+\alpha}{2}tX+\frac{1+\alpha}{2}\left(1-t\right)Y-X^{\frac{1-\alpha}{2}}\mu_p\left(X,Y,t\right)^{\frac{1+\alpha}{2}}\right]\leq\mathrm{Tr}\left[\frac{1+\alpha}{2}Y-X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}\right]$$

or

$$\mathrm{Tr}\left[\frac{1+\alpha}{2}t\left(X-Y\right)+X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}\right]\leq\mathrm{Tr}\left[X^{\frac{1-\alpha}{2}}\mu_p\left(X,Y,t\right)^{\frac{1+\alpha}{2}}\right].$$

Now, since $0<\frac{1+\alpha}{2p}\leq 1$, the function $x^{\frac{1+\alpha}{2p}}$ is operator concave. Therefore,

$$\mu_p\left(X,Y,t\right)^{\frac{1+\alpha}{2}}=\left(tX^p+\left(1-t\right)Y^p\right)^{\frac{1+\alpha}{2p}}\geq tX^{\frac{1+\alpha}{2}}+\left(1-t\right)Y^{\frac{1+\alpha}{2}}$$

and

$$\mathrm{Tr}\left[X^{\frac{1-\alpha}{2}}\mu_p\left(X,Y,t\right)^{\frac{1+\alpha}{2}}\right]\geq\mathrm{Tr}\left[tX+\left(1-t\right)X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}\right].$$

Hence, the proof is completed if we can verify

$$\mathrm{Tr}\left[tX+\left(1-t\right)X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}\right]\geq\mathrm{Tr}\left[\frac{1+\alpha}{2}t\left(X-Y\right)+X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}\right],$$

or equivalently,

$$\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}\right)\leq\mathrm{Tr}\left[\frac{1-\alpha}{2}X+\frac{1+\alpha}{2}Y^{\frac{1+\alpha}{2}}\right].$$

However, this is nothing but a consequence of the matrix Young inequality [5].

Now, we will show that $\Psi_\alpha\left(X,\mu_p\left(X,Y,t\right)\right)\leq\Psi_\alpha\left(X,Y\right)$, or equivalently,

$$\begin{aligned}&\mathrm{Tr}X+\mathrm{Tr}\mu_p\left(X,Y,t\right)-\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}\mu_p\left(X,Y,t\right)^{\frac{1+\alpha}{2}}\right)-\mathrm{Tr}\left(X^{\frac{1+\alpha}{2}}\mu_p\left(X,Y,t\right)^{\frac{1-\alpha}{2}}\right)\\&\leq\mathrm{Tr}X+\mathrm{Tr}Y-\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}\right)-\mathrm{Tr}\left(X^{\frac{1+\alpha}{2}}Y^{\frac{1-\alpha}{2}}\right)\end{aligned}$$

and

$$\begin{aligned}&\mathrm{Tr}\mu_p\left(X,Y,t\right)-\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}\mu_p\left(X,Y,t\right)^{\frac{1+\alpha}{2}}\right)-\mathrm{Tr}\left(X^{\frac{1+\alpha}{2}}\mu_p\left(X,Y,t\right)^{\frac{1-\alpha}{2}}\right)\\&\leq\mathrm{Tr}Y-\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}\right)-\mathrm{Tr}\left(X^{\frac{1+\alpha}{2}}Y^{\frac{1-\alpha}{2}}\right).\end{aligned}$$

Note that since $1\leq\frac{1}{p}\leq 2$, the function $x^{\frac{1}{p}}$ is operator convex. Hence,

$$\mu_p\left(X,Y,t\right)=\left(tX^p+\left(1-t\right)Y^p\right)^{\frac{1}{p}}\leq tX+\left(1-t\right)Y$$

and

$$\mathrm{Tr}\mu_p\left(X,Y,t\right)\leq\mathrm{Tr}\left[tX+\left(1-t\right)Y\right].$$

Hence, it now is enough to prove

$$\begin{aligned}&\mathrm{Tr}\left[tX+\left(1-t\right)Y\right]-\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}\mu_p\left(X,Y,t\right)^{\frac{1+\alpha}{2}}\right)-\mathrm{Tr}\left(X^{\frac{1+\alpha}{2}}\mu_p\left(X,Y,t\right)^{\frac{1-\alpha}{2}}\right)\\&\leq\mathrm{Tr}Y-\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}\right)-\mathrm{Tr}\left(X^{\frac{1+\alpha}{2}}Y^{\frac{1-\alpha}{2}}\right)\end{aligned}$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

or

$$\begin{aligned}&\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}\mu_p(X,Y,t)^{\frac{1+\alpha}{2}}\right)+\mathrm{Tr}\left(X^{\frac{1+\alpha}{2}}\mu_p(X,Y,t)^{\frac{1-\alpha}{2}}\right)\\&\geq \mathrm{Tr}\left[t(X-Y)+X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}+X^{\frac{1+\alpha}{2}}Y^{\frac{1-\alpha}{2}}\right].\end{aligned}$$

Note that since $0<\frac{1+\alpha}{2p}\leq 1$ and $0<\frac{1-\alpha}{2p}\leq 1$, the function $x^{\frac{1+\alpha}{2p}}$ and $x^{\frac{1-\alpha}{2p}}$ are operator concave. Hence,

$$\begin{aligned}\mu_p(X,Y,t)^{\frac{1+\alpha}{2}}&=(tX^p+(1-t)Y^p)^{\frac{1+\alpha}{2p}}\geq tX^{\frac{1+\alpha}{2}}+(1-t)Y^{\frac{1+\alpha}{2}},\\ \mu_p(X,Y,t)^{\frac{1-\alpha}{2}}&=(tX^p+(1-t)Y^p)^{\frac{1-\alpha}{2p}}\geq tX^{\frac{1-\alpha}{2}}+(1-t)Y^{\frac{1-\alpha}{2}}\end{aligned}$$

and

$$\begin{aligned}\mathrm{Tr}\left[X^{\frac{1-\alpha}{2}}\mu_p(X,Y,t)^{\frac{1+\alpha}{2}}\right]&\geq \mathrm{Tr}\left[tX+(1-t)X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}\right],\\ \mathrm{Tr}\left(X^{\frac{1+\alpha}{2}}\mu_p(X,Y,t)^{\frac{1-\alpha}{2}}\right)&\geq \mathrm{Tr}\left[tX+(1-t)X^{\frac{1+\alpha}{2}}Y^{\frac{1-\alpha}{2}}\right].\end{aligned}$$

Hence, it is enough to show

$$\begin{aligned}&\mathrm{Tr}\left[tX+(1-t)X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}\right]+\mathrm{Tr}\left[tX+(1-t)X^{\frac{1+\alpha}{2}}Y^{\frac{1-\alpha}{2}}\right]\\&\geq \mathrm{Tr}\left[t(X-Y)+X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}+X^{\frac{1+\alpha}{2}}Y^{\frac{1-\alpha}{2}}\right].\end{aligned}$$

Equivalently,

$$\mathrm{Tr}\left[X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}+X^{\frac{1+\alpha}{2}}Y^{\frac{1-\alpha}{2}}\right]\leq \mathrm{Tr}\left[X+Y\right],$$

or

$$\Psi_\alpha(X,Y)\geq 0.$$

This is obvous by the matrix Young inequality. ∎

5. The limiting cases $\alpha=\pm1$.

Recall that for $\alpha\neq\pm1$,

$$\Phi_\alpha(X,Y)=\frac{4}{1-\alpha^2}\left[\frac{1-\alpha}{2}\mathrm{Tr}X+\frac{1+\alpha}{2}\mathrm{Tr}Y-\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}\right)\right]$$

and

$$\Psi_\alpha(X,Y)=\frac{2}{1-\alpha^2}\left[\mathrm{Tr}X+\mathrm{Tr}Y-\mathrm{Tr}\left(X^{\frac{1-\alpha}{2}}Y^{\frac{1+\alpha}{2}}\right)-\mathrm{Tr}\left(X^{\frac{1+\alpha}{2}}Y^{\frac{1-\alpha}{2}}\right)\right].$$

Using L'Hôpital's rule, when $\alpha=\pm1$, we consider

$$\begin{aligned}\Phi_1(X,Y)&=\mathrm{Tr}(X-Y-Y\ln X+Y\ln Y),\\ \Phi_{-1}(X,Y)&=\mathrm{Tr}(Y-X-X\ln Y+X\ln X)=\Phi_1(Y,X),\\ \Psi_1(X,Y)&=\Psi_{-1}(X,Y)=\frac{1}{2}\mathrm{Tr}(X\ln X+Y\ln Y-X\ln Y-Y\ln X).\end{aligned}$$

Obviously, they are quantum divergences by the Klein inequality for the strictly convex function $f(t)=t\ln t$. See [14]. We now will show that they also satisfy the properties (D1.1) and (D1.2).

THEOREM 5.1. $\Phi_{\pm1}$ *and* $\Psi_{\pm1}$ *satisfy the properties (D1.1) and (D1.2).*

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

Proof. We recall that

$$\ln X = \int_0^\infty (t+1)^{-1}\,\mathbb{I}_n - (t\mathbb{I}_n + X)^{-1}\,dt$$

and

$$D(\ln X)(B) = \int_0^\infty (t\mathbb{I}_n + X)^{-1} B (t\mathbb{I}_n + X)^{-1}\,dt,$$

$$D^2(\ln X)(B,B) = -2\int_0^\infty (t\mathbb{I}_n + X)^{-1} B (t\mathbb{I}_n + X)^{-1} B (t\mathbb{I}_n + X)^{-1}\,dt.$$

Also,

$$D(\mathrm{Tr}(f(X)))(B) = \mathrm{Tr}(Bf'(X)).$$

Hence, for $f(t) = t\ln t$, we get

$$\begin{aligned}\mathrm{Tr}(B(\ln X + \mathbb{I}_n)) &= D(\mathrm{Tr}(X\ln X))(B)\\ &= \mathrm{Tr}\left(B\ln X + X\int_0^\infty (t\mathbb{I}_n + X)^{-1} B (t\mathbb{I}_n + X)^{-1}\,dt\right).\end{aligned}$$

Hence, we can deduce

$$\mathrm{Tr}B = \mathrm{Tr}\left(X\int_0^\infty (t\mathbb{I}_n + X)^{-1} B (t\mathbb{I}_n + X)^{-1}\,dt\right). \tag{5.3}$$

Then, a direct computation shows

$$\begin{aligned}\frac{\partial\Phi_1}{\partial Y}(X,Y)(B) &= \mathrm{Tr}(-B - B\ln X + B(\ln Y + \mathbb{I}_n))\\ &= \mathrm{Tr}(B(\ln Y - \ln X)),\end{aligned}$$

$$\frac{\partial\Phi_{-1}}{\partial Y}(X,Y)(B) = \mathrm{Tr}\left(B - X\int_0^\infty (t\mathbb{I}_n + Y)^{-1} B (t\mathbb{I}_n + Y)^{-1}\,dt\right),$$

and

$$\frac{\partial\Psi_{\pm1}}{\partial Y}(X,Y)(B) = \frac{1}{2}\mathrm{Tr}\left(B(\ln Y + \mathbb{I}_n) - X\int_0^\infty (t\mathbb{I}_n + Y)^{-1} B (t\mathbb{I}_n + Y)^{-1}\,dt - B\ln X\right).$$

Hence, using (5.3), we get

$$\frac{\partial\Phi_1}{\partial Y}(X,Y=X)(B) = \frac{\partial\Phi_{-1}}{\partial Y}(X,Y=X)(B) = \frac{\partial\Psi_{\pm1}}{\partial Y}(X,Y=X)(B) = 0.$$

Now, since

$$\frac{\partial\Phi_1}{\partial Y}(X,Y)(B) = \mathrm{Tr}(B(\ln Y - \ln X)),$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

we get

$$\begin{aligned}\frac{\partial^2\Phi_1}{\partial Y^2}(X,Y=X)(B,B) &= \mathrm{Tr}\left(B\int_0^\infty (t\mathbb{I}_n+X)^{-1}B(t\mathbb{I}_n+X)^{-1}dt\right)\\ &= \int_0^\infty \mathrm{Tr}\left(B(t\mathbb{I}_n+X)^{-1}B(t\mathbb{I}_n+X)^{-1}\right)dt\\ &\geq 0\end{aligned}$$

since $B(t\mathbb{I}_n+X)^{-1}B\geq 0$ and $(t\mathbb{I}_n+X)^{-1}\geq 0$.

Also, since

$$\frac{\partial\Phi_{-1}}{\partial Y}(X,Y)(B) = \mathrm{Tr}\left(B - X\int_0^\infty (t\mathbb{I}_n+Y)^{-1}B(t\mathbb{I}_n+Y)^{-1}dt\right),$$

we deduce

$$\frac{\partial^2\Phi_{-1}}{\partial Y^2}(X,Y=X)(B,B) = 2\mathrm{Tr}\left(X\int_0^\infty (t\mathbb{I}_n+X)^{-1}B(t\mathbb{I}_n+X)^{-1}B(t\mathbb{I}_n+X)^{-1}dt\right)\geq 0$$

since

$$(t\mathbb{I}_n+X)^{-1}B(t\mathbb{I}_n+X)^{-1}B(t\mathbb{I}_n+X)^{-1} = \left(B(t\mathbb{I}_n+X)^{-1}\right)^*(t\mathbb{I}_n+X)^{-1}B(t\mathbb{I}_n+X)^{-1}\geq 0$$

and $X\geq 0$.

Finally, from

$$\frac{\partial\Psi_{\pm 1}}{\partial Y}(X,Y)(B) = \frac{1}{2}\mathrm{Tr}\left(B(\ln Y+\mathbb{I}_n) - X\int_0^\infty (t\mathbb{I}_n+Y)^{-1}B(t\mathbb{I}_n+Y)^{-1}dt - B\ln X\right),$$

we have

$$\begin{aligned}&\frac{\partial^2\Psi_{\pm 1}}{\partial Y^2}(X,Y=X)(B,B)\\ &= \frac{1}{2}\mathrm{Tr}\begin{pmatrix} B\int_0^\infty (t\mathbb{I}_n+X)^{-1}B(t\mathbb{I}_n+X)^{-1}dt \\ +2X\int_0^\infty (t\mathbb{I}_n+X)^{-1}B(t\mathbb{I}_n+X)^{-1}B(t\mathbb{I}_n+X)^{-1}dt\end{pmatrix}\\ &= \frac{1}{2}\mathrm{Tr}\left(\int_0^\infty B(t\mathbb{I}_n+X)^{-1}B(t\mathbb{I}_n+X)^{-1}dt\right)\\ &+\mathrm{Tr}\left(X\int_0^\infty (t\mathbb{I}_n+X)^{-1}B(t\mathbb{I}_n+X)^{-1}B(t\mathbb{I}_n+X)^{-1}dt\right)\\ &\geq 0. \qquad \square\end{aligned}$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

Next, we will show that the weighted power mean satisfies in-betweenness w.r.t. Φ_{-1} for $\frac{1}{2} \leq p \leq 1:$

THEOREM 5.2. *The weighted p-power mean* $\mu_p(X,Y,t) = (tX^p + (1-t)Y^p)^{\frac{1}{p}}$, $0 \leq t \leq 1$, *satisfies in-betweenness w.r.t.* Φ_{-1} *for* $\frac{1}{2} \leq p \leq 1$.

Proof. We need to show that $\Phi_{-1}(X, \mu_p(X,Y,t)) \leq \Phi_{-1}(X,Y)$. Equivalently,

$$\mathrm{Tr}\left(\mu_p(X,Y,t) - X - X\ln\mu_p(X,Y,t) + X\ln X\right) \leq \mathrm{Tr}\left(Y - X - X\ln Y + X\ln X\right)$$

and

$$\mathrm{Tr}\left(\mu_p(X,Y,t) - X\ln\mu_p(X,Y,t)\right) \leq \mathrm{Tr}\left(Y - X\ln Y.\right).$$

Note that since $1 \leq \frac{1}{p} \leq 2$, the function $x^{\frac{1}{p}}$ is operator convex. Hence,

$$\mu_p(X,Y,t) = (tX^p + (1-t)Y^p)^{\frac{1}{p}} \leq tX + (1-t)Y$$

and

$$\mathrm{Tr}\mu_p(X,Y,t) \leq \mathrm{Tr}\left[tX + (1-t)Y\right].$$

Hence, it now is enough to prove

$$\mathrm{Tr}\left[tX + (1-t)Y - X\ln\mu_p(X,Y,t)\right] \leq \mathrm{Tr}\left[Y - X\ln Y\right],$$

or

$$\mathrm{Tr}\left[t(X-Y) + X\ln Y\right] \leq \mathrm{Tr}\left[X\ln\mu_p(X,Y,t)\right].$$

Now, since the function $\ln x$ is operator concave, we get

$$\ln\mu_p(X,Y,t) = \ln(tX^p + (1-t)Y^p)^{\frac{1}{p}} = \frac{1}{p}\ln(tX^p + (1-t)Y^p) \geq t\ln X + (1-t)\ln Y$$

and

$$\mathrm{Tr}\left[X\ln\mu_p(X,Y,t)\right] \geq \mathrm{Tr}\left[tX\ln X + (1-t)X\ln Y\right].$$

Hence, the proof is completed if we can verify

$$\mathrm{Tr}\left[tX\ln X + (1-t)X\ln Y\right] \geq \mathrm{Tr}\left[t(X-Y) + X\ln Y\right],$$

or equivalently,

$$\mathrm{Tr}\left(Y - X - X\ln Y + X\ln X\right) \geq 0.$$

However, this is nothing but a consequence of the Klein inequality. Indeed, since the function $f(x) = x\ln x$ is convex on $(0,\infty)$, by the Klein inequality ([14]), we have

$$\mathrm{Tr}\left[f(X) - f(Y) - (X-Y)f'(Y)\right] \geq 0.$$

Equivalently,

$$\mathrm{Tr}\left[X\ln X - Y\ln Y - (X-Y)(\ln Y + \mathbb{I}_n)\right] \geq 0$$

and

$$\mathrm{Tr}\left(Y - X - X\ln Y + X\ln X\right) \geq 0. \qquad \square$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

6. Barycenters. In this section, we assume that $\alpha \in (-1, 1)$ and recall that

$$\Phi_\alpha (X, Y) = \frac{4}{1-\alpha^2}\left[\frac{1-\alpha}{2}\mathrm{Tr}X + \frac{1+\alpha}{2}\mathrm{Tr}Y - \mathrm{Tr}\left(X^{\frac{1-\alpha}{2}} Y^{\frac{1+\alpha}{2}}\right)\right].$$

Let $A_1, \ldots, A_k$ be a k-tuple in $\mathbf{P}_n$ and $w_1, \ldots, w_k$ be positive weights such that $\sum_{j=1}^{k} w_j = 1$. In this section, we consider the barycenter problem

$$\min_{X \in \mathbf{P}_n} \sum_{j=1}^{k} w_j \Phi_\alpha (A_j, X). \tag{6.4}$$

We will prove the following result:

THEOREM 6.1. *The minimization problem (6.4) has a unique solution*

$$X = \left(\sum_{j=1}^{k} w_j A_j^{\frac{1-\alpha}{2}}\right)^{\frac{2}{1-\alpha}}.$$

Proof. By Theorem 1.2, the function $\Psi (X) = \sum_{j=1}^{k} w_j \Phi_\alpha (A_j, X)$ is convex. Therefore, a critical point of Ψ is the global minimum of Ψ. To find such critical points, we will solve the equation $D\Psi (X) = 0$.

Note that

$$\begin{aligned} D\Psi (X) (B) &= \sum_{j=1}^{k} w_j \frac{\partial \Phi_\alpha}{\partial X} (A_j, X) \\ &= \sum_{j=1}^{k} w_j \frac{4}{1-\alpha^2}\left[\frac{1+\alpha}{2}\mathrm{Tr}B - \mathrm{Tr}\left(A_j^{\frac{1-\alpha}{2}} D\left(X^{\frac{1+\alpha}{2}}\right)(B)\right)\right] \\ &= \frac{4}{1-\alpha^2}\left(\frac{1+\alpha}{2}\mathrm{Tr}B - \sum_{j=1}^{k} w_j \mathrm{Tr}\left(A_j^{\frac{1-\alpha}{2}} D\left(X^{\frac{1+\alpha}{2}}\right)(B)\right)\right). \end{aligned}$$

We now recall the following identity (see [9, page 143])

$$x^{\frac{1+\alpha}{2}} = \cos\left(\frac{1+\alpha}{4}\pi\right) + \frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi}\int_0^\infty \left(\frac{\lambda}{\lambda^2+1} - \frac{1}{\lambda + x}\right)\lambda^{\frac{1+\alpha}{2}} d\lambda.$$

Hence,

$$X^{\frac{1+\alpha}{2}} = \cos\left(\frac{1+\alpha}{4}\pi\right)\mathbb{I}_n + \frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi}\int_0^\infty \left(\frac{\lambda}{\lambda^2+1}\mathbb{I}_n - (\lambda\mathbb{I}_n + X)^{-1}\right)\lambda^{\frac{1+\alpha}{2}} d\lambda$$

and

$$D\left(X^{\frac{1+\alpha}{2}}\right)(B) = \frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi}\int_0^\infty (\lambda\mathbb{I}_n + X)^{-1} B (\lambda\mathbb{I}_n + X)^{-1}\lambda^{\frac{1+\alpha}{2}} d\lambda.$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

Therefore,

$$\begin{aligned}
&\sum_{j=1}^{k} w_j \mathrm{Tr}\left(A_j^{\frac{1-\alpha}{2}} D\left(X^{\frac{1+\alpha}{2}}\right)(B)\right) \\
&= \sum_{j=1}^{k} w_j \mathrm{Tr}\left(A_j^{\frac{1-\alpha}{2}} \frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi} \int_0^{\infty} (\lambda \mathbb{I}_n + X)^{-1} B (\lambda \mathbb{I}_n + X)^{-1} \lambda^{\frac{1+\alpha}{2}} d\lambda\right) \\
&= \mathrm{Tr}\left(\sum_{j=1}^{k} w_j A_j^{\frac{1-\alpha}{2}} \int_0^{\infty} (\lambda \mathbb{I}_n + X)^{-1} B (\lambda \mathbb{I}_n + X)^{-1} \frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi} \lambda^{\frac{1+\alpha}{2}} d\lambda\right) \\
&= \mathrm{Tr}\left(\int_0^{\infty} (\lambda \mathbb{I}_n + X)^{-1} \left(\sum_{j=1}^{k} w_j A_j^{\frac{1-\alpha}{2}}\right) (\lambda \mathbb{I}_n + X)^{-1} B \frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi} \lambda^{\frac{1+\alpha}{2}} d\lambda\right).
\end{aligned}$$

Hence, the equation $D\Psi(X) = 0$ is equivalent to

$$\frac{1+\alpha}{2}\mathbb{I}_n = \int_0^{\infty} (\lambda \mathbb{I}_n + X)^{-1} \left(\sum_{j=1}^{k} w_j A_j^{\frac{1-\alpha}{2}}\right) (\lambda \mathbb{I}_n + X)^{-1} \frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi} \lambda^{\frac{1+\alpha}{2}} d\lambda.$$

Now, let $X = U^* diag(x_i) U$, $Z = \left(\sum_{j=1}^{k} w_j A_j^{\frac{1-\alpha}{2}}\right)$ and $Y = UZU^*$. Then

$$\begin{aligned}
\frac{1+\alpha}{2}\mathbb{I}_n &= \int_0^{\infty} (\lambda + X)^{-1} \left(\sum_{j=1}^{k} w_j A_j^{\frac{1-\alpha}{2}}\right) (\lambda + X)^{-1} \frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi} \lambda^{\frac{1+\alpha}{2}} d\lambda \\
&= U^* \int_0^{\infty} (\lambda + diag(x_i))^{-1} U \left(\sum_{j=1}^{k} w_j A_j^{\frac{1-\alpha}{2}}\right) U^* (\lambda + diag(x_i))^{-1} \frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi} \lambda^{\frac{1+\alpha}{2}} d\lambda U \\
&= U^* \int_0^{\infty} (\lambda + diag(x_i))^{-1} Y (\lambda + diag(x_i))^{-1} \frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi} \lambda^{\frac{1+\alpha}{2}} d\lambda U.
\end{aligned}$$

Hence, we have

$$\frac{1+\alpha}{2}\mathbb{I}_n = \int_0^{\infty} (\lambda + diag(x_i))^{-1} Y (\lambda + diag(x_i))^{-1} \frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi} \lambda^{\frac{1+\alpha}{2}} d\lambda,$$

which implies

$$\int_0^{\infty} \frac{y_{ij}}{(\lambda + x_i)(\lambda + x_j)} \frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi} \lambda^{\frac{1+\alpha}{2}} d\lambda = \frac{1+\alpha}{2}\delta_{ij}.$$

This means that Y is diagonal with

$$\frac{1}{y_{ii}} = \frac{2}{1+\alpha} \int_0^{\infty} \frac{1}{(\lambda + x_i)^2} \frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi} \lambda^{\frac{1+\alpha}{2}} d\lambda.$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

Note that by taking the derivative both sides of

$$x^{\frac{1+\alpha}{2}} = \cos\left(\frac{1+\alpha}{4}\pi\right) + \frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi}\int_0^\infty \left(\frac{\lambda}{\lambda^2+1} - \frac{1}{\lambda+x}\right)\lambda^{\frac{1+\alpha}{2}}d\lambda,$$

we get

$$\int_0^\infty \frac{1}{(\lambda+x_i)^2}\frac{\sin\left(\frac{1+\alpha}{2}\pi\right)}{\pi}\lambda^{\frac{1+\alpha}{2}}d\lambda = \frac{1+\alpha}{2}x_i^{\frac{\alpha-1}{2}}.$$

Hence, $y_{ii} = x_i^{\frac{1-\alpha}{2}}$ and $UZU^* = Y = diag\left(x_i^{\frac{1-\alpha}{2}}\right) = UX^{\frac{1-\alpha}{2}}U^*$. Therefore, $Z = X^{\frac{1-\alpha}{2}}$. Equivalently,

$$\sum_{j=1}^k w_j A_j^{\frac{1-\alpha}{2}} = X^{\frac{1-\alpha}{2}}.$$

Obviously, this equation has only one solution, that is,

$$X = \left(\sum_{j=1}^k w_j A_j^{\frac{1-\alpha}{2}}\right)^{\frac{2}{1-\alpha}}. \qquad \square$$

REFERENCES

[1] S. Amari. *Differential-Geometrical Methods in Statistics*. Lecture Notes in Statistics, Vol. 28, Springer-Verlag, New York, 1985.
[2] S. Amari. *Information Geometry and its Applications*. Applied Mathematical Sciences, Vol. 194, Springer, Tokyo, 2016.
[3] K.M.R. Audenaert. In-betweenness, a geometrical monotonicity property for operator means. *Linear Algebra Appl.*, 438(4):1769–1778, 2013.
[4] K.M.R. Audenaert and N. Datta. α-z-Rényi relative entropies. *J. Math. Phys.*, 56(2):022202, 2015.
[5] T. Ando. Matrix Young inequalities. *Oper. Theory Adv. Appl.*, Operator Theory in Function Spaces and Banach Lattices, 75:33–38, Birkhäuser, Basel, 1995.
[6] N.J. Beaudry and R. Renner. An intuitive proof of the data processing inequality. *Quantum Inf. Comput.*, 12(5/6):432–441, 2012.
[7] S. Beigi. Sandwiched Rényi divergence satisfies data processing inequality. *J. Math. Phys.*, 54(12):122202, 2013.
[8] M. Berta, V.B. Scholz, and M. Tomamichel. Rényi divergences as weighted non-commutative vector-valued L_p-spaces. *Ann. Henri Poincaré*, 19(6):1843–1867, 2018.
[9] R. Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics, Vol. 169, Springer-Verlag, New York, 1997.
[10] R. Bhatia. *Positive Definite Matrices*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, 2007.
[11] R. Bhatia, S. Gaubert, and T. Jain. Matrix versions of the Hellinger distance. *Lett. Math. Phys.*, 109(8):1777–1804, 2019.
[12] R. Bhatia, T. Jain and Y. Lim. Inequalities for the Wasserstein mean of positive definite matrices. *Linear Algebra Appl.*, 576:108–123, 2019.
[13] R. Bhatia, T. Jain and Y. Lim. On the Bures-Wasserstein distance between positive definite matrices. *Expo. Math.*, 37(2):165–191, 2019.
[14] E. Carlen. Trace inequalities and quantum entropy: An introductory course. *Contemp. Math.*, Entropy and the Quantum, Amer. Math. Soc., Providence, 529:73–140, 2010.
[15] E.A. Carlen, R.L. Frank, and E.H. Lieb. Inequalities for quantum divergences and the Audenaert-Datta conjecture. *J. Phys. A*, 51(48):483001, 2018.
[16] E.A. Carlen and A. Vershynina. Recovery and the data processing inequality for quasi-entropies. *IEEE Trans. Inform. Theory*, 64(10):6929–6938, 2018.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

[17] A. Cherian, S. Sra, A. Banerjee, and N. Papanikolopoulos. Efficient Similarity Search for Covariance Matrices via the Jensen-Bregman LogDet Divergence. *Int. Conf. Computer Vision (ICCV)*, 2399–2406, 2011.

[18] A. Cichocki, S. Cruces, and S. Amari. Log-determinant divergences revisited: alpha-beta and gamma log-det divergences. *Entropy*, 17(5):2988–3034, 2015.

[19] I.S. Dhillon and J.A. Tropp. Matrix nearness problems with Bregman divergences. *SIAM J. Matrix Anal. Appl.*, 29(4):1120–1146, 2007.

[20] T.H. Dinh, R. Dumitru, and J.A. Franco. On the matrix Heron means and Rényi divergences. *Linear Multilinear Algebra*, to appear.

[21] T.H. Dinh, R. Dumitru, and J.A. Franco. On the monotonicity of weighted power means for matrices. *Linear Algebra Appl.*, 527:128–140, 2017.

[22] T.H. Dinh, R. Dumitru, and J.A. Franco. Some geometric properties of matrix means in different distance functions. *Positivity*, to appear.

[23] T.H. Dinh, C.T. Le, B.-K. Vo, and T.-D Vuong. Weighted Hellinger distance and in-betweenness property. Preprint, 2020.

[24] T.H. Dinh, T.-Y. Tam, and B.-K. T. Vo. In-sphere property and reverse inequalities for matrix means. *Electron. J. Linear Algebra*, 35:35–41, 2019.

[25] R. Dumitru and J.A. Franco. Generalized Hellinger metric and Audenaert's in-betweenness. *Linear Algebra Appl.*, 585;191–198, 2020.

[26] R.L. Frank and E.H. Lieb. Monotonicity of a relative Rényi entropy. *J. Math. Phys.*, 54(12):122201, 2013.

[27] H. Hasegawa. α-divergence of the noncommutative information geometry. Proceedings of the XXV Symposium on Mathematical Physics (Toruń, 1992), *Rep. Math. Phys.*, 33(1/2):87–93, 1993.

[28] F. Hiai, M. Mosonyi, D. Petz, and C. Bény. Quantum f-divergences and error correction. *Rev. Math. Phys.*, 23(7):691–747, 2011.

[29] F. Hiai and D. Petz. From quasi-entropy to various quantum information quantities. *Publ. Res. Inst. Math. Sci.*, 48(3):525–542, 2012.

[30] F. Hiai and D. Petz. *Introduction to Matrix Analysis and Applications.* Universitext, Springer - Hindustan Book Agency, New Delhi, 2014.

[31] F. Kittaneh and Y. Manasrah. Improved Young and Heinz inequalities for matrices. *J. Math. Anal. Appl.*, 361(1):262–269,2010.

[32] P.W. Lamberti, A.P. Majtey, M. Madrid, and M.E. Pereyra. Jensen-Shannon divergence: A multipurpose distance for statistical and quantum mechanics. *Nonequilibrium Statistical Mechanics and Nonlinear Physics*, AIP Conference Proceedings, Vol. 913, American Institute of Physics, Melville, NY, 32–37, 2007.

[33] F. Leditzky, C. Rouzé, and N. Datta. Data processing for the sandwiched Rényi divergence: A condition for equality. *Lett. Math. Phys.*, 107(1):61–80, 2017.

[34] H.Q. Minh. A unified formulation for the Bures-Wasserstein and log-Euclidean/log-Hilbert-Schmidt distances between positive definite operators. *Lecture Notes in Compututer Science*, Geometric Science of Information, Vol. 11712, Springer, Cham, 475–483, 2019.

[35] M. Mosonyi and T. Ogawa. Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies. *Comm. Math. Phys.*, 334(3):1617–1648, 2015.

[36] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel. On quantum Rényi entropies: a new generalization and some properties. *J. Math. Phys.*, 54(12):122203, 2013.

[37] J. Pitrik and D. Virosztek. A divergence center interpretation of general symmetric Kubo-Ando means, and related weighted multivariate operator means, 2020. Preprint, arXiv:2002.11678.

[38] J. Pitrik and D. Virosztek. Quantum Hellinger distances revisited. *Lett. Math. Phys.*, 110(8):2039-2052, 2020.

[39] W. Pusz and S.L. Woronowicz. Functional calculus for sesquilinear forms and the purification map. *Rep. Mathematical Phys.*, 8(2):159–170, 1975.

[40] I. Sason. On data-processing and majorization inequalities for f-divergences with applications. *Entropy*, 21(10):Article 1022, 2019.

[41] B. Schumacher and M.A. Nielsen. Quantum data processing and error correction. *Phys. Rev. A*, 54 (4):2629–2635, 1996.

[42] N. Sharma. Equality conditions for the quantum f-relative entropy and generalized data processing inequalities. *Quantum Inf. Process.*, 11(1):137–160, 2012.

[43] B. Simon. *Loewner's Theorem on Monotone Matrix Functions.* Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), Vol. 354, Springer, Cham, 2019.

[44] S. Sra. Metrics induced by Jensen-Shannon and related divergences on positive definite matrices. Preprint, arXiv:1911.02643, 2019.

[45] S. Sra. Positive definite matrices and the S-divergence. *Proc. Amer. Math. Soc.*, 144(7):2787–2797, 2016.

[46] D. Virosztek. The metric property of the quantum Jensen-Shannon divergence. Preprint, arXiv:1910.10447, 2019.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 704-722, November 2020.

[47] M.M. Wilde. Optimized quantum f-divergences and data processing. *J. Phys. A*, 51(37):374002, 2018.

[48] 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(2)593–622, 2014.

[49] H. Zhang. From Wigner-Yanase-Dyson conjecture to Carlen-Frank-Lieb conjecture. *Adv. Math.*, 365:107053, 2020.