MTHSc 853: Linear Algebra

10. Positive definite matrices

Def: A self-adjoint map $H: X \to X$ is <u>positive</u> (or positive definite) if $(x, Hx) > 0$ for all $x \neq 0$. It is <u>nonnegative</u> (or positive semidefinite) if $(x, Hx) \geq 0$

We denote these as e.g., $H > 0$, and $H \geqslant 0$, resp.

<u>Theorem 10.1:</u> Let X be a Euclidean space.

(i) The identity map I is positive

(ii) If $M, N > 0$ then $M + N > 0$ and $aM > 0$ for $a > 0$.

(iii) If $H > 0$ and Q invertible, then $Q^* H Q > 0$.

(iv) $H > 0$ iff all eigenvalues are positive

(v) If $H > 0$ then H is invertible.

(vi) Every positive map has a unique positive square root.

(vii) The set of positive maps is an open subset of the space of self-adjoint maps.

(viii) The boundary points of the set of positive maps are the nonnegative maps that are not positive.

2

Proof: (i) $(x, Ix) = (x,x) > 0$ if $x \neq 0$. ✓

(ii) $(x, (M+N)x) = (x, Mx) + (x, Nx) > 0$ if $x \neq 0$. ✓

and $(x, aMx) = a(x, Mx) > 0$ ✓

(iii) $(x, Q^* H Q x) = (Qx, HQx)$. Put $y = Qx \neq 0$ if $x \neq 0$.

$= (y, Hy) > 0$. ✓

(iv) $(\Rightarrow)$ Say $H > 0$ and $Hv = \lambda v$, $v \neq 0$.

Then $(v, Hv) = (v, \lambda v) = \lambda (v,v) > 0 \Rightarrow \lambda > 0$. ✓

$(\Leftarrow)$ Let $v_1, \ldots, v_n$ be an orthonormal basis of eigenvectors with eigenvalues $\lambda_1, \ldots, \lambda_n > 0$. Pick $x \neq 0$, write $x = a_1 v_1 + \cdots + a_n v_n$, and $Hx = a_1 \lambda_1 v_1 + \cdots + a_n \lambda_n v_n$.

$(x, Hx) = \sum_{i=1}^{n} \lambda_i |a_i|^2 > 0$ for all $x \neq 0$. ✓

Note: $(x, Hx) = \sum_{i=1}^{n} \lambda_i |a_i|^2 \geq \lambda_{min} \|x\|^2$ $(*)$.

(v) If H is not invertible, then $Hu = 0$ for some $v \neq 0 \Rightarrow \lambda = 0$ ↯.

(vi) Write $\sqrt{H}\, x = \sum_{i=1}^{n} a_i \sqrt{\lambda_i}\, v_i$ (clearly positive.) ✓

(vii) Fix $H > 0$, and let N be any self-adjoint map such that $\|N - H\| < \lambda_{min}$.

Claim: N is positive.

Put $M = N - H$.

Since $\|M\| < \lambda_{min}$, we know $\|Mx\| < \lambda_{min}\|x\|$ for all $x \neq 0$.

By Cauchy-Schwarz, $|(x, Mx)| \leq \|x\| \, \|Mx\| < \lambda_{min}\|x\|^2$

Together, we get for $x \neq 0$:

$$(x, Nx) = (x, (H+M)x) = \underbrace{(x, Hx)}_{\geqslant \lambda_{min}\|x\| \text{ by } (*)} + \overbrace{(x, Mx)}^{< \lambda_{min}\|x\|, \text{ see above}} > \lambda_{min}\|x\|^2 - \lambda_{min}\|x\|^2 = 0.$$

Therefore, $N = H + M$ is positive. ✓

(viii) By definition, if K is on the boundary of positive maps, then $\exists$ sequence $H_n \longrightarrow K$ of positive maps, i.e., $\|H_n - K\| \to 0$.

By Cauchy-Schwarz, $\lim_{n\to\infty} (x, H_n x) = (x, Kx)$. ✓

□

Put a <u>partial order</u> onto the set of <u>self-adjoint maps</u>:

Say $M < N$ iff $N - M > 0$

$M \leq N$ iff $N - M \geqslant 0$.

We get the following properties (almost) for free:

Additive: $M_1 < N_1$ and $M_2 < N_2 \Rightarrow M_1 + M_2 < N_1 + N_2$

Transitive: $L < M < N \Rightarrow L < N$

Multiplicative: $M < N$, Q invt $\Rightarrow Q^* M Q < Q^* N Q$.

4

Theorem 10.2: Suppose $0 < M < N$. Then $M^{-1} > N^{-1} > 0$.

Proof: First, suppose $N = I$.

Then $M < I \Rightarrow I - M > 0$.

$\Rightarrow$ Eigenvalues of $I-M$ are positive.

Say $(I-M)v = \lambda v$ for $v \neq 0$.

Then $Mv = v - \lambda v = (1-\lambda)v$. Since $1-\lambda > 0$, $0 < \lambda < 1$.

In summary: λ eigenvalue of $I-M \Rightarrow 0 < \lambda < 1$

$\Rightarrow 1-\lambda$ eigenvalue of M, and $0 < 1-\lambda < 1$.

$\Rightarrow \frac{1}{1-\lambda}$ eigenvalue of M^{-1}, and $\frac{1}{1-\lambda} > 1$

$\Rightarrow$ Eigenvalues of $M^{-1} - I > 0$

$\Rightarrow M^{-1} - I > 0 \Rightarrow M^{-1} > I$. ✓

Now, consider arbitrary $N > M > 0$.

Factor $N = R^2$, $R > 0$ and invertible.

We'll use $0 < M < N \Rightarrow 0 < Q^*MQ < Q^*NQ$ twice.

First: $Q = R^{-1}$ (so $Q^* = R^{-1}$)

$0 < R^{-1}MR^{-1} < R^{-1}NR^{-1} = I$

From what we showed above, $(R^{-1}MR^{-1})^{-1} = RM^{-1}R > I$.

Second: $Q = R^{-1}$

$0 < I < RM^{-1}R \Rightarrow 0 < R^{-1}IR^{-1} < R^{-1}(RM^{-1}R)R^{-1}$

$\Rightarrow 0 < R^{-2} = N^{-1} < M^{-1}$. $\square$

<u>Caveat</u>: The product of self-adjoint maps is in general, not self-adjoint.

<u>Example</u>: Let $A = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}$, $B = \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix}$, $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

$Ax = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $Bx = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$, $(x, ABx) = (Ax, Bx) = -3$

<u>Def</u>: If A, B are self-adjoint, define their <u>symmetrized product</u> as $S = AB + BA$.

Note that $(x, Sx) = (x, ABx) + (x, BAx) = (Ax, Bx) + (Bx, Ax)$.

In the real case: $(x, Sx) = 2(Ax, Bx)$.

In the example above, $AB + BA = \begin{pmatrix} -6 & 0 \\ 0 & 42 \end{pmatrix}$.

Thus, it is <u>false</u> that $A > 0$, $B > 0 \implies AB + BA > 0$.

But a similar statement is true:

<u>Theorem 10.3</u>: Let A, B be self-adjoint. If $A > 0$ and $AB + BA > 0$, then $B > 0$.

<u>Proof</u>: Define $B(t) = B + tA$. (<u>Note</u>: We must show $B(0) > 0$.)

<u>Claim 1</u>: The symmetized product of A and $B(t)$ is positive for $t \geq 0$.

$S(t) = AB(t) + B(t)A = A(B + tA) + (B + tA)A = AB + BA + 2tA^2 = S + 2tA^2 > 0$

(since $S > 0$ and $2tA^2 \geq 0$) ✓

Claim 2: For t large enough, $B(t) > 0$:

$$(x, B(t)x) = (x, (B + tA)x) = (x, Bx) + t(x, Ax).$$

Recall: $(x, Ax) \geq \lambda_{min} \|x\|^2$ (min e-value of A)

and $|(x, Bx)| \leq \|x\| \, \|Bx\| \leq \|B\| \, \|x\|^2$

Together, $(x, B(t)) = t(x, Ax) + (x, Bx)$

$$\geq t\lambda_{min}\|x\|^2 - \|B\|\|x\|^2 = (t\lambda_{min} - \|B\|)\|x\|^2$$

Thus, if $t > \|B\| / \lambda_{min}$, $B(t) > 0$.

Claim 3: $B = B(0) > 0$.

If not, then for some t_0, $0 \leq t_0 \leq \|B\|/\lambda_{min}$ such that

$B(t_0)$ is on the boundary of the set of positive maps, i.e, $B(t_0) \geq 0$ but $B(t_0) \not> 0$.

Such a $B(t_0)$ has a $\lambda = 0$, so $B(t_0)y = 0$ for some $y \neq 0$.

However, $(y, S(t_0)y) = (Ay, B(t_0)y) + (B(t_0)y, Ay) = 0$. $\lightning$

Thus $B > 0$.

Corollary 10.4: If $0 < M < N$, then $0 < \sqrt{M} < \sqrt{N}$.

Proof. Put $A(t) = M + t(N - M)$.

For $0 \leq t \leq 1$, $A(t) = (1-t)M + tN > 0$, $\mathring{A}(t) = N - M > 0$.

Thus we can define $R(t) = \sqrt{A(t)}$ for $0 \leq t \leq 1$.

Since $A = R^2$, $\dot{A} = \dot{R}R + R\dot{R}$ (Symmetrized product of R and $\dot{R}$).

We know $\dot{A} = N - M > 0$ and thus so is $\dot{R} > 0$. (on $[0,1]$)

<u>Claim</u>: $R(t)$ is an <u>increasing</u> function on $[0,1]$.

<u>Pf</u>: For any $x \neq 0$, $\frac{d}{dt}(x, Rx) = (x, \dot{R}x) > 0$.

By calculus, $(x, R(t)x)$ is increasing

$\Rightarrow (x, R(s)x) < (x, R(t)x)$ for $s < t$

$\Rightarrow R(t) - R(s) > 0 \Rightarrow R(t) > R(s)$ ✓

In particular, $R(0) < R(1)$

$R(0) = \sqrt{A(0)} = \sqrt{M}$, $R(1) = \sqrt{A(1)} = \sqrt{N} \Rightarrow \sqrt{M} < \sqrt{N}$. □

<u>Def</u>: A real-valued function $f(s)$, $s > 0$ is a <u>monotone matrix function</u> (<u>mmf</u>) if for all self-adjoint mappings,

$0 < M < N \Rightarrow f(M) < f(N)$.

Recall that for H self-adjoint with eigenvalues $\lambda_1, \dots, \lambda_n$ and orthonormal eigenvectors $v_1, \dots, v_n$, we define

$F(H) = \hat{\sum_{i=1}} f(\lambda_i) P_i$ for a spectral resolution $H = \hat{\sum_{i=1}} \lambda_i P_i$.

<u>Examples</u>:

(i) $f(s) = -\frac{1}{s}$ is a mmf (immediate from Theorem 10.2;

$0 < M < N \Rightarrow M^{-1} > N^{-1} > 0 \Rightarrow -M^{-1} < -N^{-1} < 0$.

(Note that $f(M) = -M^{-1}$).

(ii) $f(s) = s^{1/2}$ is a mmf (immediate from Corollary 10.4;

$0 < M < N \implies 0 < \sqrt{M} < \sqrt{N}$. Note that $f(M) = \sqrt{M}$.)

(iii) $f(s) = s^2$ is <u>not</u> a mmf

Take any $A, B > 0$ with $S = AB + BA \not> 0$.

<u>Claim</u>: For small t, if $M = A$, $N = A + tB$, then

$0 < M < N$ but $M^2 \not< N^2$.

Note that $N^2 = A^2 + \underbrace{t(AB + BA)}_{\text{not negligible}} + \underbrace{t^2 B}_{\text{negligible for } t \approx 0}$

so $N^2 = M^2 + [\text{something nonpositive}] \not> M^2$.

(iv) $f(s) = s^{(-2^{4})}$ and $f(s) = \log s$ are mmf. (Exercise)

Additionally, positive multiples, sums, and limits of mmf's are mmf's.

For example, $-\sum \frac{m_j}{s + t_j}$ $m_j > 0$, $t_j > 0$ is an mmf,

and so is $f(s) = as + b - \int_0^\infty \frac{dm(t)}{s+t}$ $a > 0$, $b \in \mathbb{R}$, $(**)$

and $m(t)$ non-negative measure for which the integral converges.

In fact, <u>every</u> mmf has the form of $(**)$.

(Theorem of C. Loewner - very nontrivial!)

Surprisingly, functions of the form (**) are easy to characterize:

<u>Theorem</u> (Herglotz, Riesz): Every function f which is analytic on the upper half-plane with $\mathrm{Im}(f) > 0$ there, and $\mathrm{Im}(f) = 0$ on the positive real axis, has the form (**).

Conversely, <u>every</u> function of the form (**) can be extended to be analytic on the upper half-plane, with $\mathrm{Im}(f) > 0$ there.

<u>Proof</u>: See "Functional Analysis" by Peter Lax.

How to construct (all) positive matrices:

<u>Def</u>: Let $f_1, \dots, f_m$ be a sequence of vectors in a Euclidean space X. Define the $m \times m$ matrix G, where $G_{ij} = (f_i, f_j)$.

This is called the <u>Gram matrix</u> of $f_1, \dots, f_m$.

<u>Theorem 10.5</u>:

(i) Every Gram matrix is nonnegative.

(ii) The Gram matrix of a set of linearly independent vectors is positive

(iii) Every positive matrix can be represented as a Gram matrix.

Proof:

(i), (ii): $(x, Gx) = \sum_{i,j} x_i \bar{G}_{ij} \bar{x}_j = \sum_{ij} (f_i, f_j) x_i \bar{x}_j$

$$= \left(\sum_{i=1}^{m} x_i f_i , \sum_{j=1}^{m} x_j f_j \right) = \left\| \sum x_i f_i \right\|^2 .$$

(iii) Let $H = (h_{ij})$ be positive, and define the nonstandard inner product by $(x, y)_H = (x, Hy)$.

Note that under this inner product, the Gram matrix of the unit vectors $e_1, \dots, e_m$ has ij-entry

$(e_i, e_j)_H = (e_i, He_j) = h_{ij}$. $\square$

Examples:

(i) Let $X = \{ f : [0,1] \to \mathbb{R} \}$, $(f, g) := \int_0^1 f(t)\, g(t)\, dt$.

If $f_1 = 0$, $f_2 = t, \dots, f_i = t^{i-1}$, then

$G = (G_{ij})$ where $G_{ij} = \dfrac{1}{i+j-1}$.

(ii) Define $(f, g) = \int_0^{2\pi} f(\theta)\, \bar{g}(\theta)\, w(\theta)\, d\theta$, $w : \mathbb{R} \to \mathbb{R}^+$.

If $f_j = e^{ij\theta}$, $j = -n, \dots, n$, the the $(2n+1) \times (2n+1)$ corresponding Gram matrix is $G_{kj} = c_{k-j}$, where

$$c_p = \int w(\theta) e^{-ip\theta}\, d\theta .$$

Theorem 10.6 (Schur): Let $A = (A_{ij})$ and $B = (B_{ij})$ be positive matrices. Then $M = (M_{ij}) := (A_{ij} B_{ij})$ is positive.

We need to employ tensor products for the proof.

Given two vector spaces U, V, their tensor product, denoted $U \otimes V$, is a related vector space of dimension $(\dim U)(\dim V)$.

Definition 1: If $\{e_i\}_{i=1}^{n}$ is a basis of U and $\{f_j\}_{j=1}^{m}$ a basis of V, then $\{e_i \otimes f_j\}$ is a basis of $U \otimes V$.

Analogy: $\langle 1, x, \ldots, x^n \rangle$, $\langle 1, y, \ldots, y^m \rangle$ vs. $\langle x^i y^j : i \leq n, j \leq m \rangle$.

There's a better, basis-free way to construct $U \otimes V$.

Definition 2: $U \otimes V = \{ \sum u_i \otimes v_i : u_i \in U, v_i \in V \} / N$

where
$$N = \langle (u_1 + u_2) \otimes v - u_1 \otimes v - u_2 \otimes v_1, \; u \otimes (v_1 + v_2) - u \otimes v_1 - u \otimes v_2 \rangle$$

Basically, we are forcing the distributive law, i.e.

$(u_1 + u_2) \otimes v = u_1 \otimes v + u_2 \otimes v$, etc.

Theorem 10.7: There is a natural isomorphism

$$U \otimes V \longrightarrow \mathcal{L}(U', V).$$

<u>Proof</u>: (sketch). Define the map as follows:

$$U \otimes V \longrightarrow \mathcal{L}(U', V)$$

$$u \otimes v \longmapsto \{ l \mapsto (l, u) v \} \quad \text{and extend linearly.}$$

$$\begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \otimes \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \longmapsto \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} = \begin{bmatrix} v_1 u_1 & v_1 u_2 & \cdots & v_1 u_n \\ \vdots & \vdots & \ddots & \vdots \\ v_m u_1 & v_m u_2 & \cdots & v_m u_n \end{bmatrix}_{m \times n}$$

Note that $(l, u) v$ is:

$$\begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} \begin{bmatrix} l_1 \\ \vdots \\ l_n \end{bmatrix} \qquad \mathbb{R} \xrightarrow{l} \mathbb{R}^n \xrightarrow{u} \mathbb{R} \xrightarrow{v} \mathbb{R}^m$$

<u>Analogy</u>: Let $U = \langle 1, x, \dots, x^n \rangle$, $V = \langle 1, y, \dots, y^m \rangle$.

Think of $U \times V$ as $\langle (x^i, y^j) : 0 \le i \le n,\ 0 \le j \le m \rangle$

and $U \otimes V$ as $\langle x^i y^j : 0 \le i \le n,\ 0 \le j \le m \rangle$

<u>Remark</u>: We could similarly define an isomorphism $U \otimes V \to \mathcal{L}(V', U)$, where the dual of $L : U' \to V$ is a map $L' : V' \to U$.

If U, V are Euclidean spaces (so $U' = U$), there is a natural way to endow $U \otimes V$ with a Euclidean structure:

For $M, L \in \mathcal{L}(U, V)$, define $(M, L) = \operatorname{tr}(L^* M) = \sum_{i,j} l_{ji} m_{ji}$

Note that $\|M\|^2 = (M,M) = \sum_{i,j} m_{ji}^2$

So $e_i \otimes e_j$ under the isomorphism is the matrix E_{ij} (i.e, the ij-entry is 1, all others 0).

Clearly these form an orthonormal basis of $U \otimes V$.

<u>Proof of Thm 10.6</u> (Schur): $(A_{ij}), (B_{ij}) > 0 \Rightarrow M := (A_{ij} B_{ij}) > 0$.

Since every positive matrix can be written as a Gram matrix, write $A_{ij} = (u_i, u_j)$, $B_{ij} = (v_i, v_j)$ where $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are linearly independent sets.

Define $g_i \in U \otimes V$ as $g_i = u_i \otimes v_i$

Note that $(g_i, g_j) = (u_i, u_j)(v_i, v_j) = A_{ij} B_{ij}$.

Hence, M is a Gram matrix, and positive by Theorem 10.5. $\square$

<u>Another way to view tensor products</u>:

Let X be an n-dimensional real vector space

Note that $\mathbb{C}$ is a 2-dimensional real vector space ($\{1, i\}$ is a basis.

Suppose $A : X \to X$ is a linear map with minimum polynomial $A(s) = s^2 + 1$.

Then i and $-i$ are eigenvalues of A, but $i \notin \mathbb{R}$.

So if v is an eigenvector with eigenvalue $\lambda = i$, $v \notin X$.

However, v should live in some "extension" of X.

In this bigger vector space, we want to have vectors

like zv, $z \in \mathbb{C}$, $v \in X$.

What we really want is $\mathbb{C} \otimes X$.

This has basis $\{x_1, \ldots, x_n, ix_1, \ldots, ix_n\}$, where $x_1, \ldots, x_n$

is a basis of X.

Note that we need certain associativity, such as

$$(3i)v = (i3)v = i(3v).$$

i.e, $3i \otimes v = i \otimes 3v$.

But this comes for free with the construction!