MA 341: Matrix Analysis and Positivity

Naman Mishra

August 2025

Contents

Lectures

The Course

Resources:

Lecture 1.
Tuesday
August 5

(1) A. Khare. Matrix analysis and entrywise positivity preservers, 2022.

(2) Rajendra Bhatia. Matrix analysis.

(3) Rajendra Bhatia. Positive definite matrices.

Grading:

(50%) Homework + Midterm

(50%) Final presentation

Various notions of matrix positivity and maps preserving these structures.

We'll briefly see how GPS triangulation works, and also Heron's formula in n dimensions.

Over this course, we will only work over $\mathbb{R}$, and "positive" will often mean "non-negative".

Chapter I

Positivity

One easy way to generalize positivity (non-negativity) of real numbers to matrices is to consider diagonal matrices with non-negative entries.

More generally, we

Theorem I.1 (Spectral theorem)**.** *Each symmetric matrix $A \in \mathbb{R}^{n\times n}$ has an orthonormal eigenbasis, that is, there exist orthonormal vectors $u_1, \ldots, u_n$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ such that $Au_i = \lambda_i u_i$.*

We could write

$$A\begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} = \begin{pmatrix} \lambda_1 u_1 & \cdots & \lambda_n u_n \end{pmatrix} = \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix}\begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}.$$

Calling the matrix of eigenvectors U and the diagonal matrix of eigenvalues Λ, we can write $AU = U\Lambda$, so that $A = U\Lambda U^T$. Note that $U^{-1} = U^T$ since

$$\begin{pmatrix} u_1^T \\ \vdots \\ u_n^T \end{pmatrix}\begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} = (u_i^T u_j)_{ij} = \mathbb{I}_n$$

iff the u_i are orthonormal.

We call matrices of this form *positive semidefinite*. That is, a matrix A is positive semidefinite iff it can be expressed as $U\Lambda U^T$ where $U \in O(n)$ and Λ is a diagonal matrix with $\lambda_{ii} \geq 0$. The formal definition is different.

Definition I.2 (Positive semidefinite)**.** Let $A \in \mathbb{R}^{n\times n}$ and let $\kappa_A \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be the associated bilinear form $\kappa_A(v, w) = v^T A w$.

A is *positive semidefinite* iff $A = A^T$ and $\kappa_A(v, v) \geq 0$ for all $v \in \mathbb{R}^n$. We denote the set of all PSD matrices of dimension n by PSD_n.

A is *positive definite* iff $A = A^T$ and $\kappa_A(v, v) > 0$ for all $v \in \mathbb{R}^n \setminus \{0\}$. We denote the set of all PD matrices of dimension n by PD_n.

Remark. If $U = \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix}$ and $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ is a positive diagonal matrix, then

$$U \Lambda U^T = \sum_i \lambda_i u_i u_i^T$$

so $x^T(U\Lambda U^T)x = \sum_i \lambda_i \langle u_i,\, x\rangle^2 \geq 0$.

Thus the definition we discussed earliear is equivalent to the one in definition I.2.

We will over the course of the course prove most of the following.

Theorem I.3. *The following are equivalent for symmetric $A \in \mathbb{R}^{n\times n}$.*

(1) A is PSD.

(2) All eigenvalues of A are in $[0, \infty)$.

(3) $A = B^T B$ for some $B \in \mathbb{R}^{n\times n}$.

(4) A is a Gram matrix from $\mathbb{R}^n$.

(5) A is the covariance matrix of some data.

(6) A is the Cayley-Menger matrix of an $(n+1)$-point Euclidean metrix space $X \subseteq (\mathbb{R}^n, \|\cdot\|_2)$.

PSDs show up in many places, such as

- Nevanlinna-Pick condition,
- classifying Dynkin diagrams of something something,
- much earlier, using Hessians to find local minima.

Traditionally, authors like Rajendra Bhatia consider functions of the form $f(A) = f(U\Lambda U^T) = Uf(\Lambda)U^T$. Why?

- $A^2 = U\Lambda^2 U^T$. $A^3 = U\Lambda^3 U^T$, et cetera.
- For a polynomial $p(x) = \sum_{i=0}^d a_i x^i$, then $p(A) = Up(\Lambda)U^T$.

As long as the eigenvalues are compactly supported, polynomials give good approximations to all continuous functions. This study is called (holomorphic) functional calculus.

In this course, we will work with functions acting on the *entries* of the matrix, that is, functions that look like

$$f[A] = (f(a_{ij}))_{ij}.$$

For example, entrywise matrix multiplication of two PSD matrices is PSD, as we will see later. Entrywise calculus is not as well-developed.

Lecture 1

Examples.

- [(Toeplitz) cosine matrices] Let $\theta_1, \dots, \theta_2 \in \mathbb{R}$ and set $a_{ij} = \cos(\theta_i - \theta_j)$. Then $A = uu^T + vv^T$ where

  $$u = \begin{pmatrix} \cos\theta_1 \\ \dots \\ \cos\theta_n \end{pmatrix}, \qquad v = \begin{pmatrix} \sin\theta_1 \\ \dots \\ \sin\theta_n \end{pmatrix}.$$

  Thus A is symmetric and $x^T A x \geq 0$ for all x.

  A matrix is *Toeplitz* if a_{ij} depends only on $i - j$. That is, A looks like

  $$\begin{pmatrix} a_0 & a_{-1} & a_{-2} & a_{-3} \\ a_1 & a_0 & a_{-1} & a_{-2} \\ a_2 & a_1 & a_0 & a_{-1} \\ a_3 & a_2 & a_1 & a_0 \end{pmatrix}$$

  If $\theta_1, \dots, \theta_n$ are in arithmetic progression, then the cosine matrix discussed above is Toeplitz.

- A *Hankel* matrix is similar to a Toeplitz matrix, but constant along the anti-diagonals: something like

  $$\begin{pmatrix} a_0 & a_1 & a_2 & a_3 \\ a_1 & a_2 & a_3 & a_4 \\ a_2 & a_3 & a_4 & a_5 \\ a_3 & a_4 & a_5 & a_6 \end{pmatrix}.$$

  Choose $u_0 \in \mathbb{R}$ and define

  $$H_{u_0} := \begin{pmatrix} 1 & u_0 & u_0^2 & u_0^3 & \cdots \\ u_0 & u_0^2 & u_0^3 & u_0^4 & \cdots \\ u_0^2 & u_0^3 & u_0^4 & u_0^5 & \cdots \\ u_0^3 & u_0^4 & u_0^5 & u_0^6 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

  Claim. Any (finite) leading principal truncation of H_{u_0} is PSD.

  Proof. Let the truncation have size $n + 1$. Then

  $$\begin{pmatrix} 1 & u_0 & \cdots & u_0^n \\ u_0 & u_0^2 & \cdots & u_0^{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ u_0^n & u_0^{n+1} & \cdots & u_0^{2n} \end{pmatrix} = \begin{pmatrix} 1 \\ u_0 \\ \vdots \\ u_0^n \end{pmatrix} \begin{pmatrix} 1 & u_0 & \cdots & u_0^n \end{pmatrix}$$

  and hence is PSD. ∎

The *Dirac* δ measure at $u_0 \in \mathbb{R}$, denoted by $\delta_{u_0}(x) = \delta_{u_0,x}$ satisfies

$$\int_{\mathbb{R}} f d\delta_{u_0} = f(u_0).$$

The *k-th moment* of a measure $\mu \geq 0$ on $\mathbb{R}$ is

$$s_k(\mu) \coloneqq \int_{\mathbb{R}} x^k d\mu(x).$$

For example,

$$\begin{aligned} \text{mass of } \mu &= s_0(\mu) = \int_{\mathbb{R}} d\mu, \\ \text{mean of } \mu &= s_1(\mu)/s_0(\mu) \text{ (if } s_0(\mu) > 0), \\ \text{variance of } \mu &= s_2(\mu) - s_1(\mu)^2 \text{ (if } s_0(\mu) = 1). \end{aligned}$$

Lecture 2. Thursday August 7

Proposition I.4. *Let μ be a non-negative measure on $\mathbb{R}$ such that $s_k(\mu) = \int_{\mathbb{R}} x^k d\mu$ converges for all k. Then $H_\mu \coloneqq (s_{i+j}(\mu))_{i,j\geq 0}$ is PSD.*

Proof. Let

$$H_n = \begin{pmatrix} s_0 & \dots & s_n \\ \vdots & \ddots & \vdots \\ s_n & \dots & s_{2n} \end{pmatrix}.$$

We need to show that $u^T H_n u \geq 0$ for all u. Write

$$\begin{aligned} u^T H_n u &= \sum_{i,j=0}^{n} u_i s_{i+j}(\mu) u_j \\ &= \sum_{i,j} \int_{\mathbb{R}} u_i x^{i+j} u_j d\mu(x) \\ &= \int_{\mathbb{R}} \left(\sum_i u_i x^i \right)^2 d\mu(x) \geq 0. \end{aligned}$$

■

Examples.

- If $\mu = \delta_{-\pi} + 2\delta_3$, then $s_k(\mu) = (-\pi)^k + 2 \cdot 3^k$. Therefore,

$$\begin{pmatrix} 3 & 6-\pi & 18+\pi^2 & \dots \\ 6-\pi & 18+\pi^2 & \ddots & \dots \\ 18+\pi^2 & \ddots & \ddots & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

is PSD.

- *Gram matrices* from $\mathbb{R}^r$ are PSD. Given $x_1, \ldots, x_n \in \mathbb{R}^r$, their Gram matrix is
$$G(x_1, \ldots, x_n) \coloneqq (\langle x_i,\, x_j\rangle)_{i,j=1}^n$$

Proposition I.5. *The Gram matrix $G(x_1, \ldots, x_n)$ where $x_i \in \mathbb{R}^r$ has rank at most r.*

Proof. Given $u = (u_1, \ldots, u_n)^T \in \mathbb{R}^n$,
$$\begin{aligned} u^T G u &= \sum_{i,j=1}^n u_i \langle x_i,\, x_j\rangle u_j \\ &= \sum_{i,j} \langle u_i x_j,\, u_j x_j\rangle \\ &= \left\langle \sum_i u_i x_j,\, \sum_j u_j x_j \right\rangle \geq 0. \end{aligned}$$

To show that the rank is at most r, write $G = X^T X$ where $X = \begin{pmatrix} x_1 & \ldots & x_n \end{pmatrix}$. (This also immediately gives positive semidefiniteness.) Now
$$\operatorname{rank}(G) = \operatorname{rank}(X^T X) \leq \operatorname{rank}(X) \leq r.$$
This bound is tight when, for example, $x_i = e_i$. ■

Theorem I.6. *The following are equivalent for symmetric $A \in \mathbb{R}^{n\times n}$.*

(1) A is PD.

(2) All eigenvalues of A are in $(0, \infty)$.

(3) $A = B^T B$ for some $B \in \mathbb{R}^{n\times n}$ and A is full rank.

(4) A is a Gram matrix $G(x_1, \ldots, x_n)$ from $\mathbb{R}^n$ and x_i form a basis.

Proof of theorem I.3.

(1) $\implies$ **(2)** If $Av = \lambda v$ for a non-zero v, then $0 \leq v^T A v = \lambda \|v\|^2$ so $\lambda \geq 0$. If A is PD, then the inequality becomes strict.

(2) $\implies$ **(1)** By the spectral theorem, $A = U^T \Lambda U$ where $U \in O(n)$ and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. If λ_i are non-negative, then
$$u^T A u = v^T \Lambda v = \sum_i \lambda_i v_i^2 \geq 0$$
where $v = Uu$. If all λ_i are strictly positive, then the sum is zero iff $v = 0$.

(3) $\implies$ (1) If $A = B^T B$ then $u^T A u = \|Bu\|^2 \geq 0$. If A is full rank, then so is B. Thus $\|Bu\|^2 = 0$ iff $u = 0$.

(4) $\implies$ (3) In the proof of proposition I.5, we wrote $G = X^T X$.

If x_i form a basis, then X^T and X are both full rank matrices, hence so is their product.

(2) $\implies$ (4) By the spectral theorem, $A = U^T \Lambda U$. Since all eigenvalues are non-negative, $A = (\sqrt{\Lambda}U)^T(\sqrt{\Lambda}U) =: X^T X$ is a Gram matrix.

If all eigenvalues are positive, X has full rank, so its columns form a basis.

To recap, we proved

$$(1) \iff (2) \implies (4) \implies (3) \implies (1). \qquad \blacksquare$$

Corollary I.7. *Given any real symmetric matrix $A_{n\times n}$, the matrix $A - \lambda_{min}(A)\mathbb{I}_n$ is PSD.*

A *cone* in a vector space is a set S such that $cS = S$ for all $c > 0$.

Corollary I.8. PD_n *is dense in* PSD_n*. Moreover, both of these are convex cones in* $\mathbb{R}^{n^2}$*.*

Proof. Any matrix $A \in \mathsf{PSD}_n \setminus \mathsf{PD}_n$ can be approximated by the sequence of matrices $(A + \frac{1}{k}\mathbb{I}_n)_{k\geq 1}$. Each of these is in PD_n. This can be seen by definition and also by the eigenvalue characterization.

Any positive scaling of a PSD (resp. PD) matrix is PSD (resp. PD): PSD_n and PD_n are cones. Sums of PSD (resp. PD) matrices are also PSD (resp. PD). Thus the cones are convex. $\blacksquare$

Definition I.9. Given $A \in \mathbb{R}^{m\times n}$ and subsets $I \subseteq [m]$ and $J \subseteq [n]$, define $A_{I\times J}$ to be the submatrix of A whose rows and columns are indexed by I and J respectively. (The indices come from I and J. A if a function from $I \times J$ to $\mathbb{R}$.) $A_{I\times J}$ is *principal* if $I = J$

A *minor* of A is $\det(A_{I\times J})$ for some subsets I, J of equal sizes of $[m]$ and $[n]$ respectively. It is *principal* if $I = J$.

Theorem I.10 (Sylvester's criterion)**.** *A real symmetric matrix A is PSD (resp. PD) iff all its principal minors are non-negative (resp. positive).*

Lecture 3.
Tuesday
August 12

Lecture 3

Lemma I.11. *Any two norms on $\mathbb{R}^n$ are equivalent. That is, if $\|\cdot\|_1$ and $\|\cdot\|_2$ are norms on $\mathbb{R}^n$, then there exist $0 < m < M$ such that*

$$B_1(0;1) \subseteq B_2(0,m) \text{ and } B_2(0;1) \subseteq B_1(0,M).$$

Proof. Norms are continuous , so they achieve a maximum and minimum on the surface on the unit ball. ■ how?

Lemma I.12. *If A is PSD (resp. PD), then so are all the principal submatrices.*

Proof. Let A be PSD (resp. PD) Fix $J \subseteq [n]$. We have to show that $A_{J\times J}$ is PSD (resp. PD).

For any non-zero $u \in \mathbb{R}^J$, fill it with zeroes so that we get $v \in \mathbb{R}^n$ with

$$v_i = \begin{cases} u_i & \text{if } i \in J, \\ 0 & \text{otherwise.} \end{cases}$$

Then $u^T A_{J\times J} u = v^T A v \geq 0$ (resp. > 0). ■

Lemma I.13. *If all the principal minors of symmetric $A \in \mathbb{R}^{n\times n}$ are strictly positive, then A is PD.*

Proof. Induction. Trivial for $n = 1$.

Fix $n > 1$, $A = A^T \in \mathbb{R}^{n\times n}$ and assume that the statement holds for all smaller n. Again write $A = U^T \Lambda U$.

Claim. *At least $n-1$ eigenvalues of A are positive.*

Proof. Suppose for the sake of contradiction that $\lambda_1, \lambda_2 \leq 0$. Then for all $w \in \operatorname{span}(u_1, u_2)$, $w^T A w \leq 0$. That is $A|_{\operatorname{span}(u_1,u_2)}$ is negative semidefinite (that is, its negative is PSD). On the other hand, for all non-zero $x \in \mathbb{R}^{n-1}$, we have

$$\begin{pmatrix} x \\ 0 \end{pmatrix}^T A \begin{pmatrix} x \\ 0 \end{pmatrix} = x^T A_{[n-1]\times[n-1]} x.$$

Each principal minor of $A_{[n-1]\times[n-1]}$ is positive. By the induction hypothesis, this is PD. Thus there is an $(n-1)$-dimensional subspace where the quadratic form is positive, and a 2-dimensional space where it is non-positive. Contradiction. □

Note that since $\det(A) > 0$, no eigenvalue is zero. Since at least $n-1$ of the eigenvalues are positive, the last one must also be. ■

Definition I.14 (Adjugate matrix)**.** Given $A \in \mathbb{R}^{n\times n}$, its *adjugate matrix* is $\operatorname{adj}(A) := (\alpha_{ij})^T$, where

$$\alpha_{ij} := (-1)^{i+j} \det(A_{([n]\setminus\{i\})\times([n]\setminus\{j\})})$$

We will henceforth write $A_{([n]\setminus\{i\})\times([n]\setminus\{j\})}$ as $A_{i^c\times j^c}$.

Fact I.15 (Jacobi's formula)**.** $A\,\mathrm{adj}(A)=\mathrm{adj}(A)A=(\det A)\mathbb{I}_n$.

Theorem I.16. *Let $A(t)=(a_{ij}(t))\colon\mathbb{R}\to\mathbb{R}^{n\times n}$ be differentiable. Then*

$$\frac{\mathrm{d}}{\mathrm{d}t}(\det A(t))=\mathrm{trace}\left(\mathrm{adj}\,A(t)\frac{\mathrm{d}A(t)}{\mathrm{d}t}\right)$$

Proof. Let $A,B\in\mathbb{R}^{n\times n}$. Let $f(\varepsilon)=\det(A+\varepsilon B)-\det A$. This is a polynomial in ε which vanishes at 0, and $f'(0)$ is just the coefficient linear term.

Now

$$\det(A+\varepsilon B)=\sum_{\sigma\in S_n}(-1)^\sigma\prod_{k=1}^n(a_{k\sigma_k}+\varepsilon b_{k\sigma_k})$$

has $n!n$ terms which are linear in ε. We get that the coefficient of εb_{ij} is

$$\sum_{\sigma:\sigma_i=j}(-1)^\sigma\prod_{k\neq i}a_{k\sigma_k}.$$

This is the same as $\det A_{i^c\times j^c}$ upto the sign. The sign $(-1)^\sigma$ must be replaced by $(-1)^{\sigma+(n-i)+(n-j)}=(-1)^{\sigma+i+j}$. In conclusion, the linear term is

$$\sum_{i,j=1}^n(-1)^{i+j}\det(A_{i^c\times j^c})b_{ij}.$$

Noting that $\sum_{ij}c_{ji}d_{ij}=\mathrm{trace}(CD)$, we have the the linear term is $\mathrm{trace}(\mathrm{adj}(A)B)$.

Now, since $A(t)$ is differentiable,

$$A(t+\varepsilon)=A(t)+\varepsilon A'(t)+o(\varepsilon).$$

Thus

$$\frac{\mathrm{d}}{\mathrm{d}t}\det A(t)=\mathbf{WTF}$$

■

Proof of theorem I.10. Lemma I.12 immediately proves the backward direction by expressing the determinant as the product of eigenvalues.

Lemma I.12 gives the forward result for PD matrices. We again induct for the PSD case. $n=1$ is trivial.

Fix $n>1$, $A=A^T\in\mathbb{R}^{n\times n}$ and assume that all principal minors are non-negative. Fix $J\subseteq[n]$ and define $f_J(t):=\det(A_{J\times J}+t\mathbb{I}_J)$. $f_J(0)$ is a principal minor and hence non-negative. By theorem I.16,

$$f_J'(2)=\mathrm{trace}(\mathrm{adj}(A_{J\times J}+2\mathbb{I}_J)\cdot 1).$$

But any diagonal entry of got lost

Fix $t > 0$. Then $f'_J(t) > 0$ and therefore $f_J(t) > 0$ for all $J \subseteq [n]$. Thus $\det(A + t\mathbb{I})_{J\times J} > 0$ for all J. By Lemma I.12, $A + t\mathbb{I}$ is PD. As eigenvalues are continuous functions, letting $t \to 0^+$ gives that A is PSD. ■