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

Man-Chung Yue[†]

May 14, 2024

**Abstract**

We prove a generalization of the Hardy-Littlewood-Pólya rearrangement inequality to positive definite matrices. The inequality can be seen as a commutation principle in the sense of Iusem and Seeger. An important instrument in the proof is a first-order perturbation formula for a certain spectral function, which could be of independent interests. The inequality is then extended to rectangular matrices. Using our main results, we derive new inequalities for several distance-like functions encountered in various signal processing or machine learning applications.

**Keywords:** Matrix Rearrangement Inequality, Matrix Perturbation, Commutation Principle, Spectral Functions

## 1 Introduction

The well-known Hardy-Littlewood-Pólya rearrangement inequality [10] states that for any vectors $u, v \in \mathbb{R}^n$,

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

where $u^{\downarrow}$ and $v^{\downarrow}$ ($u^{\uparrow}$ and $v^{\uparrow}$) are the vectors with entries of $u$ and $v$ sorted in descending (ascending) order, respectively. For positive vectors, a generalization of the rearrangement inequality (1) is obtained in [16], see also [31, Example 3].

**Theorem 1** (London [16, Theorem 2])**.** *Let $u \in \mathbb{R}^n_{++}$, $v \in \mathbb{R}^n_{+}$ and $f : \mathbb{R}_+ \to \mathbb{R}$ be any convex function such that $f(s) \geq f(0)$ for any $s \geq 0$. Then,*

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

There are also various generalizations of inequality (1) to the matrix setting, where the entries of vectors are replaced by the eigenvalues or singular values of matrices. One such example is the following result.

[†]Musketeers Foundation Institute of Data Science and Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Hong Kong. E–mail: mcyue@hku.hk

**Theorem 2** (Carlen and Lieb [5, 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*

$$\sum_{i=1}^{n} \lambda_i^q(A)\lambda_{n-i+1}^q(B) \leq \sum_{i=1}^{n} \lambda_i^q(B^{\frac{1}{2}}AB^{\frac{1}{2}}),$$

*where $\lambda_i(\cdot)$ denotes the i-th largest eigenvalue. If $q \geq 1$ is an integer, it also holds that*

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

The Hardy-Littlewood-Pólya rearrangement inequality (1) and its generalizations are useful tools in mathematical analysis and have found many applications in both pure and applied mathematics. They have been utilized in the studies the geometry of Banach spaces [29, 5], quantum entanglement [2], covariance matrix estimation [24] and wireless communication [13, 7], to name a few.

Our main contribution is a pair of matrix rearrangement inequalities that generalize both Theorem 1 (up to differentiability requirement) and Theorem 2.

**Theorem 3.** *Let $f : \mathbb{R}_{++} \to \mathbb{R}$ be a differentiable function. Then, the following hold for any positive definite matrices $A, B \in \mathbb{R}^{n\times n}$.*

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

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

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

$$\sum_{i=1}^{n} f\left(\lambda_i(A)\lambda_{n-i+1}(B)\right) \geq \sum_{i=1}^{n} f\big(\lambda_i(B^{\frac{1}{2}}AB^{\frac{1}{2}})\big) \geq \sum_{i=1}^{n} f\left(\lambda_i(A)\lambda_i(B)\right).$$

*If $f$ is additionally defined and right-continuous at 0, then (i) and (ii) hold for any positive semidefinite matrices $A, B \in \mathbb{R}^{n\times n}$.*

The proof of Theorem 3 is based on the analysis of a certain optimization problem over orthogonal matrices and reveals that the matrices $A$ and $B$ commute at optimality. Therefore, Theorem 3 can be seen as a commutation principle for the function $X \mapsto \mathrm{Tr}(f(X))$ in the sense of [11, Lemma 4], see also [21, Theorem 7] for the generalization of [11, Lemma 4] to continuously differentiable matrix functions. Nevertheless, [21, Theorem 7] is not directly applicable to our situation as the function $X \mapsto \mathrm{Tr}(f(X))$ is not differentiable in general.

[1] In [5, Theorems 3.1-3.2], the more general inequalities

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

are obtained for $p \geq 0$ and under the same conditions on $A, B$ and $q$. However, as shown in its proof in [5], the first inequality can be reduced to the case of $p = 0$.

Moreover, our proof is different from that of [21, Theorem 7] and highlights the importance of the monotonicity of $s \mapsto sf'(s)$ and positive definiteness of $A$ and $B$.

Theorem 3 does not only generalize Theorem 1 from vector case to matrix case but also relaxes the condition on the function $f$ (up to differentiability requirement). To see this, consider a function $f$ satisfying the assumption of Theorem 1, which implies in particular that $f$ is defined and right-continuous at 0. Suppose in addition that $f$ 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 two inequality follows from the assumption of Theorem 1. 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 shows that the function $s \mapsto sf'(s)$ is monotonically increasing on $\mathbb{R}_{++}$. Furthermore, by taking $f(s) = s^q$ for $q > 0$, one readily sees that Theorem 3 recovers Theorem 2. Note also that the requirement on $q$ is less stringent than Theorem 2.

The rest of the paper is organized as follows. Section 2 prepares some auxiliary results. A key technical result in this section (Proposition 1) is a first-order perturbation formula for a certain class of matrix functions, which could be of independent interests. The main result Theorem 3 and its extension to rectangular matrices will be proved in Section 3. In Section 4, we will present several applications of our matrix rearrangement inequalities. These applications are related to Schatten quasi-norms (see Section 4.1), affine-invariant distance of positive definite matrices (see Section 4.2) and Alpha-Beta log-determinant divergences (see Section 4.3).

### 1.1 Notation

The sets of non-negative and positive real numbers are denoted by $\mathbb{R}_+$ and $\mathbb{R}_{++}$, respectively. For any vector $u \in \mathbb{R}^n$, the $n \times n$ diagonal matrix with the $i$-th diagonal entry given by $u_i$ is denoted by $\operatorname{Diag}(u)$. Also, we denote by $u^\downarrow$ and $u^\uparrow$ the vectors with entries of $u$ sorted in descending and ascending 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. For any matrix $X \in \mathbb{R}^{m \times n}$, we denote by $\sigma(X) = (\sigma_1(X), \ldots, \sigma_{\min\{m,n\}}(X))^\top$ the vector of singular values sorted in descending 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 descending order. Finally, the $n \times n$ identity matrix is denoted by $I_n$.

## 2 Auxiliary Results

In this section, we prepare a few technical tools for the proof of Theorem 3. From now on, for any function $f : J \to \mathbb{R}$ with domain $J \subseteq \mathbb{R}_+$ and matrix $X \in \mathbb{R}^{m \times n}$ with singular values in $J$, we define

$$S_f(X) = \sum_{i=1}^{\min\{m,n\}} f\left(\sigma_i(X)\right).$$

When $X \in \mathbb{R}^{n\times n}$ is a positive semidefinite matrix with eigenvalues in $J$, we have

$$S_f(X) = \sum_{i=1}^{n} f\left(\lambda_i(X)\right).$$

We next prove a first-order perturbation formula for $S_f$ based on its limiting subdifferential $\partial S_f$. For the definition of the limiting subdifferential, see, *e.g.*, [22, Definition 8.3(b)].

**Proposition 1.** *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.,* $\mathrm{Rank}(X) = \min\{m, n\}$. *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 [32, Theorem 2]. Unfortunately, the proof of [32, Theorem 2] is flawed since the full-rank assumption on $X$ was not imposed, which is necessary as explained after the proof of Proposition 1. The proof here is different from that in [32] and makes explicit connection to the limiting subdifferential $\partial S_f$, a feature absent from the proof of [32, Theorem 2].

To prove Proposition 1, we recall a few well-known results in matrix theory. The first one is a characterization of the limiting subdifferential of $S_f$.

**Proposition 2** (Lewis and Sendov [14, Theorem 7.1])**.** *Let $f : \mathbb{R}_{+} \to \mathbb{R}$ be a differentiable function. For any full-rank matrix $X \in \mathbb{R}^{m\times n}$ with $m \le n$, 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 [14, Theorem 7.1] holds more generally for non-smooth functions $f$ defined on $\mathbb{R}_+$. But the version stated in Proposition 2 suffices for our purpose. Note that if $X \in \mathbb{P}_n$, then $\partial S_f(X) \subseteq \mathbb{S}_n$.

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

$$\Xi(Y) = \begin{pmatrix} 0 & Y \\ Y^\top & 0 \end{pmatrix}.$$

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 3** (Stewart [28, Chapter I, Theorem 4.2])**.** *Let $X \in \mathbb{R}^{m\times n}$ with $m \le n$ and $X = U\left(\Sigma \;\; 0\right)\left(V^1 \;\; V^2\right)^\top$ be its singular value decomposition, 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* $\pm\sigma_1(X), \ldots, \pm\sigma_m(X)$.

We also need the following perturbation formula for singular values.

**Proposition 4** (Lewis and Sendov [15, Section 5.1])**.** *Let* $X, Y \in \mathbb{R}^{m\times n}$ *with* $m \le n$*. Suppose that* $X$ *has* $\ell + 1$ *distinct singular values for some* $\ell \in \{0, 1 \ldots, m-1\}$*, arranged as follows:*

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

*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 3) and* $m_j$ *is the multiplicity of the eigenvalue* $\sigma_{i_j}(X)$.

We now have enough tools to prove Proposition 1.

*Proof of Proposition 1.* Since singular values are invariant to matrix transposition, we can assume without loss of generality that $m \le n$. Let $\Delta \in \partial S_f(X)$. By Proposition 2, 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 the matrices whose columns are the left and right singular vectors of $X$ associated with the singular value $\sigma_{i_j}(X)$. By definitions,

$$U = \begin{pmatrix} U^0 & U^1 \cdots & U^\ell \end{pmatrix} \quad \text{and} \quad V^1 = \begin{pmatrix} V^{1,0} & V^{1,1} \cdots & V^{1,\ell} \end{pmatrix}.$$

Then, Proposition 3 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 on $\mathbb{R}_{++}$, it follows from the full-rank assumption on $X$, Proposition 4 and the Taylor's Theorem that for any 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 $i \in \{i_j, \ldots, i_{j+1} - 1\}$,

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

The second summation in the last line of (7) can be computed using (5) as follows:

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

Substituting (8) into (7) completes the proof. □

The full-rank assumption of $X$ is necessary to Proposition 1, even though Propositions 2-4 all do not require this assumption (see the remark following Proposition 2). Indeed, assuming $m \le n$, 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 $\Sigma$, $m - r$ 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 $m_\ell$ of the zero eigenvalues of $\Xi(X)$ is $n+m-2r$. Instead of equation (5), the matrix $W^\ell$ 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 Proposition 1.

The proof of Theorem 3 also relies on the following three elementary lemmas, whose proofs are included for self-containedness. The first one is a vector rearrangement inequality.

**Lemma 1.** *Let $f : \mathbb{R}_{++} \to \mathbb{R}$ be a differentiable function. Then, the following hold for any positive vectors $u, v \in \mathbb{R}^n$.*

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

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

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

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

*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. $\square$

The next one reveals the structure of matrices commuting with a diagonal matrix.

**Lemma 2.** *Let $t_1, \ldots, t_\ell \in \mathbb{R}$ be distinct real numbers, $n_1, \ldots, n_\ell$ be positive integers and $X \in \mathbb{S}_n$ with $n = n_1 + \cdots + n_\ell$. Suppose that*

$$X \begin{pmatrix} t_1 I_{n_1} & & \\ & \ddots & \\ & & t_\ell I_{n_\ell} \end{pmatrix} = \begin{pmatrix} t_1 I_{n_1} & & \\ & \ddots & \\ & & t_\ell I_{n_\ell} \end{pmatrix} X.$$

*Then, there exist symmetric matrices $\tilde{X}_1 \in \mathbb{S}_{n_1}, \ldots, \tilde{X}_\ell \in \mathbb{S}_{n_\ell}$ such that*

$$X = \begin{pmatrix} \tilde{X}_1 & & \\ & \ddots & \\ & & \tilde{X}_\ell \end{pmatrix}.$$

*Proof.* To prove the lemma, for any matrix $Y \in \mathbb{R}^{n\times n}$, we partition its entries into blocks so that for any $i, j = 1, \dots, \ell$, the $ij$-th block, denoted by $[Y]_{ij}$, is $n_i \times n_j$. Let $D \in \mathbb{R}^{n\times n}$ be the diagonal matrix such that $[D]_{ii} = t_i I_{n_i}$. Then, for any $i, j = 1, \dots, \ell$, we have

$$[XD]_{ij} = \sum_{k=1}^{\ell} [X]_{ik}[D]_{kj} = t_j [X]_{ij} \quad \text{and} \quad [DX]_{ij} = \sum_{k=1}^{\ell} [D]_{ik}[X]_{kj} = t_i [X]_{ij}.$$

The supposition then implies $t_i[X]_{ij} = t_j[X]_{ij}$. Since $t_1, \dots, t_\ell$ are distinct, we conclude that $[X]_{ij}$ is a zero matrix for $i \neq j$. This completes the proof. □

The last one concerns diagonal matrices related by orthogonal conjugation.

**Lemma 3.** *Let $D, \hat{D} \in \mathbb{R}^{n\times n}$ be two diagonal matrices and $Q \in \mathbb{O}_n$. Suppose that the diagonal entries of $D$ are distinct and $\hat{D} = QDQ^\top$. Then, $Q$ is a permutation matrix.*

*Proof.* By supposition, we have that $\hat{D}Q = QD$. Next, for any $i, j = 1, \dots, n$,

$$(\hat{D}Q)_{ij} = \sum_{k=1}^{n} \hat{D}_{ik} Q_{kj} = \hat{D}_{ii} Q_{ij} \quad \text{and} \quad (QD)_{ij} = \sum_{k=1}^{n} Q_{ik} D_{kj} = D_{jj} Q_{ij},$$

which implies that $(\hat{D}_{ii} - D_{jj})Q_{ij} = 0$. Since eigenvalues are preserved by conjugation, we know that the set of numbers on the diagonal of $\hat{D}$ must be the same as that of $D$. In other words, there exists a permutation $\pi$ on $\{1, \dots, n\}$ such that $\hat{D}_{ii} = D_{\pi(i)\pi(i)}$. Therefore, we have $(D_{\pi(i)\pi(i)} - D_{jj})Q_{ij} = 0$ for any $i, j = 1, \dots, n$. Since $D_{11}, \dots, D_{nn}$ are distinct, $Q_{ij} = 0$ unless $\pi(i) = j$, in which case $Q_{ij} = 1$ due to the orthogonality. This completes the proof. □

## 3 Main Results

We are now ready to prove Theorem 3.

*Proof of Theorem 3.* We only prove assertion (i) as assertion (ii) can be proved by using the same arguments. To prove (i), let $A, B \in \mathbb{P}_n$. Since eigenvalues $\lambda_i(\cdot)$ are continuous on $\mathbb{R}^{n\times n}$ (see, *e.g.*, [3, Corollary VI.1.6]) 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 $s_2 > s_1 > 0$ and $\epsilon > 0$,

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

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 $\mathbb{R}_{++}$.

Since $B^{\frac{1}{2}}AB^{\frac{1}{2}} \in \mathbb{P}_n$, by the definition of $S_f$, we have

$$S_f(B^{\frac{1}{2}}AB^{\frac{1}{2}}) = \sum_{i=1}^{n} f(\lambda_i(B^{\frac{1}{2}}AB^{\frac{1}{2}})).$$

We start with the lower bound 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

$$\inf_{U\in\mathbb{O}_n} S_f\big(\Sigma_B^{\frac{1}{2}}U\Sigma_A U^\top\Sigma_B^{\frac{1}{2}}\big). \tag{9}$$

By the continuity of $f$ and eigenvalues $\lambda_i(\cdot)$, the function $U\mapsto S_f(\Sigma_B^{\frac{1}{2}}U\Sigma_A U^\top\Sigma_B^{\frac{1}{2}})$ is continuous on $\mathbb{O}_n$. Since $\mathbb{O}_n$ is compact, a minimizer $Q\in\mathbb{O}_n$ to problem (9) exists. We thus have

$$S_f(B^{\frac{1}{2}}AB^{\frac{1}{2}})\ge\min_{U\in\mathbb{O}_n} S_f\big(\Sigma_B^{\frac{1}{2}}U\Sigma_A U^\top\Sigma_B^{\frac{1}{2}}\big)=S_f\big(\Sigma_B^{\frac{1}{2}}Q\Sigma_A Q^\top\Sigma_B^{\frac{1}{2}}\big).$$

Let

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

Since $A,B\in\mathbb{P}_n$, we have that $C\in\mathbb{P}_n$ and hence that $\lambda(C)=\sigma(C)$. Fix any eigenvalue decomposition $C=U_C\operatorname{Diag}(\lambda(C))U_C^\top$ and let $\Delta=U_C\operatorname{Diag}\left(f'(\lambda(C))\right)U_C^\top$. By Proposition 2, we have $\Delta\in\partial S_f(C)$. Consider the skew-symmetric matrix

$$K=\hat{A}\Sigma_B^{\frac{1}{2}}\Delta\Sigma_B^{\frac{1}{2}}-\Sigma_B^{\frac{1}{2}}\Delta\Sigma_B^{\frac{1}{2}}\hat{A}.$$

By the skew-symmetry of $K$, $\operatorname{Exp}(\epsilon K)\in\mathbb{O}_n$ for any $\epsilon\in\mathbb{R}$, where $\operatorname{Exp}(\cdot)$ denotes the matrix exponential. Recalling that $\operatorname{Exp}(X)=I_n+X+\frac{1}{2}X^2+\cdots$ for any matrix $X\in\mathbb{R}^{n\times n}$, we have

$$\begin{aligned}
&S_f\big(\Sigma_B^{\frac{1}{2}}\operatorname{Exp}(\epsilon K)\hat{A}\operatorname{Exp}(\epsilon K)^\top\Sigma_B^{\frac{1}{2}}\big)\\
=&S_f\big(\Sigma_B^{\frac{1}{2}}(I+\epsilon K)\hat{A}(I-\epsilon K)\Sigma_B^{\frac{1}{2}}\big)+o(\epsilon)\\
=&S_f\big(\Sigma_B^{\frac{1}{2}}\hat{A}\Sigma_B^{\frac{1}{2}}+\epsilon\Sigma_B^{\frac{1}{2}}(K\hat{A}-\hat{A}K)\Sigma_B^{\frac{1}{2}}\big)+o(\epsilon)\\
=&S_f\left(\Sigma_B^{\frac{1}{2}}\hat{A}\Sigma_B^{\frac{1}{2}}\right)+\epsilon\left\langle\Sigma_B^{\frac{1}{2}}K\hat{A}\Sigma_B^{\frac{1}{2}}-\Sigma_B^{\frac{1}{2}}\hat{A}K\Sigma_B^{\frac{1}{2}},\,\Delta\right\rangle+o(\epsilon)\\
=&S_f\left(\Sigma_B^{\frac{1}{2}}\hat{A}\Sigma_B^{\frac{1}{2}}\right)+\epsilon\left\langle K,\,\Sigma_B^{\frac{1}{2}}\Delta\Sigma_B^{\frac{1}{2}}\hat{A}-\hat{A}\Sigma_B^{\frac{1}{2}}\Delta\Sigma_B^{\frac{1}{2}}\right\rangle+o(\epsilon)\\
=&S_f\left(\Sigma_B^{\frac{1}{2}}\hat{A}\Sigma_B^{\frac{1}{2}}\right)-\epsilon\,\|K\|_{\mathrm{F}}^2+o(\epsilon),
\end{aligned}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm, the first and second equalities follow from the continuity of $f$ and eigenvalues $\lambda_i(\cdot)$, the third from Proposition 1, the fourth by the symmetry of $\Sigma_B^{\frac{1}{2}}$ and $\hat{A}$, and the fifth by the definition of $K$. Therefore, $K=0$ because otherwise the last display would violate the minimality of $Q$ by taking a sufficiently small $\epsilon>0$. Hence, we have that

$$\hat{A}\Sigma_B^{\frac{1}{2}}\Delta\Sigma_B^{\frac{1}{2}}=\Sigma_B^{\frac{1}{2}}\Delta\Sigma_B^{\frac{1}{2}}\hat{A},$$

which, upon multiplying both sides by $\Sigma_B^{\frac{1}{2}}$, yields

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

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 $\Delta$ 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) is equivalent to

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

In other words, $\Sigma_B$ commutes with $\hat{C}$. We claim that $\Sigma_B$ also commutes with $C$. To prove the claim, note that we can write

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

for some positive integers $\ell, n_1, \ldots, n_\ell$ with $n_1 + \cdots + n_\ell = n$ and real numbers $c_1 > \cdots > c_\ell > 0$. 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.*, for some real numbers $t_1 > \cdots > t_\ell > 0$,

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

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

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

which, upon invoking Lemma 2, implies the existence of symmetric matrices $\tilde{B}_1 \in \mathbb{R}^{n_1 \times n_1}, \ldots, \tilde{B}_\ell \in \mathbb{R}^{n_\ell \times n_\ell}$ such that

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

Observing that

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

we arrive at

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

Multiplying the last display by $U_C$ from the left and $U_C^\top$ from the right, we get

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

which proves the claim. Then, it follows from equality (15) and the definition of $C$ that $\Sigma_B \hat{A} = \hat{A}\Sigma_B$. Since $\Sigma_B$ is a diagonal matrix with distinct diagonal entries, the matrix

$\hat{A} = Q\Sigma_A Q^\top$ is also diagonal by Lemma 2. Next, using Lemma 3 and that $\Sigma_A$ is a diagonal matrix with distinct diagonal entries, the minimizer $Q$ is a permutation matrix. Hence, there exists a permutation $\pi$ on $\{1, \ldots, n\}$ such that

$$S_f(\Sigma_B^{\frac{1}{2}} Q\Sigma_A Q^\top \Sigma_B^{\frac{1}{2}}) = \sum_{i=1}^{n} f\left(\lambda_i(A)\lambda_{\pi(i)}(B)\right). \tag{16}$$

Using (16), Lemma 1 and the fact that $A, B \in \mathbb{P}_n$, we get

$$S_f(B^{\frac{1}{2}} A B^{\frac{1}{2}}) \geq S_f(\Sigma_B^{\frac{1}{2}} Q\Sigma_A Q^\top \Sigma_B^{\frac{1}{2}}) \geq \sum_{i=1}^{n} f(\lambda_i(A)\lambda_{n-i+1}(B)),$$

which yields the lower bound in (i). The upper bound of $S_f(B^{\frac{1}{2}} A B^{\frac{1}{2}})$ in (i) can be proved similarly by considering the maximization problem

$$\sup_{U \in \mathbb{O}_n} S_f(\Sigma_B^{\frac{1}{2}} U\Sigma_A U^\top \Sigma_B^{\frac{1}{2}}),$$

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

We next prove a matrix rearrangement inequality for singular values of rectangular matrices.

**Theorem 4.** *Let $f : \mathbb{R}_{++} \to \mathbb{R}$ be a differentiable function. Then, the following hold for any full-rank matrices $X, Y \in \mathbb{R}^{m \times n}$.*

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

$$\sum_{i=1}^{\min\{m,n\}} f\left(\sigma_i(X)\sigma_{n-i+1}(Y)\right) \leq \sum_{i=1}^{\min\{m,n\}} f\left(\sigma_i(X^\top Y)\right) \leq \sum_{i=1}^{\min\{m,n\}} f\left(\sigma_i(X)\sigma_i(Y)\right).$$

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

$$\sum_{i=1}^{\min\{m,n\}} f\left(\sigma_i(X)\sigma_{n-i+1}(Y)\right) \geq \sum_{i=1}^{\min\{m,n\}} f\left(\sigma_i(X^\top Y)\right) \geq \sum_{i=1}^{\min\{m,n\}} f\left(\sigma_i(X)\sigma_i(Y)\right).$$

*If $f$ is additionally defined and right-continuous at 0, then (i) and (ii) hold for any matrices $X, Y \in \mathbb{R}^{m \times n}$.*

*Proof.* Let $X, Y \in \mathbb{R}^{m \times n}$. Since $\sigma_i(X^\top Y) = \sigma_i(XY^\top)$ for $i = 1, \ldots, \min\{m, n\}$, we can assume without loss of generality that $m \leq n$. By the definition of singular values, we have that for any $i = 1, \ldots, m$,

$$\sigma_i(X^\top Y) = \lambda_i^{\frac{1}{2}}(X^\top Y Y^\top X) = \lambda_i^{\frac{1}{2}}(XX^\top YY^\top).$$

Similarly, we have that for any $i = 1, \ldots, m$,

$$\sigma_i(X) = \lambda_i^{\frac{1}{2}}(XX^\top) \quad \text{and} \quad \sigma_i(Y) = \lambda_i^{\frac{1}{2}}(YY^\top).$$

Also, $XX^\top$ and $YY^\top$ are positive definite if and only if $X$ and $Y$ have full rank. Next, let $g : \mathbb{R}_{++} \to \mathbb{R}$ be the function defined by $g(t) = f(\sqrt{t})$. Then, $g$ is differentiable on $\mathbb{R}_{++}$ and $tg'(t) = \frac{1}{2}\sqrt{t}f'(\sqrt{t})$, whose monotonicity inherits from the map $s \mapsto sf'(s)$. Moreover, if $f$ is defined and right-continuous at 0, so is $g$. Noting that $\lambda_i(B^{\frac{1}{2}}AB^{\frac{1}{2}}) = \lambda_i(AB)$ for any positive semidefinite matrices $A, B \in \mathbb{R}^{m\times m}$, applying Theorem 3 with $A = XX^\top$ and $B = YY^\top$ yields the desired conclusion. $\square$

# 4 Applications

## 4.1 Schatten Quasi-Norms

For $q > 0$ and $X \in \mathbb{R}^{m\times n}$, we denote

$$\|X\|_q = \left( \sum_{i=1}^{\min\{m,n\}} \sigma_i^q(X) \right)^{\frac{1}{q}}.$$

If $q \geq 1$, $\|X\|_q$ is the so-called Schatten-$q$ norm of $X$. The Banach space associated with the Schatten-$q$ norm is a classical subject in operator theory and has attracted much research since the forties, see [25, 9, 29]. On the other hand, if $q \in (0, 1)$, $\|X\|_q$ is no longer a norm but only a quasi-norm. Motivated by its proximity to the rank function, the Schatten-$q$ quasi-norm with $q \in (0, 1]$ has been applied to low-rank matrix recovery [23, 32].

As an application of our Theorem 4, we obtain the following inequality on $\|\cdot\|_q$ for general $q \in \mathbb{R}$, which could potentially find applications in the analysis of the statistical properties of and numerical algorithms for low-rank matrix recovery based on the Schatten-$q$ quasi-norm.

**Corollary 1.** *Let $X, Y \in \mathbb{R}^{m\times n}$ and $q > 0$. Then, it holds that*

$$\sum_{i=1}^{\min\{m,n\}} \sigma_i^q(X)\sigma_{n-i+1}^q(Y) \leq \left\|X^\top Y\right\|_q^q \leq \sum_{i=1}^{\min\{m,n\}} \sigma_i^q(X)\sigma_i^q(Y).$$

*Proof.* Let $f(s) = s^q$ for $s \in \mathbb{R}_+$. Then, the function $s \mapsto sf'(s) = qs^q$ is monotonically increasing on $\mathbb{R}_{++}$. The desired inequality then follows from Theorem 4. $\square$

## 4.2 Affine-Invariant Geometry on $\mathbb{P}_n$

It is well-known that the cone $\mathbb{P}_n$ of $n \times n$ positive definite matrices is a differentiable manifold of dimension $n(n + 1)/2$, see, *e.g.*, [4, Chapter 6]. For any $A \in \mathbb{P}_n$, 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: for any $X, Y \in T_A\mathbb{P}_n \cong \mathbb{S}_n$,

$$\langle X, Y\rangle_A := \operatorname{Tr}\left(XA^{-1}YA^{-1}\right).$$

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

$$d_{\mathbb{P}_n}(A, B) = \left\| \mathrm{Log}\left(B^{-\frac{1}{2}} A B^{-\frac{1}{2}}\right) \right\|_{\mathrm{F}},$$

where $\mathrm{Log}(\,\cdot\,)$ denotes the matrix logarithm. This distance enjoys many interesting properties [4, Chapter 6] and finds applications in diverse areas such as machine learning [19, 27], image and video processing [8, 30] and elasticity theory [17, 18].

More generally, for any $q \geq 1$ and $A, B \in \mathbb{P}_n$, we define

$$d_q(A, B) = \left\| \mathrm{Log}\left(B^{-\frac{1}{2}} A B^{-\frac{1}{2}}\right) \right\|_q.$$

It has been proved in [4, Section 6] that $d_q$ is a distance on $\mathbb{P}_n$ for any $q \geq 1$.

**Corollary 2.** *Let $A, B \in \mathbb{P}_n$ and $q \geq 1$. Then, it holds that*

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

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

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

where $f(s) = |\log s|^q$ for $s \in \mathbb{R}_{++}$. It can be readily verified that $f$ is differentiable on $\mathbb{R}_{++}$ and

$$f'(s) = \begin{cases} \frac{q}{s}(\log s)^{q-1}, & \text{if } s \geq 1, \\ -\frac{q}{s}(\log \frac{1}{s})^{q-1}, & \text{if } 0 < s < 1. \end{cases}$$

Hence,

$$s f'(s) = \mathrm{sgn}(\log s) \cdot q |\log s|^{q-1}, \tag{17}$$

where $\mathrm{sgn}(\cdot)$ denotes the sign of a real number. To show that $s \mapsto s f'(s)$ 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 $1 \geq s_2 > s_1 > 0$. For the first two cases, by (17), we have that $s_2 f'(s_2) \geq 0 \geq s_1 f'(s_1)$. For the third case of $s_2 > s_1 \geq 1$, (17) shows that the function $s \mapsto s f'(s)$ is continuous on $[1, \infty)$ and differentiable on $(1, \infty)$. Also, for any $s > 1$,

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

which implies that the function $s \mapsto s f'(s)$ is monotonically increasing on $[1, \infty)$. We therefore have $s_2 f'(s_2) \geq s_1 f'(s_1)$. Similarly, for the fourth case of $s_1 < s_2 \leq 1$, (17) shows that the function $s \mapsto s f'(s)$ is continuous on $(0, 1]$ and differentiable on $(0, 1)$. Also, for any $s \in (0, 1)$,

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

which implies that the function $s \mapsto sf'(s)$ is monotonically increasing on $(0, 1]$. We therefore have $s_2 f'(s_2) \geq s_1 f'(s_1)$. Hence, by Theorem 3(i), we obtain

$$\sum_{i=1}^{n} f\left(\lambda_i(A)\lambda_i(B^{-1})\right) \geq \sum_{i=1}^{n} f\left(\lambda_i(B^{-\frac{1}{2}}AB^{-\frac{1}{2}})\right) \geq \sum_{i=1}^{n} f\left(\lambda_i(A)\lambda_{n-i+1}(B^{-1})\right),$$

which is equivalent to

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

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

For any $A, B \in \mathbb{P}_n$, Corollary 2 with $q = 2$ immediately implies that the Riemannian distance $d_{\mathbb{P}_n}$ with respect to the affine-invariant metric satisfies the inequality

$$\sum_{i=1}^{n} \left(\log \lambda_i(A) - \log \lambda_i(B)\right)^2 \leq d_{\mathbb{P}_n}^2(A, B) \leq \sum_{i=1}^{n} \left(\log \lambda_i(A) - \log \lambda_{n-i+1}(B)\right)^2.$$

### 4.3 Alpha-Beta Log-Determinant Divergences

Divergences, which are measures of dissimilarity between two positive definite matrices, play an important role in information geometry and find applications across many areas, see [20, 6, 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 [6]. Given any $\alpha, \beta \in \mathbb{R}$ such that $\alpha\beta \neq 0$ and $\alpha + \beta \neq 0$, the AB log-det divergence with parameter $\alpha$ and $\beta$ between $A, B \in \mathbb{P}_n$ is defined as

$$D_{\alpha,\beta}(A\|B) = \frac{1}{\alpha\beta} \log\det\left(\frac{\alpha(AB^{-1})^{\beta} + \beta(AB^{-1})^{-\alpha}}{\alpha+\beta}\right)$$

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

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

For $\alpha = \beta = 0$, $D_{0,0}(A\|B) = \frac{1}{2} d_{\mathbb{P}_n}^2(A, B)$. We thus omit the discussion on this case and refer the readers to Section 4.2. Besides the squared affine-invariant Riemannian metric, many other well-known divergences are special cases of AB log-det divergences, including the S-divergence [26] where $\alpha = \beta = \frac{1}{2}$ and the Stein's loss [12] (also called the Burg divergence) where $\alpha = 0$ and $\beta = 1$. For more examples of AB log-det divergences, we refer the readers to [6, Section 3].

Note that $AB^{-1}$ is diagonalizable for any $A, B \in \mathbb{P}_n$. Therefore, as pointed out in [6], AB log-det divergences $D_{\alpha,\beta}(A\|B)$ can be expressed via the (positive) eigenvalues of the matrix $AB^{-1}$, which coincide with those of the matrix $B^{-\frac{1}{2}}AB^{-\frac{1}{2}}$:

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

The following upper and lower bounds for AB log-det divergences are generalizations of the [26, Corollary 3.8].

**Corollary 3.** *Let $A, B \in \mathbb{P}_n$ and $\alpha, \beta \in \mathbb{R}$. Then,*

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

*and*

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

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

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

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 $s>0$,

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

By Theorem 3, if $\alpha\beta>0$, then

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

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

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

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

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

We have that for any $s > 0$,

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

which is monotonically increasing regardless of the sign of $\alpha$. By Theorem 3,

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

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

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

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

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

We have that for any $s > 0$,

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

Then, for any $s > 0$,

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

By Theorem 3,

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

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. This work is supported in part by the Hong Kong Research Grants Council under the GRF project 15305321.

## 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. *Matrix Analysis*. Springer, 1997.

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

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

[6] 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.

[7] 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.

[8] 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.

[9] I. Gohberg and M. G. Kreĭn. *Introduction to the Theory of Linear Nonselfadjoint Operators*. American Mathematical Society, 1969.

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

[11] A. Iusem and A. Seeger. Angular analysis of two classes of non-polyhedral convex cones: The point of view of optimization theory. *Computational & Applied Mathematics*, 26:191–214, 2007.

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

[13] 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.

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

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

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

[17] 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.

[18] 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.

[19] 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.

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

[21] H. Ramírez, A. Seeger, and D. Sossa. Commutation principle for variational problems on Euclidean Jordan algebras. *SIAM Journal on Optimization*, 23(2):687–694, 2013.

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

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

[24] Y. Rychener, D. Kuhn, V. A. Nguyen, and M.-C. Yue. A geometric unification of distributionally robust covariance estimators: Shrinking the spectrum by inflating the ambiguity set. *arXiv*, 2024.

[25] R. Schatten. *Norm Ideals of Completely Continuous Operators*. Springer-Verlag, 1960.

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

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

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

[29] 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.

[30] 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.

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

[32] 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.