Lecture 11 :- Basic properties of PSD matrices.

Definition :- A Symmetric matrix is a matrix A s.t. $A = A^T$. A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is positive semi-definite (PSD) if $v^T A v \geq 0 \quad \forall v \in \mathbb{R}^n$.

Notation :- $A \succeq 0$.

[**Note** :- A matrix $A \in \mathbb{C}^{n\times n}$ is Hermitian if $A = A^*$ where A^* is defined by $A^*_{ij} = \overline{A_{ji}}$ for every i,j. Usually PSD ness is defined as a matrix $A \in \mathbb{C}^{n\times n}$ is PSD if $\forall v \in \mathbb{C}^n$, $v^* A v \geq 0$.

Observation :- Any matrix $A \in \mathbb{C}^{n\times n}$ that satisfies $v^* A v \geq 0 \quad \forall v \in \mathbb{C}^n$ must be Hermitian]

Fact :-
Any Hermitian matrix $A \in \mathbb{C}^{n\times n}$ decomposition of the form
$$A = \sum \lambda_i u_i u_i^* \quad \text{where}$$
$\lambda_i \in \mathbb{R}$, u_i are orthonormal vectors: ($u_i^* u_j = \delta_{ij}$ for all i,j). Further when $A \in \mathbb{R}^{n\times n}$ (and hence symmetric) u_i can be chosen to have real entries.

– When A is further PSD then $\lambda_i \geq 0$.

Fact :- $A \in \mathbb{C}^{n\times n}$ is PSD iff there are $\{\lambda_i\}$ non-negative and orthonormal unit-vectors $\{u_i\}_{i=1}^n$ such that
$$A = \sum \lambda_i u_i u_i^* .$$

Examples :- Suppose $\mathcal{D}$ is a distribution over $\mathbb{R}^d$. Consider the matrix $M \in \mathbb{R}^{d\times d}$ where

$$M_{ij} := \underset{X\sim\mathcal{D}}{\mathrm{Cor}}(X_i, X_j)$$
$$= \underset{X\sim\mathcal{D}}{E}[X_i X_j] - \underset{X\sim\mathcal{D}}{E}[X_i]\ \underset{X\sim\mathcal{D}}{E}[X_j]$$

<u>Claim</u> :- M is PSD.

<u>Proof</u> :- [First suppose $\mathcal{D}$ has finite support].
[Assume also that $\underset{X\sim\mathcal{D}}{E}[X] = 0$] [Ex: Remove this assumption]

$\mathcal{D} : \{(p_{i_n}, x^{(i)}) \mid 1 \leq i \leq n\}$: $X \sim \mathcal{D} = x^{(i)}$ with probability p_i

$$\sum_{i=1} p_i = 1$$
$$x^{(i)} \in \mathbb{R}^d.$$

Consider $A^{(k)} := x^{(k)} x^{(k)^T}$ | $\forall v \quad v^T A^{(k)} v = (v^T x^{(k)})(x^{(k)^T} v)$

$A^{(k)}_{ij} = x^{(k)}_i x^{(k)}_j$ | $= |v^T x^{(k)}|^2 \geq 0.$

$\Rightarrow A^{(k)} \succcurlyeq 0.$

$$M_{ij} = \underset{X\sim\mathcal{D}}{E}[X_i X_j] = \sum_{k=1}^{n} p_k x^{(k)}_i x^{(k)}_j = \sum_{k=1}^{n} p_k A^{(k)}_{ij}$$

$$\Rightarrow \quad M = \sum p_k A^{(k)}$$

Since $p_k \geq 0 \ \forall k$, and $A^{(k)} \succcurlyeq 0$ we get $M \succcurlyeq 0$.

[<u>Ex</u>:- Every real PSD matrix is a covariance matrix.].

<u>Example</u> :- If we have any inner product space H and $v_1 \ldots v_k \in H$ then the $k\times k$ matrix M given by
$$M_{ij} = \langle v_i, v_j \rangle \quad \text{is PSD.}$$
(<u>Ex</u>:- Again every PSD M has such a representation)

Proof:- (real case) Take any $w \in \mathbb{R}^k$.

$$
\begin{aligned}
w^T M w &= \sum_{i,j} w_i M_{ij} w_j \\
&= \sum_{i,j} w_i \langle v_i, v_j \rangle w_j \\
&= \sum_{i,j} \langle w_i v_i, w_j v_j \rangle \\
&= \sum_j \left\langle \sum_i w_i v_i, w_j v_j \right\rangle \\
&= \left\langle \sum_i w_i v_i, \sum_j w_j v_j \right\rangle \\
&= \left\langle \sum_i w_i v_i, \sum_i w_i v_i \right\rangle \\
&\geq 0.
\end{aligned}
$$

(All we really needed above is bilinear form $\langle \cdot, \cdot \rangle_H$ s.t. $\langle v, v \rangle_H \geq 0$).

[This can be used to prove (for example) that if $\lambda_1, \lambda_2, \ldots \lambda_k > 0$ then the matrix M defined by $M_{ij} = \frac{1}{\lambda_i + \lambda_j}$ is PSD.

(From

$$f_\lambda(t) := e^{-\lambda t}$$

$$\langle g, h \rangle := \int_0^\infty g(t)\, h(t)\, dt \Big)$$

].

<u>Basic properties of PSD matrices</u>

$$M = \sum \lambda_i v_i v_i^*, \quad = V \Gamma V^+, \; V \text{ unitary}$$

$\Gamma = \mathrm{diag}(\vec{\lambda})$.

$$\mathrm{Tr}(M) = \sum \lambda_i$$

Functions of Hermitian matrices :-

Suppose D is diagonal, and f is $\mathbb{C} \to \mathbb{C}$.
One can then 'naturally' define:

$$f(D)_{ii} = f(D_{ii}) \ \forall i.$$

Suppose f looked like (convergent power series in a neighborhood of 0)

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots$$

then our definition above also agrees:

$$f(D) = a_0 + a_1 D + a_2 D^2 + a_3 D^3 + \cdots$$

(in a neighborhood of the 0-matrix).

Then it seems natural to define:

$$f(M) = a_0 + a_1 M + a_2 M^2 + a_3 M^3 \cdots$$

(hope that this 'makes sense' analytically for M close to 0).

Now suppose that M was diagonalizable. Then the above series definition says that

($M = VDV^{-1}$, V invertible, D diagonal)

$$f(M) = V\,(f(D))\,V^{-1}$$

When M is Hermitian,

$M = VDV^*$ where V is unitary ($VV^\dagger = I$),

So, $f(M) = V f(D) V^\dagger$ (definition of f for Hermitian matrices)

Goal :- To understand fns. like log, exp etc. applied to PSD matrices (with a view towards proving concentration)

We will start small by looking at $A \mapsto A^{-1}$ and $A \mapsto \sqrt{A}$ (and $A \mapsto A^2$). | over positive semi-definite matrices.

(Q) :- For reals. $r^2 \geq s^2$ when $r \geq s \geq 0$.
Does $A \succcurlyeq B \succcurlyeq 0$ (this means $A - B \succcurlyeq 0$) imply that
(a) $A^2 \succcurlyeq B^2$?
(b) $A^{-1} \preccurlyeq B^{-1}$ (assume A, B are invertible)
(c) $A^{1/2} \succcurlyeq B^{1/2}$.

(a) is false. Let C, D be PSD.

Consider $f(t) = (C + tD)^2 - C^2$ $\quad (C + tD \succcurlyeq C \;\; \forall t \geq 0)$.

$$= t\,[CD + DC] + t^2 D^2.$$

$$C = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \qquad D = \begin{pmatrix} 1 & \\ & \alpha \end{pmatrix}. \qquad \alpha > 0$$

$$CD + DC = \begin{pmatrix} 1 & \alpha \\ 1 & \alpha \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ \alpha & \alpha \end{pmatrix}$$

$$= \begin{pmatrix} 2 & \alpha + 1 \\ 1 + \alpha & 2\alpha \end{pmatrix}.$$

Fact :- A 2×2 matrix with negative determinant cannot be PSD.

$$\det(CD + DC) = 4\alpha - (\alpha + 1)^2$$
$$= -(\alpha - 1)^2.$$

So, $CD + DC \not\succcurlyeq 0$ when $\alpha = 2$.

So, there exist $\gamma < 0$ and $v \in \mathbb{R}^2$, unit vector

$$v^T (CD+DC) v = \gamma .$$

$$
\begin{aligned}
v^T\left((C+tD)^2 - C^2\right) v &= t\gamma + t^2 v^T D v \\
&\leq t\gamma + 2t^2 \\
&= t(\gamma + 2t). \quad \text{Choose } t = \frac{-\gamma}{3}. \\
&= -\frac{\gamma}{3}\left(+\frac{\gamma}{3}\right) = -\frac{\gamma^2}{9} < 0 .
\end{aligned}
$$

So, $(C+tD)^2 - C^2$ is not PSD.

<u>Example</u> :- Claim : (1) $\log(A) \succcurlyeq \log(B)$ when $A \succcurlyeq B \succ 0$
(Proof in the uploaded notes)

(2) $A \mapsto \exp(A)$ is <u>not</u> monotone increasing on PSD matrices.

<u>Thm</u> :- (1) $A \succcurlyeq B$ then $\mathrm{Tr}(A) \geq \mathrm{Tr}(B)$ for all Hermitian A, B.
(2) If $f : \mathbb{R} \to \mathbb{R}$ is twice differentiable monotone on an interval $[a,b]$ and M, N are Hermitian with eigenvalues in $[a,b]$ s.t. $M \succcurlyeq N$, then $\mathrm{Tr}(f(M)) \geq \mathrm{Tr}(f(N))$.

<u>Pf</u> :- (1) $A \succcurlyeq B \Rightarrow v^T(A-B)v \geq 0 \;\; \forall\, v$.

So, $\mathrm{Tr}(A-B) = \sum_{i=1}^{n} e_i^T (A-B) e_i \geq 0$

(2) Let $B = M - N \succcurlyeq 0$ Define $g(t) = \mathrm{Tr}\, f(N + tB)$.

$g(0) = \mathrm{Tr}(f(N))$, $g(1) = \mathrm{Tr}(f(M))$. and all eigenvalues of $N+tB$ lie in $[a,b]$ when $t \in [0,1]$ (Why?)

Assuming $f'(s) \geq 0 \ \forall s \in [a,b]$, we will prove that $g'(t) \geq 0 \ \forall t \in [0,1]$. (which $\Rightarrow g(1) \geq g(0)$).

$$g(t) = \mathrm{Tr}(f(N+tB)).$$

$$g'(t) = \lim_{s \to 0} \frac{\mathrm{Tr}(f(N+(t+s)B) - f(N+tB))}{s}$$

$$= \lim_{s \to 0} \frac{\mathrm{Tr}(f'(N+tB)\, sB) + O(s^2)}{s}$$

$$= \mathrm{Tr}(f'(N+tB)\,B) \quad \text{where } B \succeq 0.$$

$$= \mathrm{Tr}(B^{1/2} f'(N+tB) B^{1/2}). \quad (\because \mathrm{Tr}(PQ) = \mathrm{Tr}(QP)).$$

Since $N+tB$ has eigenvalues in $[a,b]$, and $f'(x) \geq 0$ when $x \in [a,b]$, $f'(N+tB) \succeq 0$.

$$\Rightarrow B^{1/2} f'(N+tB) B^{1/2} \succeq 0 \quad (\text{by congruence})$$

$$\Rightarrow \mathrm{Tr}(B^{1/2} f'(N+tB) B^{1/2}) \geq 0 \quad [\text{If } A \succeq 0 \text{ and } B \text{ is any matrix then } B^* A B \text{ is PSD.}]$$

$$\Rightarrow g'(t) \geq 0.$$

$$\Rightarrow g(1) = \mathrm{Tr}(M) \geq T(N) = g(0). \quad \blacksquare$$

Thm:- (Lieb's theorem) Let H be a Hermitian matrix. Define the function

$$f_H(A) := \mathrm{Tr}\exp(H + \log A)$$

on positive definite matrices. Then f_H is concave:- $\forall A, B$ positive definite, $\lambda \in [0,1]$,

$$f_H(\lambda A + (1-\lambda)B) \geq \lambda f_H(A) + (1-\lambda) f_H(B).$$

Thm:- (We won't use, but useful to know). [Golden-Thompson inequality]. For A, B hermitian,

$$\mathrm{Tr}(e^{A+B}) \leqslant \mathrm{Tr}(e^A e^B).$$

(Not true that $\mathrm{Tr}(e^{A+B+C}) \leq \mathrm{Tr}(e^A e^B e^C)$.)

Remark:- This would be trivial if A, B were commutative, because then $\underline{e^{A+B} = e^A e^B}$.

not in general true when $AB \neq BA$.

Corollary of Lieb's theorem:- If Z is a random Hermitian matrix then for any Hermitian H

$$E[\mathrm{Tr}\exp(H+Z)] \leqslant \mathrm{Tr}\exp(H + \log E[\exp(Z)])$$

Pf:- $X := \exp(Z) \geqslant 0$. Apply Jensen's inequality to Lieb's theorem,

$$\begin{aligned} E[\mathrm{Tr}(\exp(H+Z))] &= E[\mathrm{Tr}(\exp(H+\log X))] \\ &= E[f_H(X)] \\ &\leqslant f_H(E(X)) \quad \text{(Jensen's inequality) (Check!)} \\ &= \mathrm{Tr}\exp(H + \log E[\exp(Z)]). \end{aligned}$$