A Matrix Generalization of the Hardy-Littlewood-Pólya Rearrangement Inequality and Its Applications

Man-Chung Yue[†]

June 15, 2020

Abstract

Rearrangement inequalities are important tools in mathematical analysis and have found applications to many different areas. This paper proves a new matrix rearrangement inequality:

$$\sum_{i=1}^{n} f\left(\sigma_i(A)\sigma_{n-i+1}(B)\right) \leq \sum_{i=1}^{n} f\left(\sigma_i(AB)\right) \leq \sum_{i=1}^{n} f\left(\sigma_i(A)\sigma_i(B)\right),$$

for any matrices $A, B \in \mathbb{R}^{n\times n}$ and differentiable function f such that $s \mapsto sf'(s)$ is monotonically increasing. An important tool in the proof of the above inequality is a first-order perturbation formula for a certain class of matrix functions, which could be of independent interests. Applications of the new matrix rearrangement inequality to the Schatten quasi-norms, the affine-invariant geometry and log-determinant divergences for positive definite matrices are also presented.

Keywords: Matrix Rearrangement Inequality, Matrix Perturbation, Schatten Norms, Affine-Invariant Metric, Log-Determinant Divergences

1 Introduction

The well-known Hardy-Littlewood-Pólya rearrangement inequality [9] states that for any vectors $u, v \in \mathbb{R}^n$,

$$\sum_{i=1}^{n} u_i^{\downarrow} v_i^{\uparrow} \leq \sum_{i=1}^{n} u_i v_i \leq \sum_{i=1}^{n} u_i^{\downarrow} v_i^{\downarrow}, \tag{1}$$

[†]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong. E–mail: manchung.yue@polyu.edu.hk

where $u^{\downarrow}$ and $v^{\downarrow}$ ($u^{\uparrow}$ and $v^{\uparrow}$) are the vectors with entries of u and v sorted in non-increasing (non-decreasing) order, respectively. For positive vectors, a generalization of inequality (1) is obtained in [14], see also [28, Example 3].

Theorem 1 (London [14, Theorem 2]). *Let $u \in \mathbb{R}^n_{++}$, $v \in \mathbb{R}^n_{+}$ and $f : \mathbb{R}_+ \to \mathbb{R}$ be any convex such that $f(s) \geq f(0)$ for any $s \geq 0$. Then,*

$$\sum_{i=1}^{n} f(u_i^{\downarrow} v_i^{\uparrow}) \leq \sum_{i=1}^{n} f(u_i v_i) \leq \sum_{i=1}^{n} f(u_i^{\downarrow} v_i^{\downarrow}).$$

On the other hand, there are also various generalizations of inequality (1) to the matrix case, with the entries of vectors replaced by the eigenvalues $\lambda_i(\cdot)$ or singular values $\sigma_i(\cdot)$ of matrices. One such example is the following result.

Theorem 2 (Carlen and Lieb [4, Theorems 3.1-3.2][1]). *Let $A, B \in \mathbb{R}^{n\times n}$ be positive semidefinite matrices and $q \geq 1$. Then, it holds that*

$$\sum_{i=1}^{n} \lambda_i^q(A)\lambda_{n-i+1}^q(B) \leq \mathrm{Tr}\left((B^{\frac{1}{2}}AB^{\frac{1}{2}})^q\right).$$

If $q \geq 1$ is an integer, it also holds that

$$\mathrm{Tr}\left((B^{\frac{1}{2}}AB^{\frac{1}{2}})^q\right) \leq \sum_{i=1}^{n} \lambda_i^q(A)\lambda_i^q(B).$$

The Hardy-Littlewood-Pólya rearrangement inequality (1) and its generalizations, including Theorem 1 and Theorem 2, are useful tools in mathematical analysis and have found many applications in both pure and applied mathematics. For example, they have been applied to the study the geometry of Banach spaces [26, 4], quantum entanglement [2] and capacity of wireless communication channels [11, 6], to name a few.

The main result Theorem 5 of this paper is a new matrix rearrangement inequality, which generalizes both Theorem 1 (up to smoothness requirement) and Theorem 2. For the sake of easy discussion and comparisons, here we state a simplified version of Theorem 5.

[1]In [4, Theorems 3.1-3.2], the more general inequalities

$$\sum_{i=1}^{n} \lambda_i^q(A)\lambda_{n-i+1}^{p+q}(B) \leq \mathrm{Tr}\left(B^p(B^{\frac{1}{2}}AB^{\frac{1}{2}})^q\right) \quad \text{and} \quad \mathrm{Tr}\left(B^p(B^{\frac{1}{2}}AB^{\frac{1}{2}})^q\right) \leq \sum_{i=1}^{n} \lambda_i^q(A)\lambda_i^{p+q}(B)$$

for $p \geq 0$ are obtained under the same conditions on A, B and q. However, as shown in its proof, the first inequality can be reduced to the case of $p = 0$.

Theorem 3. *Let $f : \mathbb{R}_+ \to \mathbb{R}$ be a continuous function that is differentiable on $\mathbb{R}_{++}$. Suppose that the function $s \mapsto sf'(s)$ is monotonically increasing on $\mathbb{R}_{++}$. Then, for any matrices $A, B \in \mathbb{R}^{n\times n}$, it holds that*

$$\sum_{i=1}^{n} f\left(\sigma_i(A)\sigma_{n-i+1}(B)\right) \leq \sum_{i=1}^{n} f\left(\sigma_i(AB)\right) \leq \sum_{i=1}^{n} f\left(\sigma_i(A)\sigma_i(B)\right).$$

Theorem 3 does not only generalize Theorem 1 from the vector case to the matrix case but also relaxes the condition on the function f (up to smoothness requirements). Indeed, suppose that the function f in Theorem 1 is differentiable on $\mathbb{R}_{++}$. Then, we have that for any $s > 0$,

$$f(s) \geq f(0) \geq f(s) - sf'(s),$$

where the first inequality follows from the assumption of Theorem 1 and the second inequality from the convexity of f. Therefore, $f'(s) \geq 0$ for any $s > 0$. Hence, for any $s_2 > s_1 > 0$,

$$s_2 f'(s_2) - s_1 f'(s_1) = (s_2 - s_1) f'(s_2) + s_1(f'(s_2) - f'(s_1)) \geq 0, \tag{2}$$

where we used the fact that $f'(s_2) \geq f'(s_1)$, due to the convexity of f. This implies that the function $s \mapsto sf'(s)$ is monotonically increasing on $\mathbb{R}_{++}$. Furthermore, as shown in Theorem 6 (see also the remark after the proof of Theorem 6), Theorem 3 is generalization of Theorem 2 as well.

One possibility for proving rearrangement inequalities akin to Theorem 3 is by using the majorization theory and Schur-convexity [15]. To put things into perspective, we recall the weak majorization partial order: a vector $u \in \mathbb{R}^n$ is said to be weakly majorized by another vector $v \in \mathbb{R}^n$, denoted by $u \prec_w v$, if

$$\sum_{i=1}^{k} u_i \leq \sum_{i=1}^{k} v_i \quad \forall k = 1, \ldots, n.$$

It is well-known (see, *e.g.*, [15, Chapter 9, H]) that singular values satisfy the weak majorization inequality

$$\sigma^{\downarrow}(A) \circ \sigma^{\uparrow}(B) \prec_w \sigma(AB) \prec_w \sigma^{\downarrow}(A) \circ \sigma^{\downarrow}(B),$$

where $\circ$ denotes the Hadamard (entry-wise) product. Therefore, for any function $S : \mathbb{R}^n_+ \to \mathbb{R}$ that is order-preserving with respect to the partial order $\prec_w$, we have

$$S(\sigma^{\downarrow}(A) \circ \sigma^{\uparrow}(B)) \leq S(\sigma(AB)) \leq S(\sigma^{\downarrow}(A) \circ \sigma^{\downarrow}(B)).$$

However, the results [15, Chapter 3, A.8., C.1.b. and C.1.d.] show that the function considered in this paper, *i.e.*,

$$S(u) = \sum_{i=1}^{n} f\left(u_i\right),$$

is order-preserving with respect to $\prec_w$ if and only if f is convex and increasing, which implies the condition on f in Theorem 3 by inequality (2). By contrast, functions f satisfying the condition in Theorem 3 are not necessarily convex nor increasing. This shows that our approach yields more general rearrangement inequalities for singular values than those by using the majorization theory and Schur-convexity.

The rest of the paper is organized as follows. Section 2 prepares some auxiliary results. A key technical result Theorem 4 in this section is a first-order perturbation formula for a certain class of matrix functions, which could be of independent interests. The main result Theorem 5 and its extension to rectangular matrices will be proved in Section 3. In Section 4, we will present several applications of the main result Theorem 5. These applications are related to the Schatten quasi-norm (see Section 4.1), the affine-invariant Riemannian geometry (see Section 4.2) and the Alpha-Beta log-determinant divergences (see Section 4.3) for positive definite matrices.

1.1 Notations

The sets of non-negative and positive real numbers are denoted by $\mathbb{R}_+$ and $\mathbb{R}_{++}$, respectively. Let $u \in \mathbb{R}^n$. The $n \times n$ diagonal matrix with the i-th diagonal entry given by u_i is denoted by $\mathrm{Diag}(u)$. Also, the notations $u^\downarrow$ and $u^\uparrow$ denote the vectors with entries of u sorted in non-increasing and non-decreasing orders, respectively. The sets of $n \times n$ symmetric matrices, positive definite matrices and orthogonal matrices are denoted by $\mathbb{S}_n$, $\mathbb{P}_n$ and $\mathbb{O}_n$, respectively. Occasionally, we also write $A \succ 0$ to denote that A is positive definite. Throughout the paper, we always assume that m and n are positive integers and $m \leq n$. For any matrix $X \in \mathbb{R}^{m\times n}$, we denote by $\sigma(X) = (\sigma_1(X), \ldots, \sigma_m(X))^\top$ the vector of singular values sorted in non-increasing order. Also, for any $X \in \mathbb{S}_n$, we denote by $\lambda(X) = (\lambda_1(X), \ldots, \lambda_n(X))^\top$ the vector of eigenvalues sorted in non-increasing order.

2 Auxiliary Results

In this section, we prepare a few technical tools for the proofs of our main result. For simplicity, given any real-valued function f, we let $S_f : \mathbb{R}^{m\times n} \to \mathbb{R}$ be the function defined by

$$S_f(X) = \sum_{i=1}^{m} f\left(\sigma_i(X)\right), \quad X \in \mathbb{R}^{m\times n}.$$

The following theorem provides a first-order perturbation formula for S_f, which could also be of independent interests. In the sequel, we denote by ∂S_f the limiting subdifferential (or simply called the subdifferential) of S_f. For the definition of the limiting subdifferential, see, *e.g.*, [20, Definition 8.3(b)].

Theorem 4. *Let $f : \mathbb{R}_{++} \to \mathbb{R}$ be a differentiable function and $X \in \mathbb{R}^{m\times n}$ be a full-rank matrix, i.e.,* $\mathrm{Rank}(X) = m$. *Then, for any $\Delta \in \partial S_f(X)$, $Y \in \mathbb{R}^{m\times n}$ and sufficiently small $\epsilon > 0$, we have*

$$S_f(X + \epsilon Y) = S_f(X) + \epsilon\langle \Delta, Y\rangle + o(\epsilon).$$

A similar result is obtained in [29, Theorem 2]. Unfortunately, the proof of [29, Theorem 2] is flawed since the full-rank assumption on X was not imposed, which is necessary as explained after the proof. That said, the proof here is different from that in [29] and makes explicit connection to the limiting subdifferential ∂S_f of S_f, a feature absent from the proof of [29, Theorem 2].

To prove Theorem 4, we recall a few well-known results in matrix theory. The first one is a characterization of the limiting subdifferential of S_f.

Proposition 1 (Lewis and Sendov [12, Theorem 7.1])**.** *Let $f : \mathbb{R}_{++} \to \mathbb{R}$ be a differentiable function. For any $X \in \mathbb{R}^{m\times n}$, the limiting subdifferential $\partial S_f(X)$ of S_f at X is given by*

$$\partial S_f(X) = \left\{ U \left(\mathrm{Diag}\left(f'\left(\sigma(X)\right)\right) \;\; 0\right) V^\top : (U, V) \in \mathbb{O}^X \right\},$$

where

$$\mathbb{O}^X := \left\{ (U, V) \in \mathbb{O}_m \times \mathbb{O}_n : U \left(\mathrm{Diag}(\sigma(X)) \;\; 0\right) V^\top = X \right\},$$

and $f'\left(\sigma(X)\right)$ is the vector obtained by applying f' entry-wise to the vector $\sigma(X)$.

The original result [12, Theorem 7.1] actually holds for non-smooth functions defined on $\mathbb{R}_+$. But the version stated in Proposition 1 suffices for our purpose.

Let $\Xi : \mathbb{R}^{m\times n} \to \mathbb{S}_{m+n}$ be the linear map defined by

$$\Xi(Y) = \begin{pmatrix} 0 & Y \\ Y^\top & 0 \end{pmatrix}, \quad Y \in \mathbb{R}^{m\times n}.$$

The following proposition shows that the singular values and singular vectors of Y are intimately related to the eigenvalues and eigenvectors of $\Xi(Y)$, respectively.

Proposition 2 (Stewart [25, Chapter I, Theorem 4.2])**.** *Let $X \in \mathbb{R}^{m\times n}$ be a given matrix with singular value decomposition $X = U\left(\Sigma \;\; 0\right)\left(V^1 \;\; V^2\right)^\top$, where $\Sigma = \mathrm{Diag}(\sigma(X))$, $V^1 \in \mathbb{R}^{n\times m}$ and $V^2 \in \mathbb{R}^{n\times(n-m)}$. The matrix $\Xi(X)$ admits the eigenvalue decomposition*

$$\Xi(X) = W \begin{pmatrix} \Sigma & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\Sigma \end{pmatrix} W^\top, \tag{3}$$

where

$$W = \frac{1}{\sqrt{2}} \begin{pmatrix} U & 0 & U \\ V^1 & \sqrt{2}V^2 & -V^1 \end{pmatrix} \in \mathbb{O}_{m+n}.$$

In particular, 0 is an eigenvalue of $\Xi(X)$ of multiplicity at least $n - m$, and the remaining $2m$ eigenvalues of $\Xi(X)$ are $\pm\sigma_1(X), \ldots, \pm\sigma_m(X)$.

Finally, we also need the following perturbation formula for singular values.

Proposition 3 (Lewis and Sendov [13, Section 5.1]). *Let $X, Y \in \mathbb{R}^{m\times n}$. Suppose that X has $\ell+1$ distinct singular values for some $\ell \in \{0, 1 \ldots, m-1\}$ and that they are arranged as follows:*

$$\begin{array}{rcccc} & \sigma_{i_0}(X) & = \cdots = & \sigma_{i_1-1}(X) \\ > & \sigma_{i_1}(X) & = \cdots = & \sigma_{i_2-1}(X) \\ & & \vdots & \\ > & \sigma_{i_j}(X) & = \cdots = & \sigma_{i_{j+1}-1}(X) \\ & & \vdots & \\ > & \sigma_{i_\ell}(X) & = \cdots = & \sigma_{i_{\ell+1}-1}(X), \end{array} \tag{4}$$

where $1 = i_0 < i_1 < \cdots < i_\ell < i_{\ell+1} = m+1$. Then, for any $\epsilon > 0$, $j \in \{0, 1, \ldots, \ell\}$, and $i \in \{i_j, \ldots, i_{j+1}-1\}$, we have

$$\sigma_i(X + \epsilon Y) = \sigma_i(X) + \epsilon \cdot \lambda_{i-i_j+1}\left((W^j)^\top \Xi(Y) W^j\right) + O(\epsilon^2),$$

where W^j is an $(m+n) \times m_j$ matrix whose columns are the eigenvectors associated with the eigenvalue $\sigma_{i_j}(X)$ of $\Xi(X)$ (see Proposition 2) and m_j is the multiplicity of the eigenvalue $\sigma_{i_j}(X)$.

We now have enough tools at our disposal to prove Theorem 4.

Proof of Theorem 4. Recall the notations from Propositions 1-3. Let $\Delta \in \partial S_f(X)$. By Proposition 1, there exist a matrix $V^1 \in \mathbb{R}^{n\times m}$ with orthonormal columns and an orthogonal matrix $U \in \mathbb{O}_m$ such that

$$U \operatorname{Diag}(\sigma(X)) {V^1}^\top = X \quad \text{and} \quad \Delta = U \operatorname{Diag}\left(f'(\sigma(X))\right) {V^1}^\top.$$

Next, for any $j = 0, \ldots, \ell$, let $U^j \in \mathbb{R}^{m\times(i_{j+1}-i_j)}$, $V^{1,j} \in \mathbb{R}^{n\times(i_{j+1}-i_j)}$ be matrices whose columns are the left and right singular vectors of X associated the singular value $\sigma_{i_j}(X)$. By definitions,

$$U = \left(U^0 \;\, U^1 \cdots \;\, U^\ell\right) \quad \text{and} \quad V^1 = \left(V^{1,0} \;\, V^{1,1} \cdots \;\, V^{1,\ell}\right).$$

Then, Proposition 2 and the definition of W^j imply that

$$W^j = \frac{1}{\sqrt{2}} \begin{pmatrix} U^j \\ V^{1,j} \end{pmatrix}. \tag{5}$$

Since f is differentiable, it follows from Proposition 3 and the Taylor's Theorem that for sufficiently small $\epsilon > 0$, $j \in \{0, 1, \ldots, \ell\}$, and $i \in \{i_j, \ldots, i_{j+1}-1\}$, we have

$$f(\sigma_i(X + \epsilon Y)) = f(\sigma_i(X)) + \epsilon \cdot f'(\sigma_i(X)) \cdot \lambda_{i-i_j+1}\left((W^j)^\top \Xi(Y) W^j\right) + o(\epsilon). \tag{6}$$

Summing (6) over $j \in \{0, 1, \ldots, \ell\}$ and $i \in \{i_j, \ldots, i_{j+1} - 1\}$,

$$
\begin{aligned}
&\sum_{i=1}^{m} f(\sigma_i(X + \epsilon Y)) \\
&= \sum_{j=0}^{\ell} \sum_{i=i_j}^{i_{j+1}-1} f(\sigma_i(X + \epsilon Y)) \\
&= \sum_{j=0}^{\ell} \sum_{i=i_j}^{i_{j+1}-1} f(\sigma_i(X)) + \epsilon \sum_{j=0}^{\ell} \sum_{i=i_j}^{i_{j+1}-1} f'(\sigma_i(X)) \lambda_{i-i_j+1} \left((W^j)^\top \Xi(Y) W^j \right) + o(\epsilon) \\
&= \sum_{i=1}^{m} f(\sigma_i(X)) + \epsilon \sum_{j=0}^{\ell} f'(\sigma_{i_j}(X)) \sum_{i=i_j}^{i_{j+1}-1} \lambda_{i-i_j+1} \left((W^j)^\top \Xi(Y) W^j \right) + o(\epsilon) \\
&= \sum_{i=1}^{m} f(\sigma_i(X)) + \epsilon \sum_{j=0}^{\ell} f'(\sigma_{i_j}(X)) \operatorname{Tr} \left((W^j)^\top \Xi(Y) W^j \right) + o(\epsilon),
\end{aligned}
\tag{7}
$$

where the second summation in the last line can be computed using (5) as

$$
\begin{aligned}
&\sum_{j=0}^{\ell} f'(\sigma_{i_j}(X)) \cdot \operatorname{Tr} \left((W^j)^\top \Xi(Y) W^j \right) \\
&= \frac{1}{2} \left\langle \Xi(Y), \sum_{j=0}^{\ell} f' \left(\sigma_{i_j}(X) \right) \begin{pmatrix} U^j \\ V^{1,j} \end{pmatrix} \begin{pmatrix} U^j \\ V^{1,j} \end{pmatrix}^\top \right\rangle \\
&= \left\langle Y, \sum_{j=0}^{\ell} f' \left(\sigma_{i_j}(X) \right) U^j V^{1,j} \right\rangle \\
&= \left\langle Y, U \operatorname{Diag} \left(f'(\sigma_1(X)), \ldots, f'(\sigma_m(X)) \right) {V^1}^\top \right\rangle \\
&= \left\langle Y, U \operatorname{Diag} \left(f'(\sigma(X)) \right) {V^1}^\top \right\rangle.
\end{aligned}
\tag{8}
$$

Substituting (8) into (7) yields

$$
\begin{aligned}
\sum_{i=1}^{m} f(\sigma_i(X + \epsilon Y)) &= \sum_{i=1}^{m} f(\sigma_i(X)) + \epsilon \left\langle Y, U \operatorname{Diag} \left(f'(\sigma(X)) \right) {V^1}^\top \right\rangle + o(\epsilon) \\
&= \sum_{i=1}^{m} f(\sigma_i(X)) + \epsilon \langle Y, \Delta \rangle + o(\epsilon).
\end{aligned}
$$

This completes the proof. $\square$

The non-singularity of X is necessary to Theorem 4, even though Propositions 1-3 all do not require this assumption. Indeed, if the matrix X is of rank $r < m$, then there will be $m-r$ zero eigenvalues $\sigma_{i_\ell}(X) = \cdots = \sigma_{i_{\ell+1}-1}(X)$ of $\Xi(X)$ contained in Σ, $m-r$ zero eigenvalues $-\sigma_{i_\ell}(X) = \cdots = -\sigma_{i_{\ell+1}-1}(X)$ of $\Xi(X)$ contained in $-\Sigma$, and $n-m$ zero eigenvalues if X is not a square matrix (see equation (3)). Hence, the multiplicity m_ℓ of the zero eigenvalues of $\Xi(X)$ is $n+m-2r$. Instead of equation (5), the matrix W^ℓ is given by

$$W^\ell = \frac{1}{\sqrt{2}} \begin{pmatrix} U^\ell & 0 & U^\ell \\ V^{1,\ell} & \sqrt{2}V^2 & -V^{1,\ell} \end{pmatrix} \in \mathbb{R}^{(m+n)\times(n+m-2r)}.$$

This causes the breakdown of equation (8). Therefore, we need the assumption that X is full-rank in Theorem 4.

The following vector rearrangement inequality will also be used in the proof of our main result.

Lemma 1. *Let $f : \mathbb{R}_{++} \to \mathbb{R}$ be a differentiable function. The following hold.*

(i) If the function $s \mapsto sf'(s)$ is monotonically increasing on $\mathbb{R}_{++}$, then

$$\sum_{i=1}^n f(u_i^\uparrow v_i^\downarrow) \le \sum_{i=1}^n f(u_i v_i) \le \sum_{i=1}^n f(u_i^\downarrow v_i^\downarrow) \quad \forall u, v \in \mathbb{R}^n_{++}.$$

(ii) If the function $s \mapsto sf'(s)$ is monotonically decreasing on $\mathbb{R}_{++}$, then

$$\sum_{i=1}^n f(u_i^\uparrow v_i^\downarrow) \ge \sum_{i=1}^n f(u_i v_i) \ge \sum_{i=1}^n f(u_i^\downarrow v_i^\downarrow) \quad \forall u, v \in \mathbb{R}^n_{++}.$$

Proof. We only prove (i) as (ii) can be proved similarly. It suffices to prove that

$$f(ac) + f(bd) - f(ad) - f(bc) \ge 0 \quad \forall a \ge b > 0,\ c \ge d > 0.$$

Towards that end, we define the function $g(t) = f(tc) - f(td)$ for $t > 0$. By the assumptions on f,

$$g'(t) = cf'(tc) - df'(td) = \frac{1}{t}\left(tcf'(tc) - tdf'(td)\right) \ge 0 \quad \forall t > 0.$$

Therefore, for any $c \ge d > 0$, g is monotonically increasing and hence $g(a) \ge g(b)$, which is equivalent to the inequality $f(ac) + f(bd) - f(ad) - f(bc) \ge 0$. This completes the proof. $\square$

By a similar argument as in the introduction, we see that Lemma 1 is actually a generalization of of Theorem 1, up to smoothness requirement.

3 Main Result

We are now ready to prove the main result of this paper.

Theorem 5. *Let $f : \mathbb{R}_{++} \to \mathbb{R}$ be a differentiable function. Then, the following hold for any non-singular matrices $A, B \in \mathbb{R}^{n\times n}$.*

(i) If the function $s \mapsto sf'(s)$ is monotonically increasing on $\mathbb{R}_{++}$, then

$$\sum_{i=1}^{n} f\left(\sigma_i(A)\sigma_{n-i+1}(B)\right) \le S_f(AB) \le \sum_{i=1}^{n} f\left(\sigma_i(A)\sigma_i(B)\right).$$

(ii) If the function $s \mapsto sf'(s)$ is monotonically decreasing on $\mathbb{R}_{++}$, then

$$\sum_{i=1}^{n} f\left(\sigma_i(A)\sigma_{n-i+1}(B)\right) \ge S_f(AB) \ge \sum_{i=1}^{n} f\left(\sigma_i(A)\sigma_i(B)\right).$$

If, in addition to the differentiability on $\mathbb{R}_{++}$, f is also defined and continuous from the right at 0, then (i) and (ii) hold for any matrices $A, B \in \mathbb{R}^{n\times n}$.

Proof. For the assertions (i) and (ii), we only prove (i) as (ii) can be proved by using exactly the same arguments. To prove (i), let $A, B \in \mathbb{R}^{n\times n}$ be non-singular matrices. By using the right and left polar decompositions of A and B, respectively, we can assume without loss of generality that $A, B \succ 0$. Since the singular values $\sigma_i(\cdot)$ are continuous on $\mathbb{R}^{n\times n}$ (see, *e.g.*, [25, Chapter IV, Theorem 4.11]) and f is continuous on $\mathbb{R}_{++}$, we can also assume without loss of generality that the eigenvalues of A, B are all distinct. Furthermore, suppose that (i) holds for any function f such that $s \mapsto sf'(s)$ is strictly increasing on $\mathbb{R}_{++}$. Then, for any $\tilde{f}$ such that $s \mapsto s\tilde{f}'(s)$ is only monotonically increasing on $\mathbb{R}_{++}$, we consider the perturbed function $\tilde{f}_\epsilon(s) := \tilde{f}(s) + \epsilon s$. For $\epsilon > 0$,

$$s_2\tilde{f}'_\epsilon(s_2) - s_1\tilde{f}'_\epsilon(s_1) = (s_2 - s_1)\tilde{f}'_\epsilon(s_2) + s_1(\tilde{f}'_\epsilon(s_2) - \tilde{f}'_\epsilon(s_1)) > 0 \quad \forall s_2 > s_1 > 0.$$

Therefore, $s \mapsto s\tilde{f}'_\epsilon(s)$ is strictly increasing on $\mathbb{R}_{++}$ By supposition, the inequality in (i) holds for $f = \tilde{f}_\epsilon$. Taking limit $\epsilon \searrow 0$, the rearrangement inequality in (i) then holds for $\tilde{f}$ as well. Hence, it suffices to prove (i) for functions f such that $s \mapsto sf'(s)$ is strictly increasing on $\mathbb{R}_{++}$.

We start with the lower bound of $S_f(AB)$ in (i). Let $U_A\Sigma_A U_A^\top$ and $U_B\Sigma_B U_B^\top$ be the eigenvalue decompositions of A and B, respectively. Consider the minimization problem

$$\inf_{U\in\mathbb{O}_n} S_f(\Sigma_A U \Sigma_B U^\top). \tag{9}$$

By the continuity of f and the singular values $\sigma_i(\cdot)$, the function $U \mapsto S_f\left(\Sigma_A U \Sigma_B U^\top\right)$ is continuous on $\mathbb{O}_n$. Since $\mathbb{O}_n$ is compact, a minimizer $Q \in \mathbb{O}_n$ to problem (9) exists. Then, we have

$$S_f(AB) \geq \min_{U \in \mathbb{O}_n} S_f\left(\Sigma_A U \Sigma_B U^\top\right) = S_f\left(\Sigma_A Q \Sigma_B Q^\top\right).$$

Let

$$\hat{B} := Q \Sigma_B Q^\top \quad \text{and} \quad C := \Sigma_A^{\frac{1}{2}} \hat{B} \Sigma_A^{\frac{1}{2}}.$$

Since $A, B \succ 0$, we have that $C \succ 0$ and hence that $\lambda(C) = \sigma(C)$. Therefore, by Proposition 1, we have

$$\Delta := U_C \operatorname{Diag}\left(f'(\lambda(C))\right) U_C^\top \in \partial S_f(C),$$

where $U_C \operatorname{Diag}(\lambda(C)) U_C^\top$ is the eigenvalue decomposition of C. Let K be the commutator of $\hat{B}$ and C, *i.e.*,

$$K := \hat{B} \Sigma_A^{\frac{1}{2}} \Delta \Sigma_A^{\frac{1}{2}} - \Sigma_A^{\frac{1}{2}} \Delta \Sigma_A^{\frac{1}{2}} \hat{B},$$

which is a skew-symmetric matrix. It holds that $\operatorname{Exp}(\epsilon K) \in \mathbb{O}_n$ for any $\epsilon \in \mathbb{R}$, where $\operatorname{Exp}(\cdot)$ denotes the matrix exponential. By the continuity of f and the singular values $\sigma_i(\cdot)$ and Theorem 4, we have

$$\begin{aligned}
&S_f\left(\Sigma_A \operatorname{Exp}(\epsilon K) \hat{B} \operatorname{Exp}(\epsilon K)^\top\right) \\
=&S_f\left(\Sigma_A (I + \epsilon K) \hat{B} (I - \epsilon K)\right) + o(\epsilon) \\
=&S_f\left(\Sigma_A \hat{B} + \epsilon \Sigma_A (K \hat{B} - \hat{B} K)\right) + o(\epsilon) \\
=&S_f\left(\Sigma_A^{\frac{1}{2}} \hat{B} \Sigma_A^{\frac{1}{2}} + \epsilon \Sigma_A^{\frac{1}{2}} (K \hat{B} - \hat{B} K) \Sigma_A^{\frac{1}{2}}\right) + o(\epsilon) \\
=&S_f\left(\Sigma_A^{\frac{1}{2}} \hat{B} \Sigma_A^{\frac{1}{2}}\right) + \epsilon \left\langle \Sigma_A^{\frac{1}{2}} K \hat{B} \Sigma_A^{\frac{1}{2}} - \Sigma_A^{\frac{1}{2}} \hat{B} K \Sigma_A^{\frac{1}{2}}, \Delta \right\rangle + o(\epsilon) \\
=&S_f\left(\Sigma_A^{\frac{1}{2}} \hat{B} \Sigma_A^{\frac{1}{2}}\right) + \epsilon \left\langle K, \Sigma_A^{\frac{1}{2}} \Delta \Sigma_A^{\frac{1}{2}} \hat{B} - \hat{B} \Sigma_A^{\frac{1}{2}} \Delta \Sigma_A^{\frac{1}{2}} \right\rangle + o(\epsilon) \\
=&S_f\left(\Sigma_A^{\frac{1}{2}} \hat{B} \Sigma_A^{\frac{1}{2}}\right) - \epsilon \|K\|_F^2 + o(\epsilon),
\end{aligned}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Therefore, $K = 0$ because otherwise it would violate the minimality of Q by taking a sufficiently small $\epsilon > 0$. Hence, we have that

$$\hat{B} \Sigma_A^{\frac{1}{2}} \Delta \Sigma_A^{\frac{1}{2}} = \Sigma_A^{\frac{1}{2}} \Delta \Sigma_A^{\frac{1}{2}} \hat{B},$$

which is equivalent to

$$C \Delta \Sigma_A = \Sigma_A \Delta C. \tag{10}$$

Letting

$$\hat{C} := U_C \operatorname{Diag}\left(\lambda(C) \circ f'(\lambda(C))\right) U_C^\top, \tag{11}$$

it then follows from the definition of Δ that

$$C\Delta = U_C \operatorname{Diag}\left(\lambda(C) \circ f'(\lambda(C))\right) U_C^\top = \hat{C} = U_C \operatorname{Diag}\left(f'(\lambda(C) \circ \lambda(C))\right) U_C^\top = \Delta C.$$

Thus, equality (10) becomes

$$\hat{C}\Sigma_A = \Sigma_A \hat{C}. \tag{12}$$

In other words, $\hat{C}$ commutes with Σ_A. We claim that C also commutes with Σ_A. To prove the claim, we write the diagonal matrix $\operatorname{Diag}(\lambda(C))$ into the following block-diagonal form:

$$\operatorname{Diag}(\lambda(C)) = \begin{pmatrix} c_1 I_{n_1} & & \\ & \ddots & \\ & & c_\ell I_{n_\ell} \end{pmatrix}, \tag{13}$$

where $\ell, n_1, \ldots, n_\ell$ are positive integers such that $n_1 + \cdots + n_\ell = n$, and $c_1, \ldots, c_\ell$ are distinct real numbers. Since the function $s \mapsto sf'(s)$ is strictly increasing on $\mathbb{R}_{++}$, the diagonal matrix $\operatorname{Diag}\left(\lambda(C) \circ f'(\lambda(C))\right)$ takes the same form as (13), *i.e.*, there exist distinct numbers $t_1, \ldots, t_\ell \in \mathbb{R}$ such that

$$\operatorname{Diag}\left(\lambda(C) \circ f'(\lambda(C))\right) = \begin{pmatrix} t_1 I_{n_1} & & \\ & \ddots & \\ & & t_\ell I_{n_\ell} \end{pmatrix}. \tag{14}$$

From (11), (12) and (14), we have that

$$U_C^\top \Sigma_A U_C \begin{pmatrix} t_1 I_{n_1} & & \\ & \ddots & \\ & & t_\ell I_{n_\ell} \end{pmatrix} = \begin{pmatrix} t_1 I_{n_1} & & \\ & \ddots & \\ & & t_\ell I_{n_\ell} \end{pmatrix} U_C^\top \Sigma_A U_C,$$

which implies that the matrix $U_C^\top \Sigma_A U_C$ takes the same block-diagonal form as (13), *i.e.*, there exist symmetric matrices $\tilde{A}_1 \in \mathbb{R}^{n_1 \times n_1}, \ldots, \tilde{A}_\ell \in \mathbb{R}^{n_\ell \times n_\ell}$ such that

$$U_C^\top \Sigma_A U_C = \begin{pmatrix} \tilde{A}_1 & & \\ & \ddots & \\ & & \tilde{A}_\ell \end{pmatrix}.$$

Using the equality

$$\begin{pmatrix} \tilde{A}_1 & & \\ & \ddots & \\ & & \tilde{A}_\ell \end{pmatrix} \begin{pmatrix} c_1 I_{n_1} & & \\ & \ddots & \\ & & c_\ell I_{n_\ell} \end{pmatrix} = \begin{pmatrix} c_1 I_{n_1} & & \\ & \ddots & \\ & & c_\ell I_{n_\ell} \end{pmatrix} \begin{pmatrix} \tilde{A}_1 & & \\ & \ddots & \\ & & \tilde{A}_\ell \end{pmatrix},$$

we arrive at

$$U_C^\top \Sigma_A U_C \operatorname{Diag}(\lambda(C)) = \operatorname{Diag}(\lambda(C)) U_C^\top \Sigma_A U_C.$$

Multiplying the above equality by U_C from the left and $U_C^\top$ from the right, we get

$$\Sigma_A C = C\Sigma_A, \tag{15}$$

which proves the claim. Then, it follows from equality (15) and the definition of C that $\Sigma_A \hat{B} = \hat{B}\Sigma_A$. Since Σ_A is a diagonal matrix with distinct diagonal entries, the matrix $\hat{B} = Q\Sigma_B Q^\top$ is also diagonal. Hence, the minimizer Q is a permutation matrix. So, there exists a permutation π on $\{1, \ldots, n\}$ such that

$$S_f\left(\Sigma_A Q \Sigma_B Q^\top\right) = \sum_{i=1}^{n} f\left(\lambda_i(A)\lambda_{\pi_i}(B)\right). \tag{16}$$

Using (16), Lemma 1 and the fact that $A, B \succ 0$, we get

$$S_f(AB) \geq S_f\left(\Sigma_A Q \Sigma_B Q^\top\right) \geq \sum_{i=1}^{n} f(\lambda_i(A)\lambda_{n-i+1}(B)) = \sum_{i=1}^{n} f(\sigma_i(A)\sigma_{n-i+1}(B))$$

which yields the desired lower bound. The upper bound of $S_f(AB)$ in (i) can be proved similarly by considering the maximization problem

$$\sup_{U \in \mathbb{O}_n} S_f\left(\Sigma_A U \Sigma_B U^\top\right),$$

instead of the minimization problem (9). Finally, the last statement follows from the right continuity of f at 0, the continuity of singular values $\sigma_i(\cdot)$, limiting arguments and the statements (i) and (ii). This completes the proof. $\square$

For statement (i) (statement (ii)) in Theorem 5, the upper (lower) bound actually holds even for rectangular matrices.

Corollary 1. *Let $f : \mathbb{R}_+ \to \mathbb{R}$ be a function that is continuous from the right at 0 and differentiable on $\mathbb{R}_{++}$. Then, the following hold for any matrices $A, B \in \mathbb{R}^{m \times n}$.*

(i) If the function $s \mapsto sf'(s)$ is monotonically increasing on $\mathbb{R}_{++}$, then

$$S_f\left(AB^\top\right) \leq \sum_{i=1}^{m} f\left(\sigma_i(A)\sigma_i(B)\right).$$

(ii) If the function $s \mapsto sf'(s)$ is monotonically decreasing on $\mathbb{R}_{++}$, then

$$S_f\left(AB^\top\right) \geq \sum_{i=1}^{m} f\left(\sigma_i(A)\sigma_i(B)\right).$$

Proof. We only prove (i) as (ii) can be proved by using exactly the same arguments. Let $A, B \in \mathbb{R}^{m\times n}$. Consider the $(m+n)\times(m+n)$ symmetric matrices

$$\Xi(A) = \begin{pmatrix} 0 & A \\ A^\top & 0 \end{pmatrix} \quad \text{and} \quad \Xi(B) = \begin{pmatrix} 0 & B \\ B^\top & 0 \end{pmatrix}.$$

By Proposition 2, we have that for $i = 1, \ldots, m$,

$$\sigma_{2i-1}(\Xi(A)) = \sigma_{2i}(\Xi(A)) = \sigma_i(A) \quad \text{and} \quad \sigma_{2i-1}(\Xi(B)) = \sigma_{2i}(\Xi(B)) = \sigma_i(B),$$

and that for $i = 2m+1, \ldots, m+n$,

$$\sigma_i(\Xi(A)) = \sigma_i(\Xi(B)) = 0.$$

Therefore,

$$\sum_{i=1}^{m+n} f(\sigma_i(\Xi(A))\sigma_i(\Xi(B))) = 2\sum_{i=1}^{m} f\left(\sigma_i(A)\sigma_i(B)\right) + (n-m)f(0). \tag{17}$$

On the other hand, we have

$$\Xi(A)\Xi(B) = \begin{pmatrix} AB^\top & 0 \\ 0 & A^\top B \end{pmatrix},$$

which implies that for $i = 1, \ldots, m$,

$$\sigma_{2i-1}(\Xi(A)\Xi(B)) = \sigma_{2i}(\Xi(A)\Xi(B)) = \sigma_i\left(AB^\top\right),$$

and that for $i = 2m+1, \ldots, m+n$,

$$\sigma_i(\Xi(A)\Xi(B)) = 0.$$

Therefore,

$$\sum_{i=1}^{m+n} f\left(\sigma_i(\Xi(A)\Xi(B))\right) = 2\sum_{i=1}^{m} f(\sigma_i(AB^\top)) + (n-m)f(0). \tag{18}$$

Since the function $s \mapsto sf'(s)$ is monotonically increasing on $\mathbb{R}_{++}$, by the last statement of Theorem 5, we have that

$$\sum_{i=1}^{m+n} f\left(\sigma_i(\Xi(A)\Xi(B))\right) \le \sum_{i=1}^{m+n} f(\sigma_i(\Xi(A))\sigma_i(\Xi(B))). \tag{19}$$

Substituting equalities (17) and (18) into inequality (19), we obtain

$$\sum_{i=1}^{m} f(\sigma_i(AB^\top)) \le \sum_{i=1}^{m} f\left(\sigma_i(A)\sigma_i(B)\right).$$

The completes the proof. $\square$

4 Applications

4.1 Application to the Schatten-q Quasi-Norm

For $q \geq 1$, the Schatten-q norm is defined as

$$\|X\|_q = \sum_{i=1}^{m} \sigma_i^q(X), \quad X \in \mathbb{R}^{m \times n}.$$

Here we recall that we assume $m \leq n$. The Banach space associated with the Schatten-q norm for $q \geq 1$ is a classical subject in operator theory and has attracted much research since the forties, see [22, 8, 26]. On the other hand, if $q \in (0,1)$, $\|X\|_q$ is not longer a norm but only a quasi-norm since it violates the triangle inequality. Motivated by its proximity to the rank function, the Schatten-q quasi-norm with $q \in (0,1]$ has been extensively used in the studies of low-rank matrix recovery [21, 29] over the past decade over so.

As an application of our main result Theorem 5, we obtain the following inequality on the Schatten-q quasi-norm, which could potentially find applications in the analysis of the statistical properties and numerical algorithms of low-rank matrix recovery optimization models based on the Schatten-q quasi-norm.

Theorem 6. *Let $A, B \in \mathbb{R}^{m \times n}$ and $q \in \mathbb{R}$. Then, it holds that*

$$\|AB\|_q^q \leq \sum_{i=1}^{m} \sigma_i^q(A)\sigma_i^q(B).$$

Furthermore, if $m = n$, i.e., A and B are square matrices, it also holds that

$$\sum_{i=1}^{n} \sigma_i^q(A)\sigma_{n-i+1}^q(B) \leq \|AB\|_q^q.$$

Proof. We first prove the upper bound on $\|AB\|_q^q$. The case of $q = 0$ is trivial. So, we assume that $q \neq 0$. Consider the function $f : \mathbb{R}_+ \to \mathbb{R}$ defined by $f(s) = s^q$ for $s \in \mathbb{R}_+$. Then, the function $s \mapsto sf'(s) = qs^q$ is obviously increasing on $\mathbb{R}_{++}$. The desired inequality then follows from Corollary 1(i).

The lower bound on $\|AB\|_q^q$ can be proved similarly with using Theorem 5(i) instead of Corollary 1(i). This completes the proof. $\square$

Note that Theorem 6 holds for arbitrary $q \in \mathbb{R}$, including the $q < 0$. Furthermore, Theorem 6 generalizes Theorem 2 in two senses: from $q \geq 1$ to arbitrary $q \in \mathbb{R}$ and from positive semidefinite matrices to arbitrary square or rectangular matrices.

4.2 Application to the Affine-Invariant Geometry on $\mathbb{P}_n$

It is well-known that the cone of $n \times n$ positive definite matrices, denoted by $\mathbb{P}_n$, is a differentiable manifold of dimension $n(n+1)/2$, see, *e.g.*, [3, Chapter 6]. For any point $A \in \mathbb{P}_n$, the tangent space $T_A\mathbb{P}_n$ at A can be identified with the set of $n \times n$ symmetric matrices $\mathbb{S}_n$. We can equip the cone $\mathbb{P}_n$ with a Riemannian metric called the affine-invariant metric or the Fisher-Rao metric:

$$\langle X, Y \rangle_A := \operatorname{Tr}\left(XA^{-1}YA^{-1}\right), \quad X, Y \in T_A\mathbb{P}_n \cong \mathbb{S}_n.$$

Indeed, one can easily check that, given any $A \succ 0$, the map $\langle \cdot, \cdot \rangle_A$ defines a symmetric positive definite bilinear form on $\mathbb{S}_n$. The corresponding Riemannian (geodesic) distance $d_{\mathbb{P}_n}(\cdot, \cdot)$ is given by

$$d_{\mathbb{P}_n}(A, B) = \left\| \operatorname{Log}\left(B^{-\frac{1}{2}}AB^{-\frac{1}{2}}\right) \right\|_F, \quad A, B \in \mathbb{P}_n,$$

where $\operatorname{Log}(\,\cdot\,)$ denotes the matrix logarithm. This distance enjoys many interesting properties [3, Chapter 6] and finds applications in diverse areas such as machine learning [18, 24], image and video processing [7, 27] and elasticity theory [16, 17].

We obtain the following lower and upper bounds on $d_{\mathbb{P}_n}(\cdot, \cdot)$.

Corollary 2. *For any $A, B \in \mathbb{P}_n$, it holds that*

$$\sum_{i=1}^{n} (\log \lambda_i(A) - \log \lambda_i(B))^2 \le d^2_{\mathbb{P}_n}(A, B) \le \sum_{i=1}^{n} (\log \lambda_i(A) - \log \lambda_{n-i+1}(B))^2 .$$

Corollary 2 follows immediately by taking $q = 2$ in Theorem 7 below. To state Theorem 7, for any $q \ge 1$, we define

$$d_q(A, B) = \left\| \operatorname{Log}\left(B^{-\frac{1}{2}}AB^{-\frac{1}{2}}\right) \right\|_q, \quad A, B \in \mathbb{P}_n.$$

It has been proved in [3, Section 6] that the function d_q is a distance on $\mathbb{P}_n$ for general $q \ge 1$.

Theorem 7. *Let $A, B \in \mathbb{P}_n$ and $q \ge 1$. Then, it holds that*

$$\sum_{i=1}^{n} |\log \lambda_i(A) - \log \lambda_i(B)|^q \le d_q^q(A, B) \le \sum_{i=1}^{n} |\log \lambda_i(A) - \log \lambda_{n-i+1}(B)|^q .$$

Proof. We first assume that $q > 1$. Next, we note that

$$d_q^q(A, B) = \left\| \operatorname{Log}\left(B^{-\frac{1}{2}}AB^{-\frac{1}{2}}\right) \right\|_q^q = \sum_{i=1}^{n} f\left(\sigma_i(AB^{-1})\right),$$

where $f : \mathbb{R}_{++} \to \mathbb{R}$ is the function defined by $f(s) = |\log s|^q$ for $s \in \mathbb{R}_{++}$. It can be easily seen that f is differentiable on $\mathbb{R}_{++}$ and

$$f'(s) = \begin{cases} \frac{q}{s}(\log s)^{q-1}, & \text{if } s \geq 1, \\ -\frac{q}{s}(\log \frac{1}{s})^{q-1}, & \text{if } s < 1. \end{cases}$$

Hence,

$$sf'(s) = \mathrm{sgn}(\log s) \cdot q|\log s|^{q-1}. \tag{20}$$

To show that $s \mapsto sf'(s)$ is monotonically increasing, we let $s_2 > s_1 > 0$ and consider four different cases: $s_2 \geq 1 > s_1 > 0$, $s_2 > 1 \geq s_1 > 0$, $s_2 > s_1 \geq 1$ and $s_1 < s_2 \leq 1$. For the first two cases, by (20), we have that $s_2 f'(s_2) \geq 0 \geq s_1 f'(s_1)$. For the third case of $s_2 > s_1 \geq 1$, since f is twice differentiable on $(1, \infty)$ and

$$(sf'(s))' = \left(q(\log s)^{q-1}\right)' = \frac{q(q-1)(\log s)^{q-2}}{s} \geq 0 \quad \forall s > 1,$$

we have $s_2 f'(s_2) \geq s_1 f'(s_1)$. Similarly, for the fourth case of $s_1 < s_2 \leq 1$, since f is also twice differentiable on $(0, 1)$ and

$$(sf'(s))' = \left(-q\left(\log \tfrac{1}{s}\right)^{q-1}\right)' = \frac{q(q-1)\left(\log \frac{1}{s}\right)^{q-2}}{s} \geq 0 \quad \forall s \in (0, 1),$$

we have $s_2 f'(s_2) \geq s_1 f'(s_1)$. Therefore, by Theorem 5(i), we obtain

$$\sum_{i=1}^{n} f\left(\sigma_i(A)\sigma_i(B^{-1})\right) \geq \sum_{i=1}^{n} f\left(\sigma_i\left(AB^{-1}\right)\right) \geq \sum_{i=1}^{n} f\left(\sigma_i(A)\sigma_{n-i+1}(B^{-1})\right),$$

which is equivalent to

$$\sum_{i=1}^{n} |\log \lambda_i(A) - \log \lambda_i(B)|^q \leq d_q^q(A, B) \leq \sum_{i=1}^{n} |\log \lambda_i(A) - \log \lambda_{n-i+1}(B)|^q.$$

The case of $q = 1$ follows from limiting arguments. This completes the proof. □

4.3 Application to the Alpha-Beta Log-Determinant Divergences

Divergences, which are measures of dissimilarity of positive definite matrices, play an important role in information geometry and find applications across many areas, see [19, 5, 1] and the references therein. As a unification and generalization of many existing divergences in the literature, the family of Alpha-Beta log-determinant divergences (or AB log-det divergences for short) is introduced and studied in [5].

Given any $\alpha, \beta \in \mathbb{R}$ such that $\alpha\beta \neq 0$ and $\alpha + \beta \neq 0$, the AB log-det divergence $D_{\alpha,\beta}(\cdot\|\cdot)$ with parameter α and β is defined by

$$D_{\alpha,\beta}(A\|B) = \frac{1}{\alpha\beta} \log\det\left(\frac{\alpha(BA^{-1})^{\beta} + \beta(AB^{-1})^{-\alpha}}{\alpha+\beta}\right), \quad A, B \in \mathbb{P}_n.$$

The definition of AB log-det divergences can be extended to the cases of $\alpha\beta = 0$ and/or $\alpha + \beta = 0$ by taking limits and using the L,Hôpital's rule. In particular, we have

$$D_{\alpha,\beta}(A\|B) = \begin{cases} \frac{1}{\alpha^2}\left(\mathrm{Tr}\left(\left(BA^{-1}\right)^{\alpha} - I\right) - \alpha \log\det\left(BA^{-1}\right)\right), & \text{if } \alpha \neq 0, \beta = 0, \\ \frac{1}{\beta^2}\left(\mathrm{Tr}\left(\left(BA^{-1}\right)^{\alpha} - I\right) - \beta \log\det\left(BA^{-1}\right)\right), & \text{if } \beta \neq 0, \alpha = 0, \\ \frac{1}{\alpha^2} \log\left(\frac{\det\left(BA^{-1}\right)^{\alpha}}{\det\left(I + \log\left(BA^{-1}\right)^{\alpha}\right)}\right), & \text{if } \alpha = -\beta \neq 0, \\ \frac{1}{2}\left\|\mathrm{Log}(BA^{-1})\right\|_F^2, & \text{if } \alpha = \beta = 0. \end{cases}$$

Many well-known divergences are special cases of the AB log-det divergences, including the squared affine-invariant Riemannian metric (up to re-scaling, see Section 4.2) where $\alpha = \beta = 0$, the S-divergence [23] where $\alpha = \beta = \frac{1}{2}$ and the Stein's loss [10] (also called the Burg divergence) where $\alpha = 0$ and $\beta = 1$. For more examples of AB log-det divergences, we refer the readers to [5, Section 3].

As pointed out in [5], for any $A, B \in \mathbb{P}_n$, the AB log-det divergence $D_{\alpha,\beta}(A\|B)$ can be expressed via the eigenvalues of the matrix AB^{-1}, which coincide with those of the matrix $B^{-\frac{1}{2}}AB^{-\frac{1}{2}}$ and hence are positive. More precisely, it holds that

$$D_{\alpha,\beta}(A\|B) = \begin{cases} \frac{1}{\alpha\beta}\sum_{i=1}^{n} \log\left(\frac{\alpha\lambda_i^{\beta}(AB^{-1}) + \beta\lambda_i^{-\alpha}(AB^{-1})}{\alpha+\beta}\right), & \text{if } \alpha\beta, \alpha+\beta \neq 0, \\ \frac{1}{\alpha^2}\sum_{i=1}^{n}\left(\lambda_i^{-\alpha}(AB^{-1}) - \log\lambda_i^{-\alpha}(AB^{-1}) - 1\right), & \text{if } \alpha \neq 0, \beta = 0, \\ \frac{1}{\beta^2}\sum_{i=1}^{n}\left(\lambda_i^{\beta}(AB^{-1}) - \log\lambda_i^{\beta}(AB^{-1}) - 1\right), & \text{if } \beta \neq 0, \alpha = 0, \\ \frac{1}{\alpha^2}\sum_{i=1}^{n} \log\left(\frac{\lambda_i^{\alpha}\left(AB^{-1}\right)}{1 + \log\lambda_i^{\alpha}\left(AB^{-1}\right)}\right), & \text{if } \alpha = -\beta \neq 0, \\ \frac{1}{2}\sum_{i=1}^{n}\left(\log\lambda_i(BA^{-1})\right)^2, & \text{if } \alpha = \beta = 0. \end{cases}$$

The following upper and lower bounds for the AB log-det divergences are generalizations of the [23, Corollary 3.8].

Theorem 8. *Let $A, B \in \mathbb{P}_n$ and $\alpha, \beta \in \mathbb{R}$. If $\alpha\beta \geq 0$, then*

$$D_{\alpha,\beta}(A\|B) \leq \begin{cases} \dfrac{1}{\alpha\beta}\displaystyle\sum_{i=1}^{n} \log\left(\dfrac{\alpha\lambda_i^{\beta}(A)\lambda_{n-i+1}^{-\beta}(B) + \beta\lambda_i^{-\alpha}(A)\lambda_{n-i+1}^{\alpha}(B)}{\alpha+\beta}\right), & \text{if } \alpha\beta, \alpha+\beta \neq 0, \\ \dfrac{1}{\alpha^2}\displaystyle\sum_{i=1}^{n}\left(\dfrac{\lambda_{n-i+1}^{\alpha}(B)}{\lambda_i^{\alpha}(A)} - \log\left(\dfrac{\lambda_{n-i+1}^{\alpha}(B)}{\lambda_i^{\alpha}(A)}\right) - 1\right), & \text{if } \alpha \neq 0, \beta = 0, \\ \dfrac{1}{\beta^2}\displaystyle\sum_{i=1}^{n}\left(\dfrac{\lambda_i^{\beta}(A)}{\lambda_{n-i+1}^{\beta}(B)} - \log\left(\dfrac{\lambda_i^{\beta}(A)}{\lambda_{n-i+1}^{\beta}(B)}\right) - 1\right), & \text{if } \beta \neq 0, \alpha = 0, \\ \dfrac{1}{\alpha^2}\displaystyle\sum_{i=1}^{n} \log\left(\dfrac{\lambda_i^{\alpha}(A)\lambda_{n-i+1}^{-\alpha}(B)}{1+\log\left(\lambda_i^{\alpha}(A)\lambda_{n-i+1}^{-\alpha}(B)\right)}\right), & \text{if } \alpha = -\beta \neq 0, \\ \dfrac{1}{2}\displaystyle\sum_{i=1}^{n}\left(\log\lambda_i(A) - \log\lambda_{n-i+1}(B)\right)^2, & \text{if } \alpha = \beta = 0; \end{cases}$$

and

$$D_{\alpha,\beta}(A\|B) \geq \begin{cases} \dfrac{1}{\alpha\beta}\displaystyle\sum_{i=1}^{n} \log\left(\dfrac{\alpha\lambda_i^{\beta}(A)\lambda_i^{-\beta}(B) + \beta\lambda_i^{-\alpha}(A)\lambda_i^{\alpha}(B)}{\alpha+\beta}\right), & \text{if } \alpha\beta, \alpha+\beta \neq 0, \\ \dfrac{1}{\alpha^2}\displaystyle\sum_{i=1}^{n}\left(\dfrac{\lambda_i^{\alpha}(B)}{\lambda_i^{\alpha}(A)} - \log\left(\dfrac{\lambda_i^{\alpha}(B)}{\lambda_i^{\alpha}(A)}\right) - 1\right), & \text{if } \alpha \neq 0, \beta = 0, \\ \dfrac{1}{\beta^2}\displaystyle\sum_{i=1}^{n}\left(\dfrac{\lambda_i^{\beta}(A)}{\lambda_i^{\beta}(B)} - \log\left(\dfrac{\lambda_i^{\beta}(A)}{\lambda_i^{\beta}(B)}\right) - 1\right), & \text{if } \beta \neq 0, \alpha = 0, \\ \dfrac{1}{\alpha^2}\displaystyle\sum_{i=1}^{n} \log\left(\dfrac{\lambda_i^{\alpha}(A)\lambda_i^{-\alpha}(B)}{1+\log\left(\lambda_i^{\alpha}(A)\lambda_i^{-\alpha}(B)\right)}\right), & \text{if } \alpha = -\beta \neq 0, \\ \dfrac{1}{2}\displaystyle\sum_{i=1}^{n}\left(\log\lambda_i(A) - \log\lambda_i(B)\right)^2, & \text{if } \alpha = \beta = 0. \end{cases}$$

If $\alpha\beta < 0$, the inequalities for $D_{\alpha,\beta}(A\|B)$ in the case of $\alpha\beta, \alpha+\beta \neq 0$ are reversed.

Proof. We start with the case of $\alpha\beta, \alpha+\beta \neq 0$. Consider the function $f : \mathbb{R}_{++} \to \mathbb{R}$ defined by

$$f(s) = \log\left(\frac{\alpha s^{\beta} + \beta s^{-\alpha}}{\alpha+\beta}\right), \quad s > 0.$$

We have that for any $s > 0$,

$$sf'(s) = \frac{\alpha\beta(s^{\beta} - s^{-\alpha})}{\alpha s^{\beta} + \beta s^{-\alpha}}.$$

Then, for any $s > 0$,

$$\big(sf'(s)\big)' = \frac{\alpha\beta(\alpha+\beta)^2 s^{\beta-\alpha-1}}{(\alpha s^{\beta} + \beta s^{-\alpha})^2} = \begin{cases} > 0, & \text{if } \alpha\beta > 0, \\ < 0, & \text{if } \alpha\beta < 0. \end{cases}$$

By Theorem 5, if $\alpha\beta > 0$, then

$$\begin{aligned} &\frac{1}{\alpha\beta}\sum_{i=1}^{n} \log\left(\frac{\alpha\lambda_i^{\beta}(A)\lambda_i^{-\beta}(B) + \beta\lambda_i^{-\alpha}(A)\lambda_i^{\alpha}(B)}{\alpha+\beta}\right) \\ &\quad \le D_{\alpha,\beta}(A\|B) \le \frac{1}{\alpha\beta}\sum_{i=1}^{n} \log\left(\frac{\alpha\lambda_i^{\beta}(A)\lambda_{n-i+1}^{-\beta}(B) + \beta\lambda_i^{-\alpha}(A)\lambda_{n-i+1}^{\alpha}(B)}{\alpha+\beta}\right); \end{aligned}$$

and if $\alpha\beta < 0$, then

$$\begin{aligned} &\frac{1}{\alpha\beta}\sum_{i=1}^{n} \log\left(\frac{\alpha\lambda_i^{\beta}(A)\lambda_i^{-\beta}(B) + \beta\lambda_i^{-\alpha}(A)\lambda_i^{\alpha}(B)}{\alpha+\beta}\right) \\ &\quad \ge D_{\alpha,\beta}(A\|B) \ge \frac{1}{\alpha\beta}\sum_{i=1}^{n} \log\left(\frac{\alpha\lambda_i^{\beta}(A)\lambda_{n-i+1}^{-\beta}(B) + \beta\lambda_i^{-\alpha}(A)\lambda_{n-i+1}^{\alpha}(B)}{\alpha+\beta}\right). \end{aligned}$$

For the case of $\alpha \neq 0$ and $\beta = 0$, we consider the function $f : \mathbb{R}_{++} \to \mathbb{R}$ defined by

$$f(s) = s^{-\alpha} + \alpha \log s - 1, \quad s > 0.$$

We have that for any $s > 0$,

$$sf'(s) = \alpha(1 - s^{-\alpha}).$$

Then, for any $s > 0$,

$$\big(sf'(s)\big)' = \alpha^2 s^{-\alpha-1} > 0.$$

By Theorem 5,

$$\begin{aligned} &\frac{1}{\alpha^2}\sum_{i=1}^{n} \left(\frac{\lambda_i^{\alpha}(B)}{\lambda_i^{\alpha}(A)} - \log\left(\frac{\lambda_i^{\alpha}(B)}{\lambda_i^{\alpha}(A)}\right) - 1\right) \\ &\quad \le D_{\alpha,\beta}(A\|B) \le \frac{1}{\alpha^2}\sum_{i=1}^{n} \left(\frac{\lambda_{n-i+1}^{\alpha}(B)}{\lambda_i^{\alpha}(A)} - \log\left(\frac{\lambda_{n-i+1}^{\alpha}(B)}{\lambda_i^{\alpha}(A)}\right) - 1\right). \end{aligned}$$

For the case of $\beta \neq 0$ and $\alpha = 0$, by using exactly the same argument as that for the case of $\alpha \neq 0$ and $\beta = 0$, we can prove that

$$\frac{1}{\beta^2}\sum_{i=1}^{n}\left(\frac{\lambda_i^\beta(A)}{\lambda_i^\beta(B)} - \log\left(\frac{\lambda_i^\beta(A)}{\lambda_i^\beta(B)}\right) - 1\right)$$
$$\leq D_{\alpha,\beta}(A\|B) \leq \frac{1}{\beta^2}\sum_{i=1}^{n}\left(\frac{\lambda_i^\beta(A)}{\lambda_{n-i+1}^\beta(B)} - \log\left(\frac{\lambda_i^\beta(A)}{\lambda_{n-i+1}^\beta(B)}\right) - 1\right).$$

For the case of $\alpha = -\beta \neq 0$, we consider the function $f : \mathbb{R}_{++} \to \mathbb{R}$ defined by

$$f(s) = \log\left(\frac{s^\alpha}{1+\alpha\log s}\right), \quad s > 0.$$

We have that for any $s > 0$,

$$sf'(s) = \frac{\alpha^2 \log s}{1+\alpha\log s}.$$

Then, for any $s > 0$,

$$\left(sf'(s)\right)' = \frac{\alpha^2}{s(1+\alpha\log s)^2} > 0.$$

By Theorem 5,

$$\frac{1}{\alpha^2}\sum_{i=1}^{n}\log\left(\frac{\lambda_i^\alpha(A)\lambda_i^{-\alpha}(B)}{1+\log\left(\lambda_i^\alpha(A)\lambda_i^{-\alpha}(B)\right)}\right)$$
$$\leq D_{\alpha,\beta}(A\|B) \leq \frac{1}{\alpha^2}\sum_{i=1}^{n}\log\left(\frac{\lambda_i^\alpha(A)\lambda_{n-i+1}^{-\alpha}(B)}{1+\log\left(\lambda_i^\alpha(A)\lambda_{n-i+1}^{-\alpha}(B)\right)}\right).$$

For the case of $\alpha = \beta = 0$, it is the same inequality as in Corollary 2. This completes the proof. $\square$

Acknowledgments

The author is grateful to Cheuk Ting Li, Chi-Kwong Li, Wing-Kin Ma and Viet Anh Nguyen for their valuable comments on the manuscript.

References

[1] S.-I. Amari. *Information Geometry and Its Applications*, volume 194. Springer, 2016.

[2] R. Augusiak, J. Stasińska, and P. Horodecki. Beyond the standard entropic inequalities: Stronger scalar separability criteria and their applications. *Physical Review A*, 77(1):012333, 2008.

[3] R. Bhatia. *Positive Definite Matrices*, volume 24. Princeton University Press, 2009.

[4] E. Carlen and E. H. Lieb. Some matrix rearrangement inequalities. *Annali di Matematica Pura ed Applicata*, 185(5):S315–S324, 2006.

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

[6] M. Dörpinghaus, N. Gaffke, L. A. Imhof, and R. Mathar. A log-det inequality for random matrices. *SIAM Journal on Matrix Analysis and Applications*, 36(3):1164–1179, 2015.

[7] I. L. Dryden, A. Koloydenko, and D. Zhou. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics*, 3(3):1102–1123, 2009.

[8] I. Gohberg and M. G. Kreĭn. *Introduction to the theory of linear nonselfadjoint operators*. American Mathematical Society, 1969.

[9] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, 1952.

[10] W. James and C. Stein. Estimation with quadratic loss. In *Breakthroughs in Statistics*, pages 443–460. Springer, 1992.

[11] E. A. Jorswieck and H. Boche. Performance analysis of capacity of MIMO systems under multiuser interference based on worst-case noise behavior. *EURASIP Journal on Wireless Communications and Networking*, 2004(2):670321, 2004.

[12] A. S. Lewis and H. S. Sendov. Nonsmooth analysis of singular values. Part I: Theory. *Set-Valued Analysis*, 13(3):213–241, 2005.

[13] A. S. Lewis and H. S. Sendov. Nonsmooth analysis of singular values. Part II: Applications. *Set-Valued Analysis*, 13(3):243–264, 2005.

[14] D. London et al. Rearrangement inequalities involving convex functions. *Pacific Journal of Mathematics*, 34(3):749–753, 1970.

[15] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer Science & Business Media, 2010.

[16] M. Moakher and A. N. Norris. The closest elastic tensor of arbitrary symmetry to an elasticity tensor of lower symmetry. *Journal of Elasticity*, 85(3):215–263, 2006.

[17] P. Neff, Y. Nakatsukasa, and A. Fischle. A logarithmic minimization property of the unitary polar factor in the spectral and Frobenius norms. *SIAM Journal on Matrix Analysis and Applications*, 35(3):1132–1154, 2014.

[18] V. A. Nguyen, S. S. Abadeh, M.-C. Yue, D. Kuhn, and W. Wiesemann. Calculating optimistic likelihoods using (geodesically) convex optimization. In *Advances in Neural Information Processing Systems*, pages 13920–13931, 2019.

[19] F. Nielsen and R. Bhatia. *Matrix Information Geometry*. Springer, 2013.

[20] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317. Springer Science & Business Media, 2009.

[21] A. Rohde and A. B. Tsybakov. Estimation of high-dimensional low-rank matrices. *The Annals of Statistics*, 39(2):887–930, 2011.

[22] R. Schatten. *Norm ideals of completely continuous operators*. Springer-Verlag, 1960.

[23] S. Sra. Positive definite matrices and the S-divergence. *Proceedings of the American Mathematical Society*, 144(7):2787–2797, 2016.

[24] S. Sra and R. Hosseini. Geometric optimization in machine learning. In *Algorithmic Advances in Riemannian Geometry and Applications*, pages 73–91. Springer, 2016.

[25] G. W. Stewart. *Matrix Perturbation Theory*. Academic Press, Boston, 1990.

[26] N. Tomczak-Jaegermann. The moduli of smoothness and convexity and the Rademacher averages of the trace classes S_p $(1 \leq p < \infty)$. *Studia Mathematica*, 50(2):163–182, 1974.

[27] O. Tuzel, F. Porikli, and P. Meer. Pedestrian detection via classification on Riemannian manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1713–1727, 2008.

[28] A. Vince. A rearrangement inequality and the permutahedron. *The American Mathematical Monthly*, 97(4):319–323, 1990.

[29] M.-C. Yue and A. M.-C. So. A perturbation inequality for concave functions of singular values and its applications in low-rank matrix recovery. *Applied and Computational Harmonic Analysis*, 40(2):396–416, 2016.