ECE 269- Lecture 18: Thursday March 6th

- PD matrices cont.
  - Sylvester Criterion

Recall: A symmetric/Hermitian matrix is called:

- A P.D. matrix if $x^* A x > 0$ for all $x \neq 0$.
- A P.S.D. " " $x^* A x \geq 0$ for all $x = 0$.

Fact: A Hermitian matrix A is P.D. (P.S.D.) iff all eigenvalues are positive (non-negative).

A function $f: \mathbb{R}^d \to \mathbb{R}$ is said to be PD (P.S.D) if $f(x) > 0$ ($f(x) \geq 0$) for all $x \neq 0$

Facts: A P.D. $\Rightarrow \det(A) > 0$

Pf: if $A \succ 0$, then $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0 \Rightarrow$

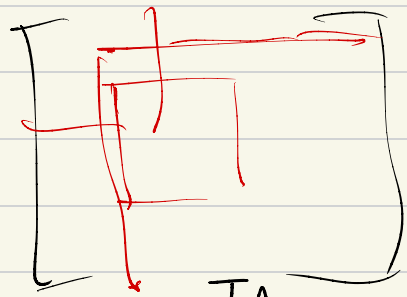
$\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n \geq 0$

For a matrix $A \in \mathbb{C}^{n \times n}$, and $\emptyset \neq S \subseteq \{1, 2, \ldots, n\}$, the matrix A^S that the matrix by dropping all rows i and columns j of A, that $i, j \notin S$ is called a principal submatrix of A.

Example: $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 3 & 1 & 2 \end{bmatrix}$, $A^{\{1\}} = [1]$, $A^{\{2,3\}} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$

We refer to $\det(A^{S})$ as a principal minor.

Fact: A (p.s.d.) p.d. $\Rightarrow$ all principals (p.s.d) P.D $\Rightarrow$ all principal minors are positive (non-negative)

$A =$ 

Pf: A p.d. $\equiv$ $x^T A x > 0$ $\forall x \neq 0$.

if we restrict this condition to all x with one non-zero element,

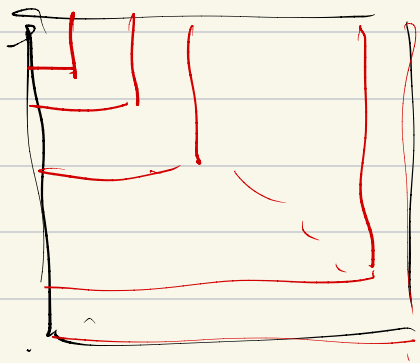
$$0 < (x_1 \ 0 \ \cdots \ 0) \, A \begin{pmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = x_1^2 A^{\{1\}} \Rightarrow \det(A^{\{1\}}) > 0 .$$

Note that for any $\emptyset \neq S \subseteq \{1, \dots, n\}$, we have a principal submatrix

$\Rightarrow$ $2^n - 1$ such submatrice.

Sylvester Criterion:

If $S = \{1, 2, \dots, i\}$, A^S is called a leading principal of A.



Fact: A is p.d. iff all leading principal minors are positive.

Example: $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$, $\det(A^{\{1\}}) = 1 > 0$, $\det(A^{\{1,2\}}) = 1 > 0 \Rightarrow A$ is p.d.

Pf (sufficiency): By induction on n:

if $n=1$, $A = \alpha$ is p.d. iff $\alpha > 0 \Rightarrow$ ✓.

Suppose that the assertion holds for all matrices of size $k \times k$.

Consider $A \in \mathbb{R}^{(k+1)\times(k+1)}$ and suppose all the leading principle minors are positive.

$$A_{(k+1)\times(k+1)} = \left(\begin{array}{c|c} B_{k\times k} & c \\ \hline c^T & \alpha \end{array}\right)$$

By the induction hypothesis B is a p.d. matrix.

By the interlacing of eigenvalues,

$$\lambda_k(A) \geq \lambda_k(B) \geq \lambda_{k+1}(A) > 0$$

$\Rightarrow \lambda_1(A) \geq \cdots \geq \lambda_k(A) > 0$, but $0 < \det(A) = \lambda_1 \cdots \lambda_k(A)\, \lambda_{k+1}(A) \Rightarrow \lambda_{k+1}(A) > 0$.

Fact: A is p.s.d. iff $A = Q^T Q$ for some $k \times n$ matrix Q.

Note that $x^T Q^T Q x = \|Qx\|^2 \geq 0 \Rightarrow$ all matrices of the form $A = Q^T Q$ is a p.s.d. matrix.

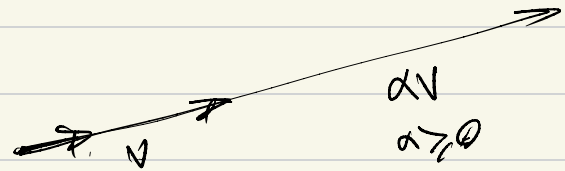
Pf: Since A is symmetric $A = U \Lambda U^T$ and since A is p.s.d. $\Lambda_{ii} \geq 0$. Therefore, letting $Q = \underbrace{\begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix}}_{\Lambda^{\frac{1}{2}}} U^T$ we have the result.

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

Such a decomposition is not necessarily unique, but if we restrict Q to be a lower triangular matrix with positive diagonal, $A = Q^T Q$ exists and it is unique. Such a decomposition is called the Cholesky decomposition of A.

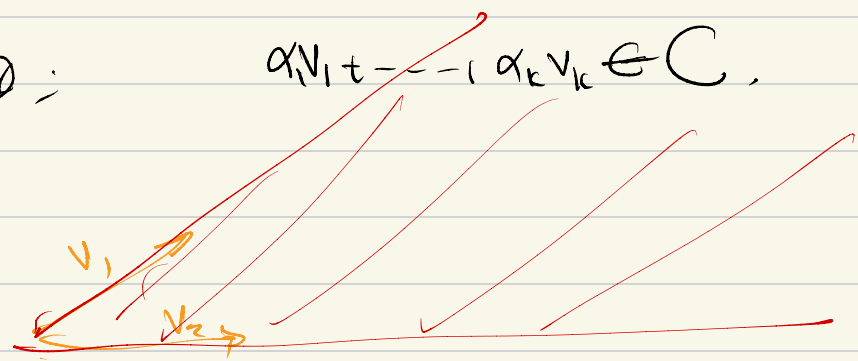
(see Prof. Kim's notes)

Def: We say a $C \subseteq \mathbb{R}^n$ is a cone, if $\forall\, \alpha \geq 0$, $\forall v \in C$
$\alpha v \in C$.



C is called a convex cone, if $\forall v_1, \dots, v_k \in C$,

$\forall \alpha_1, \dots, \alpha_k \geq 0$: $\alpha_1 v_1 + \dots + \alpha_k v_k \in C$.



(p.s.d.)

Fact: The set of p.d. matrices is a convex cone in $\mathbb{R}^{n\times n}$.
(except $\underline{\underline{0}}$)

Pf: $A_1, \ldots, A_k \succeq 0$;

Then for $B = \alpha_1 A_1 + \cdots + \alpha_k A_k$,

$$x^T B x = \sum_{i=1}^{k} \alpha_i x^T A_i x \geq 0.$$

Application: Portfolio optimization:

Suppose that we are interested in distributing our wealth among n stocks $i=1,\dots,n$, i.e., determine $x_1,\dots,x_n \geq 0$:

$$\sum_{i=1}^{n} x_i = 1$$

such that for a given performance level R, it minimizes the risk.

Let $\mu = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix}$ where X_i is the r.v. representing the return of stock i.

$Y = x_1 X_1 + \cdots + x_n X_n$ is the (random) return.

meeting performance: $\mu_1 x_1 + \cdots + \mu_n x_n = R$

One measure for risk is $Var(Y) = x^T \Sigma x$ with

$\Sigma_{ij} = E[(X_i - \bar{X}_i)(X_j - \bar{X}_j)]$ is the cov. matrix of the return.

Markowitz Portfolio Opt.

$$\min: \quad x^T \Sigma x$$

$$\text{s.t.} \quad \sum_{i=1}^{n} x_i = 1$$

$$\sum_{i=1}^{n} \mu_i x_i = R$$