

Symmetric Positive Definite Matrix (section 2.3)

Definition (1) An $n \times n$ matrix A is symmetric if $A^T = A$, i.e. $a_{ji} = a_{ij}$ if $A = [a_{ij}]_{n \times n}$.

(2) An $n \times n$, real, symmetric matrix A is symmetric positive definite (SPD), if

$$x^T A x > 0 \quad \text{for any } x \in \mathbb{R}^n, x \neq 0.$$

Remarks (1) We consider real matrices for symmetric positive definiteness.
For complex matrix, Hermitian replaces symmetric $\bar{A}^T = A$.

Also, $x^T A x$ is replaced by $x^* A x$.
 $x^* = \bar{x}^T$. x : complex.

$$(1) \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

$$x^T A x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \cdot a \left[\begin{array}{c|cc} \sum_{j=1}^n a_{1j} x_j & & \\ \hline & \ddots & \\ & & \sum_{j=1}^n a_{nj} x_j \end{array} \right] = \sum_{j=1}^n x_i \left(\sum_{j=1}^n a_{ij} x_j \right)$$

$$= \sum_{i,j=1}^n a_{ij} x_i x_j.$$

In fact, every symmetric matrix is uniquely related to a quadratic form.

$$A = [a_{ij}], \quad A^T = A.$$

$$\begin{aligned}
 x^T A x &= \sum_{i,j=1}^n a_{ij} x_i x_j \\
 &= a_{11} x_1^2 + a_{12} x_1 x_2 + \cdots + a_{1n} x_1 x_n \\
 &\quad + a_{21} x_2 x_1 + a_{22} x_2^2 + \cdots + a_{2n} x_2 x_n \\
 &\quad + \cdots + \\
 &\quad + a_{n1} x_n x_1 + a_{n2} x_n x_2 + \cdots + a_{nn} x_n^2 \\
 &= a_{11} x_1^2 + a_{22} x_2^2 + \cdots + a_{nn} x_n^2 \\
 &\quad + 2a_{12} x_1 x_2 + 2a_{13} x_1 x_3 + \cdots + 2a_{1n} x_1 x_n \\
 &\quad + \cdots + \\
 &\quad + 2a_{n-1n} x_{n-1} x_n
 \end{aligned}$$

Examples (1) $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 2x_1 x_2 - 2x_1 x_3 + 2x_2 x_3$

$$= x^T A x \quad A = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix}$$

(2) $g(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + 11x_3^2 - 2x_1 x_2 + 6x_1 x_3 - 4x_2 x_3$

$$A = \begin{bmatrix} 1 & -1 & 3 \\ -1 & 2 & -2 \\ 3 & -2 & 11 \end{bmatrix}.$$

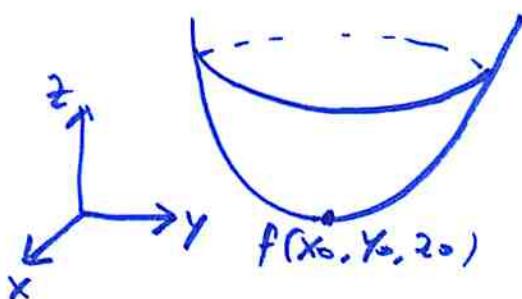
Example $f(x, y, z)$: smooth function.
 Suppose $\nabla f(x_0, y_0, z_0) = 0$, i.e. $\frac{\partial f}{\partial x}(x_0, y_0, z_0) = 0$
 $\frac{\partial f}{\partial y}(x_0, y_0, z_0) = 0$. $\frac{\partial f}{\partial z}(x_0, y_0, z_0) = 0$
 So, f may have a local maximum
 or local minimum at (x_0, y_0, z_0) .

$$\begin{aligned}
 f(x, y, z) &= f(x_0, y_0, z_0) + \frac{\partial f}{\partial x}(x_0, y_0, z_0)(x - x_0) \\
 &+ \frac{\partial f}{\partial y}(x_0, y_0, z_0)(y - y_0) + \frac{\partial f}{\partial z}(x_0, y_0, z_0)(z - z_0) \\
 &+ \frac{1}{2} \left[(x - x_0)^2 \frac{\partial^2 f}{\partial x^2}(x_0, y_0, z_0) + (y - y_0)^2 \frac{\partial^2 f}{\partial y^2}(x_0, y_0, z_0) \right. \\
 &\quad + (z - z_0)^2 \frac{\partial^2 f}{\partial z^2}(x_0, y_0, z_0) + 2(x - x_0)(y - y_0) \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0, z_0) \\
 &\quad + 2(x - x_0)(z - z_0) \frac{\partial^2 f}{\partial x \partial z}(x_0, y_0, z_0) \\
 &\quad \left. + 2(y - y_0)(z - z_0) \frac{\partial^2 f}{\partial y \partial z}(x_0, y_0, z_0) \right] + \dots \text{high order term} \\
 &= f(x_0, y_0, z_0) + \frac{1}{2} \cancel{(x-x_0)^2} \cancel{(y-y_0)^2} \cancel{(z-z_0)^2} \\
 &\quad + \frac{1}{2} (\vec{r} - \vec{r}_0) \cdot \nabla^2 f(x_0, y_0, z_0) + O(|\vec{r} - \vec{r}_0|^3)
 \end{aligned}$$

$$\vec{r} - \vec{r}_0 = \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix}$$

$$\text{Hessian matrix } \nabla^2 f(x_0, y_0, z_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{bmatrix}_{(x_0, y_0, z_0)}$$

If $\nabla^2 f(x_0, y_0, z_0)$ is SPD, then (x_0, y_0, z_0) is a local min. of f



More examples

(1) $x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$.

$$f(x) = f(x_0) + (x - x_0) \cdot \nabla f(x_0) + \frac{1}{2} (x - x_0)^T \nabla^2 f(x_0) (x - x_0) + \dots$$

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right]_{n \times n} : \text{Hessian matrix.}$$

(a) $x^T \nabla^2 f(x_0) x = x^T \nabla^2 f(x_0) x > 0 \Rightarrow x_0: \text{local min.}$

(b) $y^T \nabla^2 f(x) y \geq 0 \quad \forall y. \quad \forall x: \Rightarrow f \text{ is convex.}$

(2) Let X_1, X_2, \dots, X_n be real random variables with finite moments on some probability space. Let $\mu_i = E(X_i) = \bar{x}_i$ be the expectation of X_i .

[Quick review: (Ω, Σ, P) a probability space.
 $X: \Omega \rightarrow \mathbb{R} \cup \{\pm\infty\}$: random variable
mean or expectation: $E(X) = \int x dP =: \bar{x}$
variance $\text{Var}(X) = E(X - \bar{x})^2$
Random vector $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}: \Omega \rightarrow \mathbb{R}^n$.]

Covariance matrix of X is $A = [a_{ij}]_{n \times n}$ with

$$a_{ij} = E((X_i - \bar{x}_i)(X_j - \bar{x}_j)) = a_{ji}. \quad A^T = A.$$

Now let $z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \in \mathbb{R}^n$. Then

$$z^T A z = z \cdot A z = \sum_{i,j=1}^n a_{ij} z_i z_j$$

5

$$= \sum_{i,j=1}^n E((x_i - \bar{x}_i)(x_j - \bar{x}_j)) z_i z_j = \sum_{i,j=1}^n E(z_i(x_i - \bar{x}_i) \cdot z_j(x_j - \bar{x}_j)) \\ = E\left(\sum_{i=1}^n z_i(x_i - \bar{x}_i)\right)^2 \geq 0 = E\left(\sum_{i,j=1}^n z_i(x_i - \bar{x}_i) \cdot z_j(x_j - \bar{x}_j)\right)$$

A is symmetric, semi-definite

(3) Let $f \in C([0,1])$. Define $a_k = \int_0^1 x^k f(x) dx$.

Hausdorff moment sequence: a_0, a_1, a_2, \dots

$$\text{Let } A = \begin{bmatrix} a_0 & a_1 & a_2 & \dots & a_n \\ a_1 & a_2 & a_3 & \dots & a_{n+1} \\ a_2 & a_3 & a_4 & \dots & a_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_n & a_{n+1} & a_{n+2} & \dots & a_{2n} \end{bmatrix}$$

* Let $z = \begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_n \end{bmatrix} \in \mathbb{K}^{n+1}$

$$z^T A z = \sum_{j,k=0}^n a_{j+k} z_j z_k = \sum_{j,k=0}^n \int_0^1 x^{j+k} z_j \cdot z_k f(x) dx \\ = \int_0^1 \left(\sum_{k=0}^n z_k x^k \right)^2 f(x) dx$$

$$\begin{cases} -u''(x) = f(x) & x \in (0,1) \\ u'(0) = u'(1) = 0 \end{cases}$$

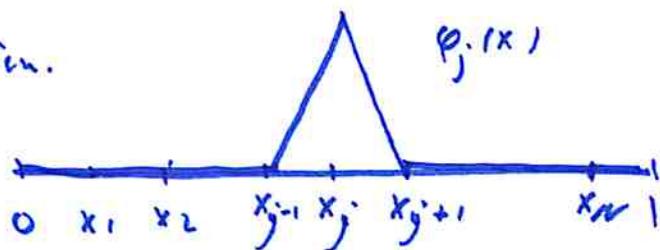
(4) Finite element method.

$$-u''\varphi = f\varphi \quad \text{for any nice } \varphi, \varphi(0) = \varphi(1) = 0$$

Weak form: $\int_0^1 u' \varphi dx = \int_0^1 f \varphi dx \quad \forall \text{ nice } \varphi, \varphi(0) = \varphi(1) = 0$

Finite element basis functions $\varphi_1, \varphi_2, \dots, \varphi_n$.

hat function.



$$\begin{aligned} h &= x_{j+1} - x_j \\ x_j &= j \Delta x \\ \Delta x &= \frac{1}{N-1} \end{aligned}$$

$\varphi_j(x)$: continuous, piecewise linear.

$$\varphi_j(x_j) = 1, \varphi_j(x_i) = 0 \quad (i \neq j).$$

$$\varphi_j(0) = \varphi_j(1) = 0$$

Solution u is approximated by

$$u_h(x) = \sum_{j=1}^N c_j \varphi_j(x)$$

Find all c_1, \dots, c_N .

$$\int_0^1 u_h' \varphi_j' = \int_0^1 f \varphi_j \quad u_h' = \sum_{i=1}^n c_i \varphi_i'$$

$$\sum_{i=1}^N c_i \underbrace{\int_0^1 \varphi_i' \varphi_j' dx}_{a_{ij}} = \int_0^1 f \varphi_j dx \quad j = 1, 2, \dots, n.$$

$$A = [a_{ij}] = \left[\int_0^1 \varphi_i' \varphi_j' dx \right].$$

Call A the stiffness matrix.

$$z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \quad z^T A z = \int_0^1 \left(\sum_{i=1}^n z_i \varphi_i' \right)^2 dx \geq 0$$

If $z^T A z = 0$, then $\sum_{i=1}^n z_i \varphi_i' = 0$

$$\left(\sum z_i \varphi_i \right)' = 0 \Rightarrow \sum z_i \varphi_i = \text{const.}$$

all $\varphi_i(0) = 0 \Rightarrow \text{const} = 0$. $\sum_{i=1}^n z_i \varphi_i(x_j) = 0 \Rightarrow z_j = 0$.

$$\text{So, } \sum z_i \varphi_i = 0.$$

[But, all $\varphi_1, \varphi_2, \dots, \varphi_n$ are linearly indep.]

[So, all $z_i = 0$.]

Why?

A stiffness matrix is SPD.

When a real symmetric matrix is positive definite?

of $A (=A^T)$

The positive definiteness is defined through the quadratic form $x^T A x$. So, let's understand it.

Example Completing the squares.

$$\begin{aligned} a^2 + 2ab + b^2 \\ = (a+b)^2 \end{aligned}$$

$$(1) f(x_1, x_2, x_3) = 2x_1^2 - 4x_1x_2 + 2x_1x_3 \\ - 3x_2^2 + 2x_2x_3 + 5x_3^2$$

$$A = \begin{bmatrix} 2 & -2 & 1 \\ -2 & -3 & 1 \\ 1 & 1 & 5 \end{bmatrix}, \quad = 2 \left(x_1^2 - 2x_1x_2 + x_1x_3 \right) \\ - 3x_2^2 + 2x_2x_3 + 5x_3^2$$

$$= 2 \left[x_1^2 + 2x_1 \left(-x_2 + \frac{1}{2}x_3 \right) + \left(-x_2 + \frac{1}{2}x_3 \right)^2 \right] \\ - 2 \left(x_2 + \frac{1}{2}x_3 \right)^2 - 3x_2^2 + 2x_2x_3 + 5x_3^2$$

$$= 2 \left[x_1 + \left(-x_2 + \frac{1}{2}x_3 \right) \right]^2 + \underbrace{g(x_2, x_3)}_{\text{no } x_1 \text{ in it}} \\ = 2(x_1 - x_2 + \frac{1}{2}x_3)^2 - 3(x_2 - \frac{x_3}{3})^2 + \frac{14}{3}x_3^2$$

$$(2) f(x_1, x_2, x_3) = 2x_1x_2 - x_1x_3 + 4x_2x_3$$

$$\text{Let } x_1 = z_1 + z_2, \quad x_2 = z_1 - z_2, \quad x_3 = z_3$$

$$f(x_1, x_2, x_3) = 2(z_1 + z_2)(z_1 - z_2) - (z_1 + z_2)z_3 \\ + 4(z_1 - z_2)z_3$$

$$= 2z_1^2 - 2z_2^2 - z_1z_3 - z_2z_3 + 4z_1z_3 - 4z_2z_3$$

$$= 2z_1^2 + 3z_1z_3 + 5z_2z_3.$$

$$(\text{Now, } a_{11} = 2 \neq 0).$$

$$= 2(z_1 + \frac{3}{4}z_3)^2 - \frac{50}{3}z_2^2 - \frac{3}{2}(z_3 + \frac{2}{3}z_2)^2$$

$$= (\sqrt{2}(z_1 + \frac{3}{4}z_3))^2 - (\sqrt{\frac{50}{3}}z_2)^2 - (\sqrt{\frac{3}{2}}(z_3 + \frac{2}{3}z_2))^2.$$

8

In general $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ $A = [a_{ij}]_{n \times n} = A^T$.

$$\begin{aligned} f(x) = x^T A x &= (c_{11}x_1 + \dots + c_{1n}x_n)^2 + \dots + (c_{p1}x_1 + \dots + c_{pn}x_n)^2 \\ &\quad - (c_{p+1,1}x_1 + \dots + c_{p+1,n}x_n)^2 - \dots - (c_{r1}x_1 + \dots + c_{rn}x_n)^2 \\ &= y_1^2 + \dots + y_p^2 - y_{p+1}^2 - \dots - y_r^2. \end{aligned}$$

$y = Cx$ $C = [c_{ij}]_{n \times n}$ invertible.

$$x^T A x = y^T \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \ddots & 1 \\ & & & & 0 & \dots & 0 \end{bmatrix} y. \begin{pmatrix} p & 1s \\ q & -1s \\ r & \dots \\ s & 0s \end{pmatrix} = \text{rank } A.$$

Let $S = C^{-1}$. $x = C^{-1}y = Sy$.

$$x^T A x = (Sy)^T A (Sy) = y^T (S^T A S) y = \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \ddots & 1 \\ & & & & 0 & \dots & 0 \end{bmatrix} y$$
 $\text{rank } A = \text{rank}(S^T A S) = \#1s + \#(-1)s = n.$

Theorem. Let $A = [a_{ij}]_{n \times n}$ be a symmetric matrix. Then, there exists an invertible $n \times n$ matrix S such that

$$S^T A S = \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \ddots & 1 \\ & & & & 0 & \dots & 0 \end{bmatrix}$$

with # of 1s = p , # of (-1)s = q . $p+q = \text{rank}(A)$.

In particular,

$$x^T A x = y_1^2 + \dots + y_p^2 - y_{p+1}^2 - \dots - y_r^2$$

where $x = Sy$.

And, A is positive definite \Leftrightarrow ~~A is full rank~~

i.e. ~~$\det A \neq 0$~~ $\Leftrightarrow S^T A S = I$ (identity matrix)

i.e. $A = C^T C$. ($C = S^{-1}$) .

Summarize: $A^T = A$ is positive definite
 $\Leftrightarrow \exists C$ ($\det C \neq 0$) such that $A = C^T C$.

In this case

$$x^T A x = y_1^2 + \dots + y_n^2$$

where $y = \begin{bmatrix} x \\ \vdots \\ x \end{bmatrix}$
Corollary A is SPD $\Rightarrow \det A > 0$. $\left[\begin{array}{l} \text{pf. } \det A = \det(C^T C) \\ = (\det C)^2 > 0 \end{array} \right]$

Note: $A = C^T C \Rightarrow x^T A x = x^T C^T C x = (Cx)^T C x$
 $= \|Cx\|^2 \geq 0$
 $\|Cx\|^2 = 0 \Leftrightarrow Cx = 0 \Leftrightarrow x = 0$.

But. Completing squares is complicated.

Observation $f(x) = x^T A x$ is positive definite
 $\Rightarrow f(x_1, \dots, x_k, 0, \dots, 0) = \begin{bmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} \cdot A \begin{bmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ is
 also positive definite in $\begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$.
 $f(x_1, \dots, x_k, 0, \dots, 0) = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} \cdot A_k \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$.

where $A = \left[\begin{array}{c|c} A_k & a_{k+1} \\ \hline a_{k+1} & a_{kk} \end{array} \right]$ A_k is the upper left $k \times k$ submatrix of A

A_k - principal submatrix.

So, $\det A_k > 0$, $k = 1, 2, \dots, n$.

Turns out the converse is also true. (This can be proved by induction.)

Theorem $A^T = A$ is SPD \Leftrightarrow All principal subdeterminants of A are positive.

In particular. $A = [a_{ij}] = A^T$ is SPD
 \Rightarrow all $a_{ii} > 0$ ($i=1, 2, \dots, n$)

Example $A = \begin{bmatrix} 5 & 2 & -4 \\ 2 & 1 & -2 \\ -4 & -2 & 5 \end{bmatrix}$ A is symmetric.

$$\det(A_{11}) = 5 > 0 \quad \det(A_{22}) = \begin{vmatrix} 5 & 2 \\ 2 & 1 \end{vmatrix} = 1 > 0$$

$$\det(A_{33}) = \det A = 1 > 0. \quad \text{So, } A \text{ is SPD.}$$

Finally, the positive definiteness of $A = A^T$ is related to eigenvalues of A .

Some facts of eigenvalues and eigenvectors of a real symmetric matrix $A = [a_{ij}]_{n \times n}$.

① eigenvalues are real.

$$A\vec{u} = \lambda \vec{u} \quad (\vec{u} \neq \vec{0}) \Rightarrow \lambda: \text{real.}$$

Rewrite $Au = \lambda u, \quad u \neq 0.$

Complex conjugate: $\overline{Au} = \overline{\lambda u} \quad A\bar{u} = \bar{\lambda}\bar{u}$.

$$\bar{u} \cdot Au = \lambda \bar{u} \cdot u = \lambda \|u\|^2 \quad \Rightarrow \lambda = \bar{\lambda}$$

$$\boxed{\bar{u} \cdot u = \overline{u \cdot u} = \bar{\lambda} \bar{u} \cdot u = \bar{\lambda} \bar{u} u = \bar{\lambda} \|u\|^2. \quad \lambda: \text{real.}}$$

Important: $x \cdot Ay = A^T x \cdot y.$

② λ, μ : eigenvalues of $A (= A^T)$.
 $\lambda \neq \mu \Rightarrow$ their eigenvectors are
orthogonal to each other.

$$Au = \lambda u \quad (\lambda \neq 0) \Rightarrow u \cdot v = 0$$

$$Av = \mu v \quad (\mu \neq 0)$$

In fact, $v \cdot Au = v \cdot \lambda u = \lambda v \cdot u$

$$\cancel{\text{---}} \quad \cancel{Av \cdot u} = Av \cdot u = \mu v \cdot u = \mu u \cdot v$$

$$\text{So, } (\lambda - \mu) u \cdot v = 0 \quad \lambda \neq \mu \Rightarrow u \cdot v = 0.$$

③ A symmetric matrix is diagonalizable.
with orthogonal eigenvectors.

Thm let A be an $n \times n$ real symmetric
matrix. Then, there exists an orthogonal
matrix R (i.e. $R^{-1} = R^T$) s.t.

$$R^T A R = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix} \in D$$

where $\lambda_1, \dots, \lambda_n$ are eigenvalues of A and
~~R consists~~ its column vectors of R are correspond.
eigenvectors.

In particular, let $x = Ry$, then

$$x^T A x = (Ry)^T A (Ry) = y^T (R^T A R) y$$

$$= y^T D y = \lambda_1 y_1^2 + \dots + \lambda_n y_n^2.$$

Theorem A real symmetric matrix is SPD
 \Leftrightarrow all its eigenvalues are positive.

Summary $A^T = A$ ($n \times n$), (real).

Def. A is SPD: $x^T A x > 0 \quad \forall x \in \mathbb{R}^n, x \neq 0$.

$\Leftrightarrow A = C^T C$ for some C , $\det(C) \neq 0$

\Leftrightarrow all $\det A_k > 0$ ($k = 1, 2, \dots, n$)

\Leftrightarrow all eigenvalues are positive.