Chap6-Week13-14

2023/05/04 09:00

△ Singular Value Decomposition

List of properties about A^TA and AA^T

① A^TA and AA^T are symmetric. Thus, orthogonally diagonalizable.

② All eigenvalues of A^TA and AA^T are nonnegative (and real).

proof: Assume that λ is an eigenvalue of A^TA with $\vec{x}$ being an eigenvector.

$$\vec{x}^T A^TA\vec{x} = \vec{x}^T\lambda\vec{x}$$

$$(A\vec{x})^T(A\vec{x}) = \lambda\,\vec{x}^T\vec{x} \rightarrow \lambda = \frac{\|A\vec{x}\|^2}{\|\vec{x}\|^2} \quad \text{since } \vec{x}\neq\vec{0},\ \|\vec{x}\|\neq 0.$$

Hence $\lambda \geq 0$ as $\|A\vec{x}\| \geq 0$.

Similarly, AA^T also contains nonnegative eigenvalues only.

③ $\text{rank}(A^TA) = \text{rank}(AA^T) = \text{rank}(A) = \text{rank}(A^T)$. (See 5.2)

(Thm: $N(A^TA) = N(A)$, $N(AA^T) = N(A^T)$)

proof: (i) $N(A^TA) \subseteq N(A)$:

$\forall\ \vec{x} \in N(A^TA)$,

$$\vec{x}^T A^TA\vec{x} = \vec{x}^T\vec{0} = 0 \rightarrow \|A\vec{x}\|^2 = 0 \rightarrow A\vec{x} = \vec{0}$$
$$\rightarrow \vec{x} \in N(A)$$

(ii) $N(A) \subseteq N(A^TA)$:

$\forall\ \vec{y} \in N(A)$, $A^TA\vec{y} = A^T\vec{0} = \vec{0} \rightarrow A^TA\vec{y} = \vec{0} \rightarrow \vec{y} \in N(A^TA)$

$N(A^TA) = N(A)$.

④ A^TA and AA^T have the same non-zero eigenvalues.

proof: Assume that $\lambda \neq 0$ is an eigenvalue of A^TA with a corresponding eigenvector $\vec{x}$.

$$A^TA\vec{x} = \lambda\vec{x}$$
$$AA^TA\vec{x} = A\lambda\vec{x}$$
$$AA^T(A\vec{x}) = \lambda(A\vec{x})$$

Since $N(A) = N(A^TA)$, then if $\vec{x} \in N(A)$, $\lambda\vec{x} = A^TA\vec{x} = \vec{0}$ which is impossible.

Hence $A\vec{x} \neq \vec{0}$ if $\vec{x}$ is an eigenvector corresponding to $\lambda \neq 0$.

Thus, λ is also an eigenvalue of AA^T with a corresponding eigenvector $A\vec{x}$.

Vice versa, any nonzero eigenvalue of AA^T is also an eigenvalue of A^TA.

⑤ $\text{rank}(A^TA)$ = the number of nonzero eigenvalues of A^TA.

> Remark: For a general square matrix B, the $\text{rank}(B)$ is NOT necessarily equal to the number of nonzero eigenvalues of B.
>
> E.g. $B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$
>
> $\text{rank}(B) = 2$, but all eigenvalues of B are 0.

proof: Since A^TA is a real symmetric matrix, then it is (orthogonally) diagonalizable by some orthogonal matrix Q.

$$A^TA = QDQ^T, \quad \text{where } D \text{ is diagonal.}$$

$$\text{rank}(A^TA) = \text{rank}(QDQ^T) = \text{rank}(D)$$

⑦ σ_i is the square root of eigenvalues of A^TA, nonnegative.

⑧ Columns of V are eigenvectors of A^TA.
Similarly, columns of U are eigenvectors of AA^T.

proved in the last lecture

⑦ σ_i is the square root of eigenvalues of A^TA, nonnegative.

⑧ Columns of V are eigenvectors of A^TA.
Similarly, columns of U are eigenvectors of AA^T.

— proved in the last lecture

⑨ Any eigenvector of AA^T belonging to a nonzero eigenvalue is in $Col(A)$

proof: Method 1: Recall the 4^{th} property: If $\lambda \neq 0$ is a nonzero eigenvalue of A^TA, with eigenvector $\vec{x}$,

then λ is also an eigenvalue of AA^T with $A\vec{x} \in Col A$.

Method 2: Assume $\mu \neq 0$ is a nonzero eigenvalue of AA^T with an eigenvector $\vec{y}$.

$$A(\underbrace{A^T\vec{y}}_{\vec{z}}) = AA^T\vec{y} = \mu\vec{y} \in Col(A) \quad \text{since } \mu \neq 0.$$

⑩ Any eigenvector of AA^T belonging to eigenvalue 0 is in $N(A^T)$.

For any eigenvector $\vec{x}$ of AA^T w.r.t. $\lambda = 0$,

$$\vec{x} \in N(AA^T - 0 \cdot I) \Rightarrow \vec{x} \in N(AA^T) = N(A^T).$$

⑪ Comparing the j^{th} column of each side of the equation

$$AV = U\Sigma$$

we get

$$A\vec{v}_j = \sigma_j \vec{u}_j \qquad j = 1, 2, \cdots, r$$

similarly,

$$A^TU = V\Sigma^T$$

and hence

$$A^T\vec{u}_j = \sigma_j \vec{v}_j, \quad j = 1, 2, \cdots, n$$

$$A^T\vec{u}_j = \vec{0}, \quad j = r+1, \cdots, m$$

$$\longrightarrow \quad A^T[\vec{u}_1 \cdots \vec{u}_m] = [\vec{v}_1 \cdots \vec{v}_n]\left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \sigma_2 & & O \\ & & \ddots & \\ & & & \sigma_n \end{array}\right]$$

$$A^T\vec{u}_j = \begin{cases} \sigma_j \vec{v}_j, & j = 1, 2, \cdots, r \\ \vec{0}, & j = r+1, \cdots, m \end{cases}$$

$$\vec{u}_j = \begin{cases} \frac{1}{\sigma_j} A\vec{v}_j, & \text{for } \sigma_j \neq 0 \\ \in N(A^T), & \sigma_j = 0. \end{cases}$$

Thm (The SVD Theorem)

If A is an $m \times n$ matrix of rank r, then A has a singular value decomposition.
That is,

$$A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V^T_{n\times n}$$

where

U is an $m\times m$ orthogonal matrix
V is an $n\times n$ orthogonal matrix

and

Σ is an $m\times n$ sparse matrix whose diagonal elements

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_{\min\{m,n\}} = 0$$

$$\Sigma = \left[\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & O_{m\times(n-m)} \\ & & \ddots & & \\ & & & \sigma_m & \end{array}\right]_{m\times n} \quad (m < n) \qquad \text{or} \qquad \left[\begin{array}{cccc} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ \hline & O_{(m-n)\times n} & & \end{array}\right]_{m\times n} \begin{matrix} \} \ n \text{ rows} \\ \\ \end{matrix} \quad (m > n)$$

The σ_i's are unique and called **the singular values of A**.

Remark: The proof of this theorem is based on the construction of U, Σ and V. (构造法)

Please read the textbook for more details.

As a result, consider $A = U\Sigma V^T$, then

(without loss of generality, assume $m>n$)

$$(A^TA)_{n\times n} = V\Sigma^T \underbrace{U^TU}_{I_{m\times m}} \Sigma V^T$$

$$= V_{n\times n}\, \Sigma^T_{n\times m}\, I_{m\times m}\, \Sigma_{m\times n}\, V^T_{n\times n}$$

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0$

$$= V_{n\times n} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_n & \end{bmatrix}_{n\times m} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & 0 & \end{bmatrix}_{m\times n} V^T$$

$$= V \begin{bmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \ddots \\ & & & \sigma_n^2 \end{bmatrix}_{n\times n} V^T$$

orthogonal diagonalization of A^TA

$\Sigma^T\Sigma$ contains eigenvalues of A^TA

Columns of V contains the corresponding eigenvectors of A^TA.

$$AA^T = U\Sigma \underbrace{V^TV}_{I} \Sigma^T U^T$$

$$= U \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & 0 & \end{bmatrix}_{m\times n} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_n & \end{bmatrix}_{n\times m} U^T$$

$$= U \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & 0 \\ & & \ddots & \\ & & & \sigma_n^2 \\ & 0 & & 0 \end{bmatrix}_{m\times m} U^T$$

there could be zero eigenvalues within $\{\sigma_1^2, \cdots \sigma_n^2\}$.

orthogonal diagonalization of AA^T

$\Sigma\Sigma^T$ contains eigenvalues of AA^T

Columns of U contains corresponding eigenvectors of AA^T.

△ the procedure of finding SVD :

$$A = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}_{m\times n} = \underbrace{\begin{bmatrix} | & | & \cdots & | \end{bmatrix}_{m\times m}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & 0 & \end{bmatrix}_{m\times n}}_{\Sigma} \underbrace{\begin{bmatrix} - \\ - \\ \vdots \\ - \end{bmatrix}_{n\times n}}_{V^T}$$

eigenvectors of AA^T (columns of U); eigenvectors of A^TA (rows of V^T)

Step 1 : Singular values σ_i

$$\sigma_i = \sqrt{\lambda_i} \quad \text{for each } i$$

where λ_i is the eigenvalue of A^TA.

let $\underbrace{\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_r}_{\text{non zero eigenvalues}} > \underbrace{\sigma_{r+1} = \cdots = \sigma_{\min\{m,n\}} = 0}_{\text{zero eigenvalues}}$

Step 2 : Columns in V are eigenvectors of A^TA corresponding to $\sigma_1^2, \sigma_2^2, \cdots \sigma_n^2$
(orthonormal)

This might involve normalization of vectors and Gram-Schmidt process.

Step 3 : Columns in U :

Method 1 : eigenvectors of AA^T corresponding to $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2, 0 \cdots 0$
(orthonormal)

Method 2 : $\vec{u}_i = \frac{1}{\sigma_i} A\vec{v}_i$ for nonzero σ_i's.

For $\sigma_j = 0$, take $\vec{u}_j \in N(A^T)$, use Gram-Schmidt.

Example

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}_{3\times 2} = U\Sigma V^T$$, notice that rank(A) = 1

Step 1 : Singular Values of A

The eigenvalues of A^TA :

$$A^TA = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}_{2\times 3} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}_{3\times 2} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}_{2\times 2}$$

rank $(A^TA) = 1$

$\lambda_1 = 4$, $\lambda_2 = 0$

$$\sigma_1 = \sqrt{\lambda_1} = 2, \quad \sigma_2 = 0, \quad \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}_{3\times 2}$$

Step 2 : Columns in V (eigenvector of A^TA)

For $\lambda_1 = 4$, $(A^TA - 4I)\vec{x}_1 = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} \vec{x}_1 = \vec{0} \rightarrow \vec{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \rightarrow \vec{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

For $\lambda_2 = 0$, $A^TA\vec{x_2} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}\vec{x_2} = \vec{0} \rightarrow \vec{x_2} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \rightarrow \vec{v_2} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$$V = \begin{bmatrix} \vec{v_1} & \vec{v_2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}.$$

Step 3: Columns in U

$\sigma_1 = 2$ -- $\vec{u_1} = \frac{1}{\sigma_1}A\vec{v_1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \in Col(A)$

$\sigma_2 = 0$: $\vec{u_2} \in N(A^T)$, $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow \vec{y} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \alpha\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + \beta\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

let $\vec{u_2} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, $\vec{u_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ (Since $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \perp \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, we don't need to perform Gram-Schmidt)

Then

$$U = \begin{bmatrix} \vec{u_1} & \vec{u_2} & \vec{u_3} \end{bmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Exercise:

Also in $Col(A)$ — Also in $col(A^T)$ — Also in $N(A)$

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \sqrt{2} & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix}$$

eigenvectors of AA^T belonging to 2

eigenvectors of A^TA belonging to 2

eigenvectors of A^TA belonging to $\lambda = 0$.

Example: U and V are not unique.

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$$

orthonormal basis of $Col(A)$ — orthonormal basis of $N(A^T)$ — orthonormal basis of $Col(A^T)$

$Col(A) \oplus N(A^T) = \mathbb{R}^3$

$Col(A^T) \oplus N(A) = \mathbb{R}^2$

$span\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\} \oplus \{\vec{0}\}$

In addition, if $\underbrace{\{\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}, \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}\}}_{}$ (eigenvectors of A^TA) forms an orthonormal basis of $Col(A^T)$,

then $\{A\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}, A\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}\}$ forms an orthogonal basis of $Col(A)$.

If $\{\vec{v_1}, \vec{v_2}, \cdots, \vec{v_n}\}$ in $Col(A^T)$, eigenvectors of A^TA

then $\{\underbrace{\frac{1}{\sigma_1}A\vec{v_1}, \frac{1}{\sigma_2}A\vec{v_2}, \cdots, \frac{1}{\sigma_r}A\vec{v_r}}_{\text{orthonormal basis of } Col(A)}, \underbrace{\vec{u}_{r+1}, \cdots, \vec{u}_m}_{\text{Orthonormal basis of } N(A^T)}\}$

A Brief Summary:

Square Matrices : $A \in \mathbb{C}^{n\times n}$
(general matrix)

$A \in \mathbb{R}^{n\times n}$

Eigenvalues : could be complex, could be real

Eigenvectors : For $\lambda_i \neq \lambda_j$, $\vec{v_i}$ and $\vec{v_j}$ are linearly independent

Diagonalizable ? Some yes, some no

$A \in \mathbb{R}^{n\times n}$, $A = PDP^{-1}$ $A = CJC^{-1}$

Symmetric Matrices : $A \in \mathbb{R}^{n\times n}$, $A^T = A$

Eigenvalues : All real

Eigenvectors : For $\lambda_i \neq \lambda_j$, $\vec{v_i} \perp \vec{v_j}$

Diagonalizable? Yes! Orthogonally diagonalizable.

$A = QDQ^T$

$A^T = A \Leftrightarrow A$ is orthogonally diagonalizable

Non-square Matrices : $A \in \mathbb{R}^{m\times n}$

Eigenvalues ? No eigenvalues, but it has singular values

$\sigma_i = \sqrt{\text{eigenvalues of } A^TA}$

$$A = U\Sigma V^T$$

△ Quadratic form (Sec. 6.6)

o. Quadratic Equation of two variables x, y.

$\underbrace{ax^2 + by^2 + cxy}_{} + \underbrace{dx + ey}_{} + f = 0$ $\rightarrow \underbrace{[x\ y]}_{\vec{x}^T}\begin{bmatrix} a & c/2 \\ c/2 & b \end{bmatrix}\underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{} + [d\ e]\underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\vec{x}} + f = 0$

- Quadratic Equation of two variables x, y.

$$ax^2+by^2+cxy+dx+ey+f=0 \quad\to\quad \underbrace{[x\ y]}_{\vec{x}^T}\underbrace{\begin{bmatrix} a & c/2 \\ c/2 & b\end{bmatrix}}_{A}\underbrace{\begin{bmatrix} x \\ y\end{bmatrix}}_{\vec{x}} + [d\ e]\underbrace{\begin{bmatrix} x \\ y\end{bmatrix}}_{\vec{x}} + f = 0$$

The term

$$ax^2+by^2+cxy \quad\to\quad \underbrace{[x\ \ y]_{1\times 2}}_{\vec{x}^T}\underbrace{\begin{bmatrix} a & c/2 \\ c/2 & b\end{bmatrix}_{2\times 2}}_{A}\underbrace{\begin{bmatrix} x \\ y\end{bmatrix}_{2\times 1}}_{\vec{x}}$$

is called the quadratic form of the above equation.

- Quadratic Equation of three variables x, y, z.

$$ax^2+by^2+cz^2+dxy+exz+fyz+gx+hy+kz+l=0 \quad\to$$

The term

$$ax^2+by^2+cz^2+dxy+exz+fyz \quad\to\quad \underbrace{[x\ \ y\ \ z]}_{\vec{x}^T}\underbrace{\begin{bmatrix} a & d/2 & e/2 \\ d/2 & b & f/2 \\ e/2 & f/2 & c\end{bmatrix}_{3\times 3}}_{A}\underbrace{\begin{bmatrix} x \\ y \\ z\end{bmatrix}}_{\vec{x}}$$

is called the quadratic form of the above eqn.

- A Quadratic Equation in n variables $x_1, x_2, \cdots, x_n$ has the form

$$\vec{x}^TA\vec{x} + B\vec{x} + \alpha = 0$$

where $\vec{x} = (x_1, x_2 \cdots x_n)^T$, A is an $n\times n$ symmetric matrix, B is a $1\times n$ matrix and $\alpha \in \mathbb{R}$.

The vector function $f(\vec{x})$ is a quadratic form in n variables associating with the quadratic eqn.

$$f(\vec{x}) = \vec{x}^TA\vec{x} = [x_1 \cdots x_n]\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn}\end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_n\end{bmatrix} = \sum_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij}x_i\right)x_j$$

Example $x_1x_2 - x_1^2 + 2x_2 = 0$ is a quadratic eqn. and $x_1x_2 - x_1^2$ is its quadratic terms.

$x_1x_2^2 - x_1^2 + 2x_2 = 0$ is not a quadratic eqn.

$$f_1(\vec{x}) = x_1^2 + 2x_1x_2 + x_2^2 = [x_1\ \ x_2]\begin{bmatrix} 1 & 1 \\ 1 & 1\end{bmatrix}\begin{bmatrix} x_1 \\ x_2\end{bmatrix} \quad \text{is a quadratic form.}$$

$$f_2(\vec{x}) = x_1^2 + x_2^2 + 2x_1x_2 - 4x_1x_3 + 2\sqrt{2}\,x_2x_3 = [x_1\ \ x_2\ \ x_3]\begin{bmatrix} 1 & 1 & -2 \\ 1 & 1 & \sqrt{2} \\ -2 & \sqrt{2} & 0\end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3\end{bmatrix} \quad \text{is a quadratic form.}$$

Remark: It is possible to write the quadratic form as $\vec{x}^TA\vec{x}$ for some non-symmetric matrix A.

For example,

$$f_1(\vec{x}) = \vec{x}^T\begin{bmatrix} 1 & 0 \\ 2 & 1\end{bmatrix}\vec{x}.$$

However, we prefer symmetry due to a lot of good properties.

Exercise Let $Q(\vec{x}) = 2x_1^2 - 3x_2^2 + 2x_3^2 + 2x_1x_3 = \vec{x}^T\begin{bmatrix} 2 & & 1 \\ & -3 & \\ 1 & & 2\end{bmatrix}_{3\times 3}\vec{x}$

Chap6.6-6.7 Definite Matrices

1. Definitions of different types of definite matrices
2. Some important theorems
3. Applications: Hessian matrix and identifying local extrema/saddle point

Recall:

Thm A is a real symmetric matrix

$\Leftrightarrow$ A is orthogonally diagonalizable.

Thm If A is a real symmetric matrix, and $\lambda \neq \mu$ are two distinct eigenvalues of A, $\vec{v}$ and $\vec{w}$ are corresponding eigenvectors.

Then $\vec{v} \perp \vec{w}$.

Example (Motivating) Consider the function

$$q(\vec{x}) = 8x_1^2 - 4x_1x_2 + 5x_2^2$$

from $\mathbb{R}^2$ to $\mathbb{R}$. Determine whether $q(0,0)$ is a global max/min or neither.

Note that $q(\vec{x}) = [x_1\ \ x_2]\begin{bmatrix} 8 & -2 \\ -2 & 5\end{bmatrix}\begin{bmatrix} x_1 \\ x_2\end{bmatrix} = \vec{x}^TA\vec{x}$

Rotation

where A is symmetric and orthogonally diagonalizable by

$$Q = [\vec{v}_1 \quad \vec{v}_2] = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}, \quad D = \begin{bmatrix} 9 & \\ & 4 \end{bmatrix}$$

Denote $\mathcal{B} = \{\vec{v}_1, \vec{v}_2\}$ as a ordered orthonormal basis for $\mathbb{R}^2$, then write

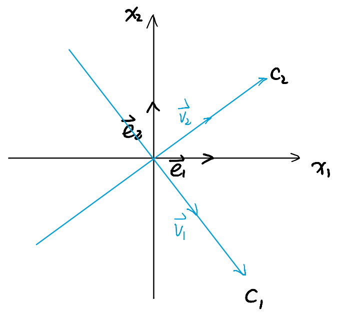
$$\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 = [\vec{v}_1 \quad \vec{v}_2]\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = Q\vec{c}$$

Then

$$q_2(\vec{x}) = \vec{x}^T A \vec{x} = \underbrace{\vec{c}^T Q^T}_{\vec{x}^T}\underbrace{QDQ^T}_{A}\underbrace{Q\vec{c}}_{\vec{x}} = \vec{c}^T D \vec{c} = [c_1 \quad c_2]\begin{bmatrix} 9 & \\ & 4 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

$$= 9c_1^2 + 4c_2^2 > 0 \qquad \text{for nonzero } \vec{x} \in \mathbb{R}^2.$$

So $q_2\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right) = 0$ is a global minimum.



Change of coordinates
from $\{\vec{e}_1, \vec{e}_2\}$ to $\mathcal{B} = \{\vec{v}_1, \vec{v}_2\}$

Notice that $q_2(\vec{x}) = 9c_1^2 + 4c_2^2$ is easier to work with than $8x_1^2 - 4x_1x_2 + 5x_2^2$

since there's no term involving c_1c_2.

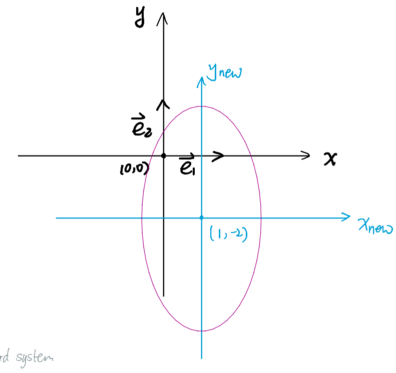
Example: Consider the equation

$$9x^2 - 18x + 4y^2 + 16y - 11 = 0$$

then in matrix form:

$$\underbrace{\vec{x}^T\begin{bmatrix} 9 & \\ & 4 \end{bmatrix}\vec{x}}_{\text{no } xy \text{ term}} + \underbrace{[-18 \quad 16]\,\vec{x}}_{\text{but } 1^{st} \text{ order term exists}} - 11 = 0$$

Translation



Complete the squares to absorb the 1^{st} order terms:

$$9x^2 - 18x + 9 + 4y^2 + 16y + 16 - 36 = 0$$

$$\rightarrow \quad 9(x-1)^2 + 4(y+2)^2 - 36 = 0$$

or $\quad \frac{(x-1)^2}{4} + \frac{(y+2)^2}{9} = 1 \qquad$ an ellipse centered at $(1,-2)$ in the standard system

Set $\quad x_{new} = x - 1, \quad y_{new} = y + 2, \qquad$ translation (平移), not a linear transformation

$$\frac{x_{new}^2}{2^2} + \frac{y_{new}^2}{3^2} = 1 \qquad \text{an ellipse centered at the origin in } x_{new}\text{-}y_{new}\text{-plane}$$

By rotating and translating the axes, it is possible to rewrite the equation into the standard form:

(i) $\quad \lambda_1(x_{new})^2 + \lambda_2(y_{new})^2 + \alpha_{new} = 0 \quad$, when A is nonsingular

or

(ii) $\quad \lambda_1(x_{new})^2 + \beta_{new}\, y_{new} + \alpha_{new} = 0 \quad$ or $\quad \lambda_2(y_{new})^2 + \beta_{new}\, x_{new} + \alpha_{new} = 0$, when A is singular.

E.g. $\quad A = \begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix}$

check:

$\forall$ nonzero $\vec{x} \in \mathbb{R}^2$,

$$\vec{x}^T A \vec{x} = [x_1 \quad x_2]\begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 5x_1^2 + 4x_1x_2 + 5x_2^2$$

$$= 5\left(x_1^2 + \tfrac{4}{5}x_1x_2 + \tfrac{4}{25}x_2^2\right) + \left(5 - \tfrac{4}{5}\right)x_2^2$$

$$= 5\left(x_1 + \tfrac{2}{5}x_2\right)^2 + \tfrac{21}{5}x_2^2 > 0$$

$\Rightarrow A$ is positive definite.

Exercise: $\quad A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ is negative definite

$\forall$ nonzero $\vec{x} \in \mathbb{R}^2 \qquad \vec{x}^T A \vec{x} = -(x_1^2 + x_2^2) < 0$

$B = \begin{pmatrix} -1 & 0 \\ 0 & 3 \end{pmatrix}$ is indefinite.

△ **Definition of Definite Matrices**

A real symmetric matrix $A_{n\times n}$ is said to be

(i) **positive definite** if $\vec{x}^T A \vec{x} > 0$ for all nonzero $\vec{x} \in \mathbb{R}^n$.

(ii) **negative definite** if $\vec{x}^T A \vec{x} < 0$ for all nonzero $\vec{x} \in \mathbb{R}^n$.

(iii) **positive semidefinite** if $\vec{x}^T A \vec{x} \geq 0$ for all nonzero $\vec{x} \in \mathbb{R}^n$.

(iv) **negative semidefinite** if $\vec{x}^T A \vec{x} \leq 0$ for all nonzero $\vec{x} \in \mathbb{R}^n$.

(v) **indefinite** if $\vec{x}^T A \vec{x}$ takes on values that differ in sign.

Thm Let A be a real symmetric $n\times n$ matrix. Then A is positive definite if and only if all its eigenvalues are positive.

Thm Let A be a real symmetric $n \times n$ matrix. Then A is positive definite if and only if all its eigenvalues are positive.

proof: (i) Let A be positive definite.

Assume that $\vec{x}$ is an eigenvector of A w.r.t λ, then $\vec{x} \neq \vec{0}$

$$\vec{x}^T A \vec{x} = \vec{x}^T (\lambda \vec{x}) = \lambda \vec{x}^T \vec{x}$$

$$\rightarrow \lambda = \frac{\overset{>0}{\vec{x}^T A \vec{x}}}{\underset{>0}{\vec{x}^T \vec{x}}} > 0$$

since $\vec{x}^T \vec{x} = \|\vec{x}\|^2 > 0$

(ii) All eigenvalues of A are positive.

Let $A = QDQ^T$ for an orthogonal matrix Q and

$$D = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

$\forall$ nonzero $\vec{x} \in \mathbb{R}^n$,

$$\vec{x}^T A \vec{x} = \underbrace{\vec{x}^T Q}_{\vec{y}^T} D \underbrace{Q^T \vec{x}}_{\vec{y}_{n\times 1}} = \vec{y}^T D \vec{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

$$\geq \min\{\lambda_i\}_{i=1}^n (y_1^2 + y_2^2 + \cdots + y_n^2) = \min\{\lambda_i\}_{i=1}^n \cdot \|\vec{y}\|^2 > 0$$

Since $\min\{\lambda_i\}_{i=1}^n$ is positive and $\vec{y} = Q^T \vec{x} \neq \vec{0}$ for nonsingular Q^T.

E.g. $A = \begin{bmatrix} 3 & 2 \\ 2 & 0 \end{bmatrix}$

Symmetric A $\Rightarrow$ All eigenvalues real $\begin{cases} \lambda_1 + \lambda_2 = Tr(A) = 3 \\ \lambda_1 \cdot \lambda_2 = det(A) = -6 \end{cases}$

$\rightarrow \begin{cases} \lambda_1 = 4 \\ \lambda_2 = -1 \end{cases}$ A is indefinite

△ Eigenvalue Test

(i) A is positive definite $\Leftrightarrow$ All eigenvalues of A are positive.

(ii) A is negative definite $\Leftrightarrow$ All eigenvalues of A are negative.

(iii) A is indefinite $\Leftrightarrow$ Some eigenvalues of $A > 0$ and some < 0.

(iv) A is positive semidefinite $\Leftrightarrow$ All eigenvalues of A are non-negative.

(v) A is negative semidefinite $\Leftrightarrow$ All eigenvalues of A are non-positive.

E.g. $A = \begin{bmatrix} 6 & 4 & -2 \\ 4 & 5 & 3 \\ -2 & 3 & 6 \end{bmatrix}$

By eigenvalue test,

$$\det(A - \lambda I) = \begin{vmatrix} 6-\lambda & 4 & -2 \\ 4 & 5-\lambda & 3 \\ -2 & 3 & 6-\lambda \end{vmatrix} = (6-\lambda)[(5-\lambda)(6-\lambda)-9] - 4[4(6-\lambda)+6] - 2[12 + 2(5-\lambda)]$$

$$= (6-\lambda)(\lambda^2 - 11\lambda + 30 - 25) - 4(5-\lambda) - 48$$

$$= -\lambda^3 + 17\lambda^2 + 67\lambda - 38$$

$\lambda_1 = -0.5015$, $\lambda_2 = 7.8569$, $\lambda_3 = 9.6465$

$\Rightarrow$ A is indefinite.