POSITIVE DEFINITE MATRIX - SOME BASICS

Paul Simanjuntak

24/10/2019

Note. This is a talk I gave in Graduate Analysis Seminar on 17 September 2019. The talk is taken mostly from "Matrix Analysis" by Horn and Johnson, Chapter 7.

We begin by the definition:

Definition 1. A Hermitian matrix $A \in M_n(\mathbb{R})$ is positive (semi)-definite if $x * Ax > 0$ (respectively, ≥ 0) for all $x \neq 0$.

Recall here that a matrix A is Hermitian if $A = A^*$, its conjugate transpose. Thus the notion of positive-definite matrix is well-defined, because if A is Hermitian, $x * Ax$ is real. In particular, this means all the eigenvalues of A are real. Of course in what follows, we will conveniently forget the distinction between the positive and non-negative definitions.
Let's begin with some basic properties of and observations about positive-definite matrices:

(a) **The eigenvalues of A are non-negative**

Let (λ, x) be an eigenpair of A, then

$$x^*Ax = x^*\lambda x = \lambda(x^*x)$$

and hence (because $x \neq 0$)

$$\lambda = \frac{x^*Ax}{x^*x} \geq 0$$

In particular, we can see that if $A \in M_n(\mathbb{C})$ is Hermitian, A is positive definite implies that all of its eigenvalues are non-negative. The converse is actually also true, which we can see from the next observation.

(b) **Positive definiteness is invariant under conjugation congruence**

Given $C \in M_n(\mathbb{R})$ and A to be any positive definite matrix, we claim that C^*AC is also positive definite. Indeed, let $\lambda_1, \cdots, \lambda_n$ be the eigenvalues of A and let D to be the diagonal matrix with the eigenvalues of A as its diagonal entries. Clearly D is positive definite. Then for some $P \in M_n(\mathbb{R})$, we have $A = P^*DP$. If Q is any other matrix, we have $Q^*AQ = (PQ)^*D(PQ)$. But the right side is positive definite (because conjugation preserves eigenvalue), so Q^*AQ also has only non-negative eigenvalues, i.e. it is positive definite.

(c) **Principals of positive definite matrix are also positive definite**

Recall that a principal of a matrix is obtained by deleting rows and columns from the original matrix. We will not prove this fact, but we'll note that it gives our first 'non-obvious' way to check for positive definiteness:

Proposition 2 (Sylvester's Criterion)**.** *If the leading principal minors (the determinants of the n corner principals) of A are positive, then A is positive definite.*

(d) **Positive definite matrices are like positive numbers**

This is the real reason we like positive definite matrices. Still, why we can say this? For one, it should be clear (from the definition) that given A and B positive definite matrices, $A + B = 0$ iff $A = B = 0$.

There's another less trivial reason why we can claim parallels with positive numbers. Let $A \in M_n(\mathbb{R})$ (or complex, but we'll keep it simple) be positive definite and let r to be the rank of A. Then we may write $A = UDU^*$, where U is an unitary matrix and $D = \operatorname{diag}(\lambda_1, \cdots, \lambda_r) \oplus 0_{n-r}$, where λ_i is an eigenvalue of A. Let $B = UD^{1/k}U^*$, where

$$D^{1/k} = \operatorname{diag}(\lambda_1^{1/k}, \cdots, \lambda_r^{1/k}) \oplus 0_{n-r}$$

Then $B^k = A$, so we'll say B is the k-th root of A and write $A = B^{1/k}$. Here we note two things. First, such B is well-defined, because the eigenvalues of A are non-negative (by positive definiteness). Second, such B is also unique: we can show there exists an unique polynomial $p(x) \in \mathbb{R}[x]$ so that $B = p(A)$. Moreover, this means B commutes with any matrix that commutes with A. Note here that if $B = A^{1/2}$, all these observations mean $A = B^*B$. This gives another characterization: A is positive definite iff it can be written as a product B^*B for some B.

Now the preliminary is out of the way, we may apply our knowledge about positive definite matrix. One of the most important application is the concept of polar and singular value decomposition of a matrix. The claim is as follows: given $A \in M_{n,m}(\mathbb{C})$

(i) If $m = n$, we may write $A = PU = UQ$, where P and Q are positive semi-definite and U is an unitary matrix. In fact, such P and Q are unique: the last comment on (d) gives $P = (AA^*)^{1/2}$ and $Q = (A^*A)^{1/2}$. If A is non-singular, the matrix U is also unique.

(ii) If $n < m$, we may write $A = PU$, with P and U as above.

(iii) If $n > m$, we may write $A = UQ$, with Q and U as above.

Where is the parallel with 'regular' number here? We may think of any matrix as a complex number z. We know z has a polar representation $r\,e^{i\theta}$, where $r \geq 0$ and $e^{i\theta}$ has modulus 1. Similarly here, we can draw an analogy with A, P (or Q), and U corresponding to z, r, and $e^{i\theta}$ (for Q, this analogy works because $\det Q = 1$).
The proof of this fact is simple. The main idea uses singular value decomposition (SVD) of matrices. Given $A \in M_{n,m}$, let $q = \min\{m, n\}$ and σ_i to be the i-th eigenvalue of A^*A (arranged in descending order). Then by SVD, there exists unitary matrices $V \in M_{n,q}$ and $W \in M_{q,m}$ so that $A = V\Sigma W^*$, where $\Sigma = \operatorname{diag}(\lambda_1, \cdots, \lambda_q)$.
In this set up, we may then rewrite A as

$$A = V\Sigma W^* = \underbrace{(V\Sigma V^*)}_{\substack{\text{positive}\\ \text{definite}}} \underbrace{(VW^*)}_{\text{unitary}}$$

Alternatively, we take σ_i to be the i-th eigenvalue of AA^* and write $A = (VW^*)(W\Sigma W^*)$. This gives the necessary decomposition of A. Such V and W are not mysterious and we can prove they're actually unique. We can do it as such: if $n \geq m$, let $P = (A^*A)^{1/2}$, which by SVD can be written as $W\Sigma W^*$. Then there exists $V \in M_{n,m}$ with orthonormal columns so that $A = V\Sigma W^*$.
We may say a bit more about the singular values σ_i behaves in relation to the eigenvalues λ_i of A. For one, eventhough the σ_i don't always play well with λ_i (they can be totally different), we can find a matrix which eigenvalues are exactly the σ_i's: the eigenvalues of the matrix

$$\begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}$$

are precisely

$$-\sigma_1 \leq \cdots \leq -\sigma_q \leq \underbrace{0 = \cdots = 0}_{(n-m)-\text{times}} \leq \sigma_q \leq \cdots \leq \sigma_1$$

We also have a version of the Courant-Fisher min-max theorem (which gives analogous statement for the eigenvalues. Note that Courant-Fisher actually implies the one we're going to state):

$$\begin{aligned} \sigma_k^2(A) &= \min_{\langle w_1,\cdots,w_{k-1}\rangle} \max_{x\perp\langle w_1,\cdots,w_{k-1}\rangle} \frac{x^*A^*Ax}{x^*x} \\ &= \min_{\langle w_1,\cdots,w_{n-k}\rangle} \max_{x\perp\langle w_1,\cdots,w_{n-k}\rangle} \frac{x^*A^*Ax}{x^*x} \end{aligned}$$

In either case, we're optimizing the quantity $\dfrac{\|Ax\|_2}{\|x\|_2}$.
For this last part, we'll give an application of the singular values and positive definite matrices. We will need the following theorem, which we will not prove

Theorem 3. *Given $A, B \in M_{m,n}(\mathbb{R})$. Let $q = \min\{m, n\}$ and $\sigma_i(A)$, $\sigma_i(B)$ to be the singular values of A and B, respectively, written in descending order. Then*

$$Re\, tr AB^* \leq \sum_{i=1}^{q} \sigma_i(A)\,\sigma_i(B)$$

We can use the theorem to show the following estimates:

Proposition 4. *Given $A, B \in M_{m,n}(\mathbb{R})$, $q = \min\{m, n\}$, and $\sigma_i(A), \sigma_i(B)$ the singular values of A and B, respectively, written in descending order, we have*

i) $\|A - B\|_2^2 \geq \sum_{i=1}^{q}(\sigma_i(A) - \sigma_i(B))^2$

ii) $\sum_{i=1}^{q} \sigma_i(A) = \max_{\text{unitary } U\in M_p} Re\, tr(AU)$ *where* $p = \max\{m, n\}$.

iii) $\sum_{i=1}^{q} \sigma_i(AB^*) \leq \sum_{i=1}^{q} \sigma_i(A)\,\sigma_i(B)$.

Here $\|\cdot\|_2$ *is the usual Euclidean norm.*

Proof. We'll just prove i) and iii). The first one is just by calculation

$$\begin{aligned}
\|A-B\|_2^2 &= \langle A-B, A-B\rangle \\
&= \langle A, A\rangle - 2\mathrm{Re}\,\langle A, B^*\rangle + \langle B, B\rangle \\
&\geq \sum_{i=1}^{q} \sigma_i(A)^2 - 2\sum_{i=1}^{q} \sigma_i(A)\,\sigma_i(B^*) + \sum_{i=1}^{q} \sigma_i(B)^2 \\
&= \sum_{i=1}^{q} (\sigma_i(A) - \sigma_i(B))^2
\end{aligned}$$

where we used Theorem 3 and the fact $\langle A, A\rangle = \mathrm{tr}(A^*A)$.
For iii), let PU be the polar decomposition of AB^*. Again using the theorem,

$$\sum_{i=1}^{q} \sigma_i(AB^*) = \mathrm{tr}P = \mathrm{Re}\,\mathrm{tr}(AB^*U) \leq \sum_{i=1}^{q} \sigma_i(A)\,\sigma_i(U^*B) = \sum_{i=1}^{q} \sigma_i(A)\,\sigma_i(B)$$

□

We use this proposition to answer the following question: given a non-singular matrix A, what is the nearest singular matrix? Let B to be singular, so $\sigma_n(B) = 0$ (because B^*B is also singular, at least one of the singular value has to be zero, so the last one, σ_n, must be zero). Then by i), $\|A-B\|_2^2 \geq \sigma_n^2(A)$, so if the SVD of A is $V\Sigma W^*$, we may take $B_0 = V\Sigma_0 W^*$, where $\Sigma_0 = \mathrm{diag}\,(\sigma_i(A))_{i=1}^{n-1} \oplus 0$, so B_0 is a singular matrix so that the minimum distance to A (in Euclidean sense) is reached.
This method does not necessarily give an unique answer. If $\sigma_{n-1}(A) = \sigma_n(A)$, we may let $\widehat{\Sigma}_0 = (\sigma_1, \cdots, \sigma_{n-2}, 0, \sigma_n)$. Then with $C_0 = V\widehat{\Sigma}_0 W^*$, the Euclidean distance $\left\|A - C_0\right\|_2^2 \geq \sigma_{n-1}^2(A) = \sigma_n^2(A)$, so we get two different singular matrices with the same (minimum) distance from A.