POSITIVE DEFINITE MATRICES: SOME NOTES

PIYUSH SRIVASTAVA

The goal of these notes is to have a self contained derivation of some of the operator inequalities needed in proofs of the "standard" matrix concentration inequalities. All material here is standard: more details can be found in the book by Bhatia [2007], and in lecture notes by Carlen [2010]. An important topic currently missing from these notes is Lieb's theorem [Lieb, 1973] that for every Hermitian matrix H, the map $A \mapsto \operatorname{tr}\exp(H + \log A)$ is concave on the set of positive definite matrices.

1. Notation and Preliminaries

All matrices, unless otherwise stated, will be assumed to be square. For a vector or matrix A, A^* denotes its complex conjugate transpose. A matrix $M \in \mathbb{C}^{n\times n}$ is said to be positive semi-definite (PSD) if $v^*Av \geq 0$ for all $v \in \mathbb{C}^n$; it is said to be positive definite (PD) if $v^*Av > 0$ for all $v \neq 0 \in \mathbb{C}^n$. Note that this definition implies that any positive semi-definite A must be Hermitian, $A = A^*$, and thus must admit n orthonormal eigenvectors with non-negative (or, in the case of positive definite matrices, positive) eigenvalues.

For a matrix A, $A^\dagger$ denotes its Moore-Penrose pseudo-inverse: for a diagonal matrix D, $D^\dagger$ is defined as the diagonal matrix obtained by inverting all the non-zero entries of D, while when A is not diagonal and has a singular value decomposition $A = UDV^*$, we define $A^\dagger = UD^\dagger V^*$ (note that this definition indeed uniquely defines $A^\dagger$, even when U and V are not uniquely determined by A).

Proposition 1.1. *A is positive definite if and only if A^{-1} exists and is positive definite. A is positive semi-definite if and only if $A^\dagger$ is positive semi-definite. Further, whenever A is Hermitian, $AA^\dagger = A^\dagger A$ is the orthogonal projection onto the range of A.*

The following properties of PD matrices follow easily from the definition.

Proposition 1.2.
(1) *If A and B are PSD, then $A + B$ is PSD. If at least one of A and B is, in addition, PD, then $A + B$ is also PD.*
(2) *If s is a positive real number and A is PSD (respectively PD), then sA is PSD (respectively, PD).*
(3) *Is A is a PSD (respectively, PD) matrix, then every principal submatrix of A is PSD (respectively, PD).*
(4) *If A is Hermitian, then $A \preceq I$ if and only if all eigenvalues of A are at most 1.*

Definition 1.1 (Congruence). Two matrices A and B are said to be *congruent*, denoted $A \sim B$, if there exists an invertible matrix X such that $A = X^*BX$.[1]

The following proposition again follows easily from the definition.

Proposition 1.3. *Suppose A and B are congruent matrices. Then A is positive definite if and only if B is positive definite. Similarly, A is positive semi-definite if and only if B is positive semi-definite.*

Fact 1.4. *Let M be the block matrix $\begin{bmatrix} I & A \\ 0 & I \end{bmatrix}$, where all blocks are assumed to be squares of the same dimensions. Then M is invertible with inverse $\begin{bmatrix} I & -A \\ 0 & I \end{bmatrix}$.*

[1] Compare with the notion of *similarity*: matrices A and B are similar if there is an invertible matrix X such that $X^{-1}AX = B$. If A and B are similar, then they have the same multiset of eigenvalues. However, two congruent matrices need not have the same multiset of eigenvalues.

2. Block matrices

The following two innocuous (and, perhaps, contrived) looking propositions will turn out to be extremely useful.

Proposition 2.1. *Let A be PSD, B PD, and X any matrix, all of the same dimension. Then, the matrix $\begin{bmatrix} A & X \\ X^* & B \end{bmatrix}$ is PSD (respectively, PD) if and only if $A - XB^{-1}X^*$ is PSD (respectively, PD).*

Proof. We have

$$\begin{bmatrix} A & X \\ X^* & B \end{bmatrix} \sim \begin{bmatrix} A - XB^{-1}X^* & O \\ O & B \end{bmatrix},$$

since

$$\begin{bmatrix} A - XB^{-1}X^* & O \\ O & B \end{bmatrix} = \begin{bmatrix} I & -XB^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & X \\ X^* & B \end{bmatrix} \begin{bmatrix} I & O \\ -B^{-1}X^* & I \end{bmatrix}.$$

Now, when B is PD, the latter matrix is PSD (respectively, PD) if and only if $A - XB^{-1}X^*$ is PSD (respectively, PD). This completes the proof. □

The same proof can also be arranged in a slightly different manner, so that it does not require A and B to be PD.

Proposition 2.2. *Let A, B and X be matrices of the same dimension. Then, the matrix $\begin{bmatrix} A & X \\ X^* & B \end{bmatrix}$ is PSD if and only if $A - XB^\dagger X$ and B are PSD and $XB^\dagger B = X$.*

Proof. Suppose first that $A - XB^\dagger X$ and B are PSD. Thus, they are also Hermitian, and, in particular $B^{\dagger *} = B^\dagger$. We then have

$$M := \begin{bmatrix} A & X \\ X^* & B \end{bmatrix} \sim \begin{bmatrix} A - XB^\dagger X^* & X(I - B^\dagger B) \\ (I - BB^\dagger)X^* & B \end{bmatrix} =: N,$$

since

$$\begin{bmatrix} A - XB^\dagger X^* & X(I - B^\dagger B) \\ (I - BB^\dagger)X^* & B \end{bmatrix} = \begin{bmatrix} I & -XB^\dagger \\ 0 & I \end{bmatrix} \begin{bmatrix} A & X \\ X^* & B \end{bmatrix} \begin{bmatrix} I & O \\ -B^\dagger X^* & I \end{bmatrix}. \tag{1}$$

(The above calculation uses the identity $B^\dagger - B^\dagger BB^\dagger = O$.) Now, suppose also that $XB^\dagger B = X$. Then $N = \begin{bmatrix} A - XB^\dagger X^* & O \\ O & B \end{bmatrix}$ is a block diagonal matrix which is PSD since both its blocks $A - XB^\dagger X$ and B are PSD. Since M is congruent to N, we see from Proposition 1.3 that M is also PSD. We have thus established that

$$A - XB^\dagger X^* \geq 0,\ B \geq 0 \text{ and } XB^\dagger B = X \quad \Longrightarrow \quad M \geq 0.$$

This proves one side of the equivalence. Now, suppose that M is PSD. Then, A and B are PSD (and hence Hermitian), so that the congruence in eq. (1) is again valid, and implies by Proposition 1.3 that N is PSD. This then implies that $A - XB^\dagger X^*$ is also PSD. Now, suppose, if possible, that $X \neq XB^\dagger B$. Since $B^\dagger B$ is an orthogonal projector onto the range of the Hermitian matrix B, $X \neq XB^\dagger B$ implies that there exists a vector v such that $Bv = 0$ and $Xv \neq 0$. Now, consider the vector $w := \begin{bmatrix} \tau Xv \\ v \end{bmatrix}$, where τ is a negative real number to be fixed later. We then compute

$$w^* N w = 2\tau \|Xv\|_2^2 + \tau^2 \gamma,$$

where $\gamma := v^* X^* (A - XB^\dagger X^*) Xv$ is a real number. But this gives a contradiction, since by choosing τ to be a negative number of small enough magnitude we can force $w^* N w < 0$, which contradicts the deduction above that N is PSD. Thus, we must also have $XB^\dagger B = X$. This proves the second part of the equivalence. □

From symmetry, Proposition 2.2 yields the following.

Proposition 2.3. *Let A, B and X be square matrices of the same dimension. Then, the following are equivalent.*

(1) $\begin{bmatrix} A & X \\ X^* & B \end{bmatrix}$ *is PSD.*
(2) $A - XB^\dagger X^*$ *and* B *are PSD, and* $XB^\dagger B = X$.
(3) $B - X^* A^\dagger X$ *and* A *are PSD, and* $AA^\dagger X = X$.

For PD matrices, we have the following simpler version.

Proposition 2.4. *Let* A, B *and* X *be square matrices of the same dimension. Then, the following are equivalent.*

(1) $\begin{bmatrix} A & X \\ X^* & B \end{bmatrix}$ *is PD.*
(2) $A - XB^{-1}X^*$ *and* B *are PD.*
(3) $B - X^* A^{-1} X$ *and* A *are PD.*

3. Functions of matrices

Let $f : \mathbb{C} \to \mathbb{C}$ be any given function. A perhaps natural method to extend such a function to a *diagonal* matrix is to simply apply the function separately to each of the diagonal entries. An advantage of this choice is that it keeps all power series representations of f valid as well.

However, once we have made this choice for diagonal matrices, it extends naturally to all diagonalizable, and in particular, Hermitian matrices. In particular, if $A = UDU^{-1}$, where D is diagonal (and therefore consists of the eigenvalues of A), we define $f(A) := Uf(D)U^{-1}$, where $f(D)$ is defined by applying f separately to each non-diagonal entry. Note that this definition agrees with the usual definition of A^n when n is an integer.

Our goal now is to understand the monotonicity and convexity properties of such functions. We first study the map $A \mapsto A^{-1}$.

Corollary 3.1. *Let* M *and* N *be PD matrices such that* $M \geq N$. *Then,* $N^{-1} \geq M^{-1}$. *When* M *and* N *are PSD with identical ranges,* $M \geq N$ *implies* $N^\dagger \geq M^\dagger$.

Proof. Note that when S is PD, $S^\dagger = S^{-1}$. The first claim now follows from the equivalence of items 2 and 3 of Proposition 2.4, applied with $X = I$, $A = M$ and $B = N^{-1}$. For the second claim, we recall that for a Hermitian matrix S, $S^\dagger S = SS^\dagger$ is an orthogonal projection onto the range of S. Since M and N have identical ranges we let P denote the orthogonal projection onto this common range subspace. The second claim now follows from the equivalence of items 2 and 3 of Proposition 2.3, applied with $X = P$, $A = M$ and $B = N^\dagger$. □

Corollary 3.2. *The map* $A \mapsto A^{-1}$ *is convex on the set of PD matrices, in the sense that for* $s \in [0, 1]$, *and PD matrices* A *and* B,

$$(sA + (1-s)B)^{-1} \leq sA^{-1} + (1-s)B^{-1}.$$

Proof. From items 1 and 2 of Proposition 2.3, we see that the matrices

$$M := \begin{bmatrix} A & I \\ I & A^{-1} \end{bmatrix} \text{ and } N := \begin{bmatrix} B & I \\ I & B^{-1} \end{bmatrix}$$

are PSD. This implies that the matrix

$$sM + (1-s)N = \begin{bmatrix} sA + (1-s)B & I \\ I & sA^{-1} + (1-s)B^{-1} \end{bmatrix}$$

is PSD as well. Applying items 1 and 3 of Proposition 2.3, we then see that

$$sA^{-1} + (1-s)B^{-1} \geq (sA + (1-s)B)^{-1}.$$

□

We now consider the map $A \mapsto A^{1/2}$ defined for positive semi-definite matrices.

Proposition 3.3. *Let* A *and* B *be PSD matrices such that* $B \leq A$. *Then* $B^{1/2} \leq A^{1/2}$.

Proof. Assume first that A and B are PD, and hence invertible. Then, by congruence, we have

$$B \preceq A \Leftrightarrow A^{-1/2}BA^{-1/2} \preceq I.$$

Now, $M = A^{-1/2}BA^{-1/2} = NN^*$ (where $N := A^{-1/2}B^{1/2}$) is a PD matrix (by congruence, since A and B are PD). Thus $M \preceq I$ implies that all eigenvalues of M are in the interval $(0, 1]$. This implies that all singular values of N are in the interval $(0, 1]$ (since these are square-roots of the eigenvalues of M).

Now, note that $N = A^{-1/4}WA^{1/4}$, where $W := A^{-1/4}B^{1/2}A^{-1/4}$ is a PD matrix (by congruence, since A and B are PD). Thus, N is diagonalizable with positive real eigenvalues. Since all the singular values of N lie in $(0, 1]$, it then follows that all its eigenvalues must also lie in the same interval. Since N and W have the same eigenvalues, the eigenvalues of W are therefore also in $(0, 1]$. Since W is Hermitian, this implies that $W = A^{-1/4}B^{1/2}A^{-1/4} \preceq I$. By congruence (since A and B are PD), this implies $B^{1/2} \preceq A^{1/2}$.

We have thus proved that whenever A and B are PD such that $A \succeq B$, we also have $A^{1/2} \succeq B^{1/2}$. When A and B are only PSD (so that they may not be invertible), we consider, for each $\epsilon > 0$, the matrices $A_\epsilon := A + \epsilon I$ and $B_\epsilon := B + \epsilon I$. Since these matrices are PD, we get $A_\epsilon \succeq B_\epsilon$ for each $\epsilon > 0$. The claim $A \succeq B$ then follows by taking limits on both sides of the inequality $A_\epsilon^{1/2} \succeq B_\epsilon^{1/2}$ as $\epsilon \downarrow 0$, and using the fact that the map $A \to A^{1/2}$ is continuous on the set of PSD matrices. □

Remark 3.1. It is a good exercise to check which step in the above proof fails for the map $A \mapsto A^2$. Indeed, the map $A \mapsto A^2$ is *not* monotone on PSD (or PD) matrices.

The above two monotonicity results are special cases of the Löwner-Heinz theorem (see Carlen [2010, Theorem 2.6]).

4. The matrix logarithm

In this section, we will prove that the logarithm is both monotone and concave on PD matrices. For this, it will be convenient to define, for every positive definite matrix, the following functions on the interval $[0, 1]$:

$$\begin{aligned} f_A(t) &:= \log(I + tA) \\ g_A(t) &:= A(I + tA)^{-1}. \end{aligned} \tag{2}$$

Note that

$$\frac{\mathrm{d}f_A(t)}{\mathrm{d}t} = g_A(t),$$

where the derivative of a matrix-valued function is defined entry-wise. Similarly, by entry-wise integration, we have

$$\log(I + A) = \log(I + A) - \log I = f_A(1) - f_A(0) = \int_0^1 g_A(t)dt.$$

From the above considerations, we get the following proposition:

Proposition 4.1. *Suppose that for given PSD matrices A, B such that $A \succeq B$, and for all $t \in [0, 1]$, $g_A(t) \succeq g_B(t)$. Then, we also have*

$$\log(I + A) \succeq \log(I + B).$$

Similarly, suppose that for PSD matrices A and B, $\lambda \in [0, 1]$, and all $t \in [0, 1]$, we have

$$g_{\lambda A+(1-\lambda)B}(t) \succeq \lambda g_A(t) + (1 - \lambda)g_B(t).$$

Then, we also have

$$\log(I + \lambda A + (1 - \lambda)B) \succeq \lambda \log(I + A) + (1 - \lambda)\log(I + B).$$

To apply this proposition, we now prove the required properties of the function g.

Proposition 4.2. *Let A, B be PSD matrices, and suppose $\lambda \in [0, 1]$. Then, for all $t \in [0, 1]$, we have the following:*

(1) $g_{\lambda A+(1-\lambda)B}(t) \geq \lambda g_A(t) + (1-\lambda) g_B(t)$.
(2) *if* $A \succeq B$ *then* $g_A(t) \succeq g_B(t)$.

Proof. Note that both items are trivially true when $t = 0$. For $t \in (0, 1]$, we use the representation

$$g_A(t) := A(I + tA)^{-1} = \frac{1}{t}(I + tA - I)(I + tA)^{-1} = \frac{1}{t}\left\{I - (I + tA)^{-1}\right\}.$$

Now, when $t \geq 0$ and $A \succeq B \succeq 0$, we have $I + tA \succeq I + tB \succ 0$. Since the map $A \mapsto A^{-1}$ is monotone decreasing for positive definite matrices, we get $(I + tA)^{-1} \preceq (I + tB)^{-1}$. This immediately gives

$$g_A(t) = \frac{1}{t}\left\{I - (I + tA)^{-1}\right\} \succeq \frac{1}{t}\left\{I - (I + tB)^{-1}\right\} = g_B(t),$$

which establishes the second item. For the first item, we use the convexity of the map $A \mapsto A^{-1}$ (Corollary 3.2) for PD matrices to get

$$(I + t(\lambda A + (1-\lambda)B))^{-1} = (\lambda(I + tA) + (1-\lambda)(I + tB))^{-1} \preceq \lambda(I + tA)^{-1} + (1-\lambda)(I + tB)^{-1}$$

This gives

$$\begin{aligned} g_{\lambda A+(1-\lambda)B}(t) = \frac{1}{t}\left(I - (I + t(\lambda A + (1-\lambda)B))^{-1}\right) &\succeq \frac{1}{t}\left(I - \lambda(I + tA)^{-1} - (1-\lambda)(I + tB)^{-1}\right) \\ &= \frac{1}{t}\left(\lambda(I - (I + tA)^{-1}) + (1-\lambda)(I - (I + tB)^{-1})\right) \\ &= \lambda g_A(t) + (1-\lambda) g_B(t), \end{aligned} \tag{3}$$

and this establishes the first item. □

Now, we can prove the required monotonicity and concavity properties of the matrix logarithm.

Theorem 4.3. *Let A, B be strictly positive definite matrices, and fix $\lambda \in [0, 1]$. Then, we have the following:*
(1) $\log(\lambda A + (1-\lambda)B) \succeq \lambda \log A + (1-\lambda) \log B$.
(2) *if* $A \succeq B$ *then* $\log A \succeq \log B$.

Proof. Assume first that $A, B \succeq I$. From Propositions 4.1 and 4.2, we then get

$$\log(\lambda A + (1-\lambda)B) = \log(I + \lambda(A - I) + (1-\lambda)(B - I)) \succeq \lambda \log A + (1-\lambda) \log B, \tag{4}$$

To prove this for general positive definite matrices, we note that when A and B are strictly positive definite, there exists a positive real c such that $cA, cB \succeq I$. Equation (4) applied to cA, cB then gives the first item of the theorem.

For the second item, we again consider first the case $A \succeq B \succeq I$. Again from Propositions 4.1 and 4.2, we then get

$$\log A = \log(I + (A - I)) \succeq \log(I + (B - I)) = \log B. \tag{5}$$

To prove this for general positive definite matrices, we again note that when A and B are strictly positive definite, there exists a positive real c such that $cA, cB \succeq I$. Equation (5) applied to cA, cB then gives the second item of the theorem. □

Remark 4.1. Bhatia [2007] proves the above properties of the logarithm as direct corollaries of the fact that for every $s \in (0, 1]$, the functions $A \mapsto A^s$ are monotone non-decreasing and concave over the set of PSD matrices. I recommend going through that approach as well; especially since it involves the interesting notion of the matrix geometric mean. We, however, will not be needing these ideas for our immediate goal of studying matrix concentration inequalities, and hence, I prefer the above, perhaps more direct, approach.

References

R. Bhatia. *Positive definite matrices*. Hindustan Book Agency and Princeton University Press, 2007.

E. Carlen. Trace inequalities and quantum entropy: An introductory course. In R. Sims and D. Ueltschi, editors, *Entropy and the Quantum*, volume 529 of *Contemporary Mathematics*, pages 73–140. American Mathematical Society, 2010. ISBN 978-0-8218-8208-5.

E. H. Lieb. Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Advances in Mathematics*, 11(3):267–288, Dec. 1973. ISSN 0001-8708. doi: 10.1016/0001-8708(73)90011-X. URL https://www.sciencedirect.com/science/article/pii/000187087390011X.