Proof of a trace inequality in matrix algebra

Howard E. Haber

Santa Cruz Institute for Particle Physics
University of California, Santa Cruz, CA 95064, USA

Abstract

Given two positive definite matrices X and Y, we prove that $\text{Tr}\,[(XY)^r] \leq \left[\text{Tr}\,(X^{2r})\right]^{1/2}\left[\text{Tr}\,(Y^{2r})\right]^{1/2}$ for any real number r. For $Y = X^*$ (where X^* is the complex conjugate of X) and $r = -1/2$, this inequality reduces to $\text{Tr}\,(XX^*)^{-1/2} \leq \text{Tr}\,(X^{-1})$. The implications of the latter inequality are exhibited.

Consider an $n \times n$ matrix X. By definition, X is positive definite if

$$\sum_{ij} X_{ij} a_i^* b_j > 0\,, \tag{1}$$

for all complex vectors a_i and b_j. One can easily prove that if X is positive definite then X is hermitian (see, *e.g.*, Ref. [1], p. 65). Since the eigenvalues of hermitian matrices are real, it is easy to prove that the eigenvalues of positive definite matrices are real and positive. Moreover, a positive definite matrix is invertible, since it does not possess a zero eigenvalue. Note that a non-hermitian matrix whose eigenvalues are all strictly real and positive is not positive definite. Finally, we note that for any matrix S, if X is positive definite, then $S^\dagger XS$ is also positive definite. This follows from eq. (1) by replacing a and b with Sa and Sb, respectively.

Refs. [2] and [3] define a weakly positive definite matrix A to be a matrix that can be written as $A = SXS^{-1}$ for some non-singular matrix S, where X is positive definite.[1] Since two matrices related by a similarity transformation possess an identical eigenvalue spectrum, it follows that all the eigenvalues of A are real and positive. Moreover, since X is hermitian, it is also diagonalizable, and it then follows that A is diagonalizable as well.[2]

[1] Note that a weakly positive definite hermitian matrix is a positive definite matrix.

[2] In particular, a non-diagonalizable matrix with only real positive eigenvalues is not weakly positive definite.

Consider two positive definite matrices X and Y. Then XY is positive definite if and only if X and Y commute (in which case XY is hermitian and X and Y are simultaneously diagonalizable). The case where X and Y do not commute is covered by the following lemma [2, 3, 4]:

Lemma 1: A matrix is weakly positive definite if and only if it can be written as the product of two positive definite matrices.

Thus, if X and Y are positive definite, then XY and YX are both weakly positive definite. Moreover, if X and Y are positive definite, then all the eigenvalues of XY and YX, respectively, are real and positive. The proof of the lemma is straightforward. First, note that $A = SXS^{-1} = P_1P_2$, where $P_1 = SS^\dagger$ and $P_2 = [S^{-1}]^\dagger XS^{-1}$, where P_1 and P_2 are clearly positive definite if X is positive definite.

To prove the converse, we first define $X^{1/2}$ to be the unique positive definite square root[3] of X. Then, $X^{1/2}YX^{1/2} = [X^{1/2}]^\dagger YX^{1/2}$ is positive definite. If we now write:

$$XY = X^{1/2}[X^{1/2}YX^{1/2}][X^{1/2}]^{-1}\,, \tag{2}$$

it follows that XY and the positive definite matrix $X^{1/2}YX^{1/2}$ are related by a similarity transformation. Consequently, all the eigenvalues of XY are real and positive. A similar proof yields the same conclusion for YX.

We next consider a number of important inequalities involving eigenvalues and singular values of a complex matrix. For any matrix A, its singular values are defined as the positive square roots of the eigenvalues of $AA^\dagger$ (or equivalently of $A^\dagger A$). We shall denote the eigenvalues of A by λ_i and the singular values by σ_i.

Lemma 2: For any non-singular[4] $n \times n$ matrix A,

$$\sum_{i=1}^{n} |\lambda_i|^r \leq \sum_{i=1}^{n} \sigma_i^r\,, \tag{3}$$

for any real number r.

This result was first stated and proved by Weyl [5]. We shall apply Lemma 2 to a weakly positive definite matrix A. Since the eigenvalues of A

[3]For a proof that a positive definite matrix has a unique positive definite square root, see, *e.g.*, Ref. [1], p. 162 or Ref. [4], p.405.

[4]If A is singular, then Lemma 2 holds for non-negative real r.

are real and non-negative, it follows that:

$$\sum_{i=1}^{n} |\lambda_i|^r = \sum_{i=1}^{n} \lambda_i^r = \text{Tr}\,(A^r)\,. \tag{4}$$

To obtain the last equality above, recall that A is diagonalizable, so that $A^r = SD^rS^{-1}$, where $D = S^{-1}AS = \text{diag}(\lambda_1, \lambda_2 \ldots, \lambda_\text{n})$. In defining A^r, we use $D^r = \text{diag}(\lambda_1^\text{r}, \lambda_2^\text{r} \ldots, \lambda_\text{n}^\text{r})$, where $\lambda_i^r = \exp(r \ln \lambda_i) > 0$ by taking the principal value of the logarithm. Thus, A^r is uniquely defined and is weakly positive definite. Hence, for any weakly positive definite $n \times n$ matrix A,

$$\text{Tr}\,(A^r) \leq \sum_{i=1}^{n} \sigma_i^r\,. \tag{5}$$

Lemma 3: For any two non-singular $n \times n$ matrices A and B,

$$\sum_{i=1}^{n} \sigma_i^r(AB) \leq \sum_{i=1}^{n} \sigma_i^r(A)\sigma_i^r(B)\,, \tag{6}$$

for any real number r where the $\sigma_i(A)$ are the singular values of A and $\sigma_i^r(A) \equiv [\sigma_i(A)]^r$, *etc.*

This is Theorem 3.3.14(e) of Ref. [6]. We can derive another useful inequality as follows. First, we square both sides of eq. (6). The inequality is preserved since all terms in the sums are positive. We then make use of one further inequality:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_i^r(A)\sigma_i^r(B)\sigma_j^r(A)\sigma_j^r(B) \leq \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_i^{2r}(A)\sigma_j^{2r}(B)\,. \tag{7}$$

This inequality is a trivial consequence of the identity:

$$\begin{aligned} &\sum_{i=1}^{n}\sum_{j=1}^{n} \left[\sigma_i^{2r}(A)\sigma_j^{2r}(B) - \sigma_i^r(A)\sigma_i^r(B)\sigma_j^r(A)\sigma_j^r(B)\right] \\ &\qquad = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \left[\sigma_i^r(A)\sigma_j^r(B) - \sigma_j^r(A)\sigma_i^r(B)\right]^2 \geq 0\,. \end{aligned} \tag{8}$$

Hence, eqs. (6) and (7) imply that:

$$\left[\sum_{i=1}^{n} \sigma_i^r(AB)\right]^2 \leq \left[\sum_{i=1}^{n} \sigma_i^{2r}(A)\right]\left[\sum_{j=1}^{n} \sigma_j^{2r}(B)\right]. \tag{9}$$

Note that $\sigma_i(A^{2r}) = \sigma_i^{2r}(A)$ for any non-singular[5] matrix A. This follows from:

$$\sigma(A^{2r}) = \{\lambda[(AA^\dagger)^{2r}]\}^{1/2} = [\lambda(AA^\dagger)]^r = [\sigma(A)]^{2r} \,. \tag{10}$$

The next to last equality is obtained by diagonalizing the positive definite matrix $AA^\dagger$. Hence, we can rewrite eq. (9) as:

$$\left[\sum_{i=1}^{n} \sigma_i^r(AB)\right]^2 \leq \left[\sum_{i=1}^{n} \sigma_i(A^{2r})\right]\left[\sum_{j=1}^{n} \sigma_j(B^{2r})\right] \,. \tag{11}$$

The case of $r = -1/2$ is posed as problem 25 on p. 190 of Ref. [6].

We now apply eq. (11) to the case of positive definite matrices X and Y. In this case, we can combine eq. (11) with eq. (5) [since XY is weakly positive definite] to obtain:

$$\text{Tr}\,[(XY)^r] \leq \left[\sum_{i=1}^{n} \sigma_i(X^{2r})\right]^{1/2} \left[\sum_{j=1}^{n} \sigma_j(Y^{2r})\right]^{1/2} \,. \tag{12}$$

Recall that if X is positive definite, then $\sigma_i(X) = \lambda_i(X)$. Thus, eq. (12) is equivalent to the following result:

Theorem 1: If X and Y are positive definite $n \times n$ matrices, then

$$\text{Tr}\,[(XY)^r] \leq \left[\text{Tr}\,(X^{2r})\right]^{1/2} \left[\text{Tr}\,(Y^{2r})\right]^{1/2} \,, \tag{13}$$

for any real number r, where $(XY)^r$ is defined to be the weakly positive definite rth power of XY.

Theorem 1 is proved for the case of integer r in Ref. [7]. However, the proof given in this paper appears to hold for any real number r.

We now consider the case of $Y = X^*$, where X^* is the complex conjugate of X. Note that if A is positive definite, then X^* is also positive definite. It then follows that XX^* is diagonalizable, and all of its eigenvalues are real and positive. If we diagonalize $XX^* = QDQ^{-1}$, where D is a diagonal matrix with real positive diagonal elements, then we can define a unique[6] square root

[5] As in footnote 4, for singular matrices, one must restrict the real number r to non-negative values.

[6] According to a theorem proved in Ref. [8], for any non-singular matrix with the property that none of its eigenvalues lie on the negative real axis, there exists a unique square root whose eigenvalues lie completely in the open right complex half plane (i.e., the eigenvalues λ_i satisfy Re $\lambda_i > 0$ for all i). It follows that $(XX^*)^{1/2}$, whose eigenvalues are all real and positive, is uniquely defined.

$(XX^*)^{1/2} = QD^{1/2}Q^{-1}$, where $D^{1/2}$ is the diagonal matrix consisting of the positive square roots of the diagonal elements of D. Note that since XX^* is a complex symmetric matrix, it follows that Q is in general a complex orthogonal matrix. Hence, $(XX^*)^{1/2}$ is also a complex symmetric matrix. We are now ready to state the second theorem of interest to this paper.

Theorem 2: Let X be a positive definite matrix. The following inequality holds:

$$\text{Tr } (XX^*)^{-1/2} \leq \text{Tr } (X^{-1})\,, \tag{14}$$

where $(XX^*)^{1/2}$ is the unique complex symmetric matrix whose eigenvalues are all real and positive.

Theorem 2 is a trivial consequence of putting $r = -1/2$ and $Y = X^*$ in eq. (13), and noting that $\text{Tr } (X^{-1}) = \text{Tr } [(X^*)^{-1}]$ if X is positive definite.

One can reformulate Theorem 2 in the two different ways. We begin with the definition of the *adjugate* of a matrix Z, which is defined by

$$Z^A \equiv (\det\, Z) Z^{-1}\,. \tag{15}$$

It is easy to prove the following result. Given a matrix Z whose characteristic polynomial is given by:

$$\chi_Z(s) = (s-\lambda_1)(s-\lambda_2)\cdots(s-\lambda_n) = s^n + \beta_1 s^{n-1} + \cdots + \beta_{n-1}s + \beta_n\,, \tag{16}$$

then $(-1)^n\beta_n = \lambda_1\lambda_2\cdots\lambda_n = \det\, Z$ and [9]

$$(-1)^{n-1}\beta_{n-1} = \lambda_1\lambda_2\cdots\lambda_n \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \cdots + \frac{1}{\lambda_n}\right) = \text{Tr } Z^A\,. \tag{17}$$

Consequently, eq. (14) is equivalent to the inequality:

$$\text{Tr } \{[(XX^*)^{1/2}]^A\} \leq \text{Tr } [X^A]\,, \tag{18}$$

where we have used eq. (15) and the fact that $\det\, (XX^*) = \det\, (X^2) = (\det\, X)^2$ if X is positive definite. Equivalently,

$$\rho_1\rho_2\cdots\rho_n \left(\frac{1}{\rho_1} + \frac{1}{\rho_2} + \cdots + \frac{1}{\rho_n}\right) \leq \lambda_1\lambda_2\cdots\lambda_n \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \cdots + \frac{1}{\lambda_n}\right)\,, \tag{19}$$

where the ρ_i are the eigenvalues of $(XX^*)^{1/2}$ and the λ_i are the eigenvalues of X.

We can express β_{n-1} (or equivalently Tr Z^A) in terms of $t_k \equiv \text{Tr}\,(Z^k)$, where $k = 1, 2, \ldots, n-1$ by using the following recursion relation [10]:

$$t_k + \beta_1 t_{k-1} + \cdots + \beta_{k-1}\, t_1 + k\beta_k = 0\,, \qquad k = 1, 2, \ldots, n-1\,. \tag{20}$$

An explicit formula can be found in [11]:

$$\text{Tr}\ Z^A = (-1)^{n-1} \sum_{(p_1, p_2, \ldots, p_{n-1}) \in \mathcal{S}} \prod_{m=1}^{n-1} \frac{1}{p_m!} \left[-\frac{t_m}{m} \right]^{p_m}\,, \tag{21}$$

where $\mathcal{S}$ is the set of all non-negative integer solutions $\{(p_1, p_2, \ldots, p_{n-1})\}$ of the equation $p_1 + 2p_2 + \ldots (n-1)p_{n-1} = n-1$. Explicit results can also be found in [12].[7] However, it is not likely that these results can be used for a direct proof of the inequality given in eq. (18).

APPENDIX

Theorem 2 can be directly proven in the case of 3×3 matrices. Any 3×3 matrix Z satisfies its characteristic equation,

$$Z^3 - (\text{Tr}\ Z)Z^2 + \tfrac{1}{2}[(\text{Tr}\ Z)^2 - \text{Tr}\ (Z^2)]Z - \det\ Z = 0\,. \tag{22}$$

Multiplying eq. (22) by Z^{-1} and taking the trace yields:

$$(\det\ Z)\text{Tr}\ Z^{-1} = \tfrac{1}{2}[(\text{Tr}\ Z)^2 - \text{Tr}\ (Z^2)]\,. \tag{23}$$

In this case, eq. (14) is equivalent to the statement that

$$[\text{Tr}\ (B^{1/2})]^2 - \text{Tr}\ B \,\geq\, [\text{Tr}\ (A^{1/2})]^2 - \text{Tr}\ A\,, \tag{24}$$

where $A = XX^*$ and $B = XX^\dagger = X^2$ [and we have used $\det A = \det B = (\det X)^2$]. We can rearrange eq. (24) into the following form:

$$[\text{Tr}\ (B^{1/2}) - \text{Tr}\ (A^{1/2})][\text{Tr}\ (B^{1/2}) + \text{Tr}\ (A^{1/2})] \geq \text{Tr}\ (B) - \text{Tr}\ (A)\,, \tag{25}$$

or equivalently,

$$[\text{Tr}\ (B^{1/2} - A^{1/2})][\text{Tr}\ (B^{1/2} + A^{1/2})] \geq \text{Tr}\ (B - A) \tag{26}$$

To prove eq. (24), we make use of the following lemma [14]:

Lemma 4: If V and W are non-negative definite matrices, then

$$[\text{Tr}\ (V)][\text{Tr}\ (W)] \geq |\text{Tr}\ (VW)|\,. \tag{27}$$

Choose $V = B^{1/2} - A^{1/2}$ and $W = B^{1/2} + A^{1/2}$. Then, it easily follows that $\text{Tr}\ (VW) = \text{Tr}\ (B - A)$, where we have used the fact that $\text{Tr}\ (A^{1/2}B^{1/2}) = \text{Tr}\ (B^{1/2}A^{1/2})$. Using eq. (27), we immediately verify eq. (26).

[7]These results are known as Newton's identities [13].

References

[1] Fuzhen Zhang, *Matrix Theory: Basic results and techniques* (Springer-Verlag, New York, 1999).

[2] E.P. Wigner, *Canad. J. Math*, **15**, 313 (1963).

[3] M.L. Mehta, *Matrix Theory: selected topics and useful results* (Les Editions de Physique, Les Ulis, France and Hindustan Publishing Company, Delhi, India, 1977). See, in particular Lemma 9.18.3 on p. 211.

[4] Roger A. Horn and Charles R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, UK, 1985).

[5] H. Weyl, *Proc. Nat. Acad. Sci. USA*, **35**, 408 (1949).

[6] Roger A. Horn and Charles R. Johnson, *Topics in Matrix Analysis* (Cambridge University Press, Cambridge, UK, 1991).

[7] X.M. Yang, X.Q. Yang and K.L. Teo, *J. Math. Anal. App.* **263**, 327 (2001).

[8] C.R. Johnson, K. Okubo and R. Reams, *Linear Algebra Appl.* **323**, 51 (2001).

[9] Dennis S. Bernstein, *Matrix Mathematics* (Princeton University Press, Princeton, NJ, 2005). See proposition 4.4.5 on pp. 132–133.

[10] R.R. Silva, *J. Math. Phys.* **39**, 6206 (1998).

[11] H.-H. Zhang, W.-B. Yan and X.-S. Li, arXiv:hep-th/0701116 (2007).

[12] V. Tapia, arXiv:math-ph/0702001.

[13] K.S. Miller, *Some Eclectic Matrix Theory* (Krieger, 1987).

[14] V.V. Prasolov, *Problems and Theorems in Linear Algebra* (American Mathematical Society, Providence, RI, 1994), problem 34.3 on p. 148 (solution given on p. 164).