

The Rearrangement Inequality and Its Generalizations

Su Jianlin

2026-09-21

The rearrangement inequality is a classic inequality that is not hard to understand in itself; many readers may have already encountered it in middle school, especially those who prepared for mathematics competitions. But the rearrangement inequality is far more than what it appears to be on the surface: it has a profound counterpart in the world of matrices—the Von Neumann trace inequality; and it can be further generalized along the directions of “multiple sequences” and “multiple matrices.”

This article follows this thread and strings these topics together.

1 Basic Form

Let there be two real sequences sorted in descending order, $a_1 \geq a_2 \geq \cdots \geq a_n$ and $b_1 \geq b_2 \geq \cdots \geq b_n$; note that no non-negativity constraint is required here. Let $\tau(1), \tau(2), \cdots, \tau(n)$ denote an arbitrary permutation of $1, 2, \cdots, n$. Then the rearrangement inequality holds:

$$\sum_{i=1}^n a_i b_{n+1-i} \leq \sum_{i=1}^n a_i b_{\tau(i)} \leq \sum_{i=1}^n a_i b_i \quad (1)$$

that is, “reverse-order product $\leq$ arbitrary-order product $\leq$ same-order product.” The standard proof of the rearrangement inequality is the local adjustment (swapping) method: suppose $a_i \geq a_j$ and $b_i \geq b_j$, then

$$(a_i b_i + a_j b_j) - (a_i b_j + a_j b_i) = (a_i - a_j)(b_i - b_j) \geq 0 \quad (2)$$

In other words, if there is a pair of positions in the sequences with opposite order, then swapping one side makes them aligned in order, and the corresponding inner product will not get worse. Hence the maximum must be attained when the orders are fully aligned. Repeating the same procedure for $a_1, a_2, \cdots, a_n$ and $-b_1, -b_2, \cdots, -b_n$ yields the minimum side.

Note that the rearrangement inequality itself does not require the sequences to be non-negative, but later when we generalize it, we will add the non-negativity constraint. It is worth pointing out that adding the non-negativity constraint does not reduce the coverage

of the current rearrangement inequality, because any finite real sequence can be turned into a non-negative sequence by adding a sufficiently large constant, without changing the order, and after expanding both ends, the rearrangement inequality without the non-negativity constraint is recovered.

2 Trace Inequality

The rearrangement inequality has a very important counterpart in matrices: let $\mathbf{A}, \mathbf{B}$ be two $n \times n$ real symmetric matrices, and let $\lambda_i(\cdot)$ denote the i -th eigenvalue sorted in descending order. Then the inequality

$$\sum_{i=1}^n \lambda_i(\mathbf{A})\lambda_{n+1-i}(\mathbf{B}) \leq \text{tr}(\mathbf{AB}) \leq \sum_{i=1}^n \lambda_i(\mathbf{A})\lambda_i(\mathbf{B}) \quad (3)$$

holds. Clearly, when $\mathbf{A}, \mathbf{B}$ are diagonal matrices, this is equivalent to the rearrangement inequality of the previous section, so this is its matrix generalization, called the “Von Neumann’s Trace Inequality,” also known as “Ky Fan’s Trace Inequality”; in this article we uniformly abbreviate it as the “trace inequality.”

When $\mathbf{A}, \mathbf{B}$ are general $n \times m$ real matrices, the inequality

$$|\text{tr}(\mathbf{A}^\top \mathbf{B})| \leq \sum_{i=1}^{\min(n,m)} \sigma_i(\mathbf{A})\sigma_i(\mathbf{B}) \quad (4)$$

holds, where $\sigma_i(\cdot)$ denotes the i -th singular value sorted in descending order. The trace inequality has wide applications in mathematics and physics; for example, in deriving the Muon optimizer, knowing this inequality in advance can significantly reduce the difficulty of the proof. The general version of the inequality can be reduced to the symmetric version via the Jordan–Wielandt symmetrization trick, constructing

$$\tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^\top & \mathbf{0} \end{pmatrix}, \quad \tilde{\mathbf{B}} = \begin{pmatrix} \mathbf{0} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{0} \end{pmatrix} \quad (5)$$

so we only need to focus on the proof of the symmetric version.

3 Orthogonal Optimization

In this section we prove the symmetric version (3). Without loss of generality, we may assume $\mathbf{A}$ is diagonal; if not, we can decompose it as $\mathbf{A} = \mathbf{U}_A \mathbf{\Lambda}_A \mathbf{U}_A^\top$, and then use $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{U}_A \mathbf{\Lambda}_A \mathbf{U}_A^\top \mathbf{B} \mathbf{U}_A \mathbf{U}_A^\top) = \text{tr}(\mathbf{\Lambda}_A \mathbf{U}_A^\top \mathbf{B} \mathbf{U}_A)$, where $\mathbf{\Lambda}_A$ is diagonal, and the eigenvalues of $\mathbf{U}_A^\top \mathbf{B} \mathbf{U}_A$ coincide with those of $\mathbf{B}$.

There are many ways to prove Von Neumann's trace inequality; here we consider a rather elegant proof that is closer to Von Neumann's original idea. Introducing an $n \times n$ orthogonal matrix $\mathbf{O}$, we obviously have

$$\min_{\mathbf{O}} \text{tr}(\mathbf{A}\mathbf{O}^\top \mathbf{B}\mathbf{O}) \leq \text{tr}(\mathbf{A}\mathbf{B}) \leq \max_{\mathbf{O}} \text{tr}(\mathbf{A}\mathbf{O}^\top \mathbf{B}\mathbf{O}) \quad (6)$$

Next we need to prove that the maximum and minimum of $\text{tr}(\mathbf{A}\mathbf{O}^\top \mathbf{B}\mathbf{O})$ are exactly the upper and lower bounds we expect. First consider the maximum; suppose it is $\text{tr}(\mathbf{A}\mathbf{O}_*^\top \mathbf{B}\mathbf{O}_*)$. Note that for any antisymmetric matrix $\mathbf{S}$, the matrix exponential $e^{t\mathbf{S}}$ is orthogonal. Consider the perturbation $\mathbf{O}_* \rightarrow \mathbf{O}_* e^{t\mathbf{S}}$; by the optimality assumption on $\mathbf{O}_*$, the derivative at $t = 0$ should be 0. Direct computation gives

$$\frac{d}{dt} \text{tr}(\mathbf{A}(\mathbf{O}_* e^{t\mathbf{S}})^\top \mathbf{B}(\mathbf{O}_* e^{t\mathbf{S}})) = \text{tr}([\mathbf{A}, (\mathbf{O}_* e^{t\mathbf{S}})^\top \mathbf{B}(\mathbf{O}_* e^{t\mathbf{S}})]\mathbf{S}) \quad (7)$$

where $[\mathbf{X}, \mathbf{Y}] \triangleq \mathbf{X}\mathbf{Y} - \mathbf{Y}\mathbf{X}$ is commonly called the commutator; the antisymmetry of $\mathbf{S}$ was also used in the simplification. Substituting $t = 0$, setting it to zero, and using the arbitrariness of $\mathbf{S}$, we know $[\mathbf{A}, \mathbf{O}_*^\top \mathbf{B}\mathbf{O}_*] = \mathbf{0}$, i.e., $\mathbf{A}$ commutes with $\mathbf{O}_*^\top \mathbf{B}\mathbf{O}_*$. The same argument applies to the minimum.

Note that $\mathbf{A}$ is diagonal; further assume that its diagonal entries are pairwise distinct (if not, it can always be expressed as the limit of a sequence of matrices satisfying the condition). Then the matrices commuting with $\mathbf{A}$ are also diagonal, so $\mathbf{O}_*$ must be an eigenmatrix of $\mathbf{B}$, or in other words, $\mathbf{O}_*^\top \mathbf{B}\mathbf{O}_*$ must be a diagonalization of $\mathbf{B}$. Thus the problem becomes finding the upper and lower bounds of the inner product of the eigenvalue vectors of two matrices, which follows from the rearrangement inequality.

4 Multiple Sequences

A very natural question is: can the rearrangement inequality be generalized to multiple sequences? Take three sequences as an example: let there be non-negative sequences $a_1 \geq \dots \geq a_n \geq 0$, $b_1 \geq \dots \geq b_n \geq 0$, and $c_1 \geq \dots \geq c_n \geq 0$, where $\tau(1), \dots, \tau(n)$ and $\pi(1), \dots, \pi(n)$ denote two arbitrary permutations of $1, \dots, n$. Then the maximum is still attained when the orders are aligned:

$$\sum_{i=1}^n a_i b_{\tau(i)} c_{\pi(i)} \leq \sum_{i=1}^n a_i b_i c_i \quad (8)$$

The proof is not complicated and can be written in one line:

$$\sum_{i=1}^n a_i b_{\tau(i)} c_{\pi(i)} = \sum_{k=1}^n (a_k - a_{k+1}) \sum_{i=1}^k b_{\tau(i)} c_{\pi(i)} \leq \sum_{k=1}^n (a_k - a_{k+1}) \sum_{i=1}^k b_i c_i = \sum_{i=1}^n a_i b_i c_i \quad (9)$$

The first equality is an identity transformation using summation by parts (setting $a_{n+1} = 0$); the second equality transforms back using summation by parts; the key is the inequality sign $\leq$: under the condition $a_k \geq a_{k+1}$, it applies the rearrangement inequality to $\sum_{i=1}^k b_{\tau(i)} c_{\pi(i)}$. Note that b_i, c_i are already sorted in descending order, so $\sum_{i=1}^k b_i c_i$ is necessarily maximal, and no matter how $\sum_{i=1}^k b_{\tau(i)} c_{\pi(i)}$ is re-paired, it cannot exceed it, hence this inequality.

This proof and result can be recursively generalized to any number of non-negative sequences. Note that non-negativity is necessary, because “no matter how $\sum_{i=1}^k b_{\tau(i)} c_{\pi(i)}$ is re-paired, it cannot exceed $\sum_{i=1}^k b_i c_i$ ” is only guaranteed for non-negative sequences.

5 Automatic Balancing

However, the minimum side cannot be generalized as simply as the maximum side.

The reason is intuitive: “same order” is transitive—the Hadamard product of two monotonically decreasing non-negative sequences is still monotonically decreasing, so we can repeatedly reuse the binary rearrangement inequality to obtain the maximum. However, “reverse order” is inherently defined for two sequences; for multiple sequences it is hard to define what “reverse order” means, and the Hadamard product of an increasing and a decreasing non-negative sequence has no fixed monotonicity.

Nevertheless, we can still get some signal about the minimum from the AM–GM inequality:

$$\sum_{i=1}^n a_i b_{\tau(i)} c_{\pi(i)} \geq n \left(\prod_{i=1}^n a_i b_{\tau(i)} c_{\pi(i)} \right)^{1/n} = n \left(\prod_{i=1}^n a_i b_i c_i \right)^{1/n} \quad (10)$$

The equality condition is that all $a_i b_{\tau(i)} c_{\pi(i)}$ are equal. Of course, since we only changed the orders of b, c , it is hard to achieve the condition that all $a_i b_{\tau(i)} c_{\pi(i)}$ are equal, but it tells us a qualitative direction: the ordering corresponding to the minimum will try to make all $a_i b_{\tau(i)} c_{\pi(i)}$ balanced.

Experienced readers may keenly realize that solving such discrete balanced pairing problems is usually rather difficult. This is indeed the case: it can essentially be abstracted as a three-dimensional axial assignment problem, which is known to be NP-Hard. In other words, the minimum side of the multi-sequence rearrangement inequality has no general closed-form solution; this is the essential difficulty of the problem.

6 Multiple Matrices

Similarly, can the trace inequality be generalized to multiple matrices? Consider three $n \times n$ matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$; we have

$$|\text{tr}(\mathbf{ABC})| \leq \sum_{i=1}^n \sigma_i(\mathbf{A}) \sigma_i(\mathbf{B}) \sigma_i(\mathbf{C}) \quad (11)$$

The proof is similar:

$$\begin{aligned}
|\text{tr}(\mathbf{A}\mathbf{B}\mathbf{C})| &\leq \sum_{i=1}^n \sigma_i(\mathbf{A})\sigma_i(\mathbf{B}\mathbf{C}) \\
&= \sum_{k=1}^n (\sigma_k(\mathbf{A}) - \sigma_{k+1}(\mathbf{A})) \sum_{i=1}^k \sigma_i(\mathbf{B}\mathbf{C}) \\
&\leq \sum_{k=1}^n (\sigma_k(\mathbf{A}) - \sigma_{k+1}(\mathbf{A})) \sum_{i=1}^k \sigma_i(\mathbf{B})\sigma_i(\mathbf{C}) \\
&= \sum_{i=1}^n \sigma_i(\mathbf{A})\sigma_i(\mathbf{B})\sigma_i(\mathbf{C})
\end{aligned} \tag{12}$$

that is, first tighten once with the binary trace inequality, then again summation by parts, tightening, and summation by parts, where the second $\leq$ uses the following inequality

$$\sum_{i=1}^k \sigma_i(\mathbf{B}\mathbf{C}) \leq \sum_{i=1}^k \sigma_i(\mathbf{B})\sigma_i(\mathbf{C}), \quad k = 1, 2, \dots, n \tag{13}$$

This can also be viewed as a generalization of Eq. (4); the proof idea is to first convert the left-hand side into a trace, and then apply Eq. (4). Specifically, let the SVD of $\mathbf{B}\mathbf{C}$ be $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$; then

$$\sum_{i=1}^k \sigma_i(\mathbf{B}\mathbf{C}) = \text{tr}(\mathbf{U}_{[:, :k]}^\top \mathbf{B}\mathbf{C}\mathbf{V}_{[:, :k]}) \leq \sum_{i=1}^k \sigma_i(\mathbf{U}_{[:, :k]}^\top \mathbf{B})\sigma_i(\mathbf{C}\mathbf{V}_{[:, :k]}) \leq \sum_{i=1}^k \sigma_i(\mathbf{B})\sigma_i(\mathbf{C}) \tag{14}$$

The last inequality sign uses $\sigma_i(\mathbf{X}\mathbf{Y}) \leq \sigma_i(\mathbf{X})\sigma_1(\mathbf{Y})$ and the orthogonality of $\mathbf{U}_{[:, :k]}$, $\mathbf{V}_{[:, :k]}$. The same induction can continue, yielding the version for any number of matrices.

7 Another Generalization

Equation (11) generalizes Eq. (4), so how can Eq. (3) be generalized? Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ be $n \times n$ positive semidefinite matrices and consider $\text{tr}(\mathbf{A}\mathbf{B}\mathbf{C})$. The maximum side is trivial, namely Eq. (11); but can the minimum side similarly be written in the form $\sum_{i=1}^n \lambda_i(\mathbf{A})\lambda_{\tau(i)}(\mathbf{B})\lambda_{\pi(i)}(\mathbf{C})$?

The answer is no. The eigenvalues of positive semidefinite matrices are necessarily non-negative, so $\sum_{i=1}^n \lambda_i(\mathbf{A})\lambda_{\tau(i)}(\mathbf{B})\lambda_{\pi(i)}(\mathbf{C})$ is necessarily non-negative, but the trace of the product of three positive semidefinite matrices can be negative! Therefore such a form of the minimum cannot hold. Where does the problem lie? Note that for positive semidefinite matrices $\mathbf{A}, \mathbf{B}$, we have

$$\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}((\mathbf{A}^{1/2}\mathbf{B}^{1/2})^\top (\mathbf{A}^{1/2}\mathbf{B}^{1/2})) = \|\mathbf{A}^{1/2}\mathbf{B}^{1/2}\|_F^2 \geq 0 \tag{15}$$

which guarantees non-negativity. From this one may conjecture that a reasonable three-matrix generalization might be $\|\mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{C}^{1/2}\|_F^2$, rather than the simpler $\text{tr}(\mathbf{ABC})$. Unfortunately, although this expression looks more “correct,” it can be shown by counterexample that its minimum is not necessarily equal to some $\sum_{i=1}^n \lambda_i(\mathbf{A})\lambda_{\tau(i)}(\mathbf{B})\lambda_{\pi(i)}(\mathbf{C})$, so it also cannot serve as the generalization of the minimum side of Eq. (3).

However, $\|\mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{C}^{1/2}\|_F^2$ does have some properties similar to the minimum side of the multi-sequence rearrangement inequality; for example,

$$\begin{aligned}\|\mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{C}^{1/2}\|_F^2 &= \sum_{i=1}^n \sigma_i^2(\mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{C}^{1/2}) \\ &\geq n \left(\prod_{i=1}^n \sigma_i(\mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{C}^{1/2}) \right)^{2/n} \\ &= n[\det(\mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{C}^{1/2})]^{2/n} \\ &= n[\det(\mathbf{A})\det(\mathbf{B})\det(\mathbf{C})]^{1/n}\end{aligned}\tag{16}$$

with equality when all singular values of $\mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{C}^{1/2}$ are equal. This means that when $\|\mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{C}^{1/2}\|_F^2$ attains its minimum, all singular values of $\mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{C}^{1/2}$ tend toward a uniform distribution—“automatic balancing” appears again!

8 Summary

Starting from the rearrangement inequality, this article reviewed its main generalizations along two threads: “matrixization” and “multi-sequencing.”

When reprinting, please include the address of this article:

<https://kerue.fm/archives/11910>

For more detailed reprinting matters, please refer to: "Scientific Spaces" FAQ