Section 5.1 Orthogonal Subspaces

Definition 5.1.1 Inner product in R^n (dot product)

Let $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T$ and $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)^T$ be vectors in R^n. Then

$$\boldsymbol{x}^T\boldsymbol{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n. = y^Tx$$

Given a vector $\boldsymbol{x}$ in R^n, its length is defined by

lenth $$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^T\boldsymbol{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

$= \sqrt{\vec{x}\cdot\vec{x}} \geq 0$ equality holds if $\vec{x} = 0$

Definition 5.1.2 angle

The angle θ between two nonzero vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in R^n is defined by

$$-1 \leq \cos\theta = \frac{\boldsymbol{x}^T\boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|} \leq 1 \quad 0 \leq \theta \leq \pi.$$

$$-\|x\|\,\|y\| \leq x^Ty = \|x\|\,\|y\|\cos\theta \leq \|x\|\,\|y\|$$

Theorem 5.1.3

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ vectors in R^n, then

$$|\boldsymbol{x}^T\boldsymbol{y}| \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|$$

Remark 1:

with equality holding if and only if one of the vectors is **0** or one vector is a multiple of the other.

Remark 2: $\vec{x} \perp \vec{y}$

If $\boldsymbol{x}^T\boldsymbol{y} = 0$, then either one of the vectors is the zero vector or $\cos\theta = 0$. If $\cos\theta = 0$, the angle between the vectors is a right angle.

Definition 5.1.4 orthogonal vector

The (column) vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in R^n are said to be **orthogonal** if $\boldsymbol{x}^T\boldsymbol{y} = 0$.

Def. Distance between two vector in R^n

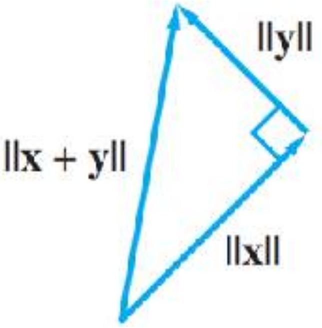
$\forall\ \vec{x}, \vec{y} \in R^n$ the distance $\vec{x}, \vec{y}$ is

$\|\vec{x} - \vec{y}\| = \sqrt{(\vec{x}-\vec{y})^T(\vec{x}-\vec{y})} = \sqrt{(x^T - y^T)(\vec{x}-\vec{y})}$

$= \sqrt{x^Tx - x^Ty - y^Tx + y^Ty} = \sqrt{\|\vec{x}\|^2 - 2x^Ty + \|\vec{y}\|^2}$

$= \sqrt{\|\vec{x}\|^2 - 2\|x\|\|y\|\cos\theta + \|\vec{y}\|^2}$

$\theta = \frac{\pi}{2} \Rightarrow \|\vec{x} - \vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2$



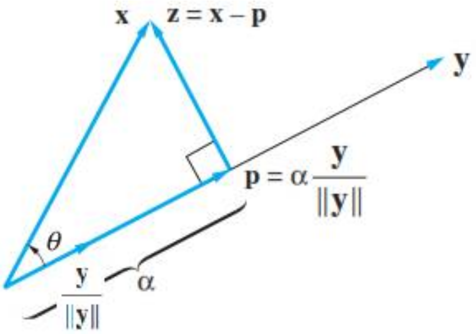
$$\|\mathbf{x}+\mathbf{y}\|^2 = (\mathbf{x}+\mathbf{y})^T(\mathbf{x}+\mathbf{y}) = \|\mathbf{x}\|^2 + 2\mathbf{x}^T\mathbf{y} + \|\mathbf{y}\|^2$$

$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \quad \Rightarrow \quad \mathbf{x}^T\mathbf{y} = 0.$$

Definition 5.1.5

Scalar projection of **x** onto **y**: $\alpha = \dfrac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|}$ $\alpha = \|\vec{x}\| \cdot \cos\theta = \|\vec{x}\| \cdot \dfrac{x^T y}{\|x\|\,\|y\|} = \dfrac{x^T y}{\|y\|} = x^T \dfrac{y}{\|y\|}$

Vector projection of **x** onto **y**: $\mathbf{p} = \alpha \dfrac{1}{\|\mathbf{y}\|}\mathbf{y} = \dfrac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}}\mathbf{y}$ $p = \alpha \cdot \dfrac{\vec{y}}{\|y\|} = x^T \dfrac{y}{\|y\|} \cdot \dfrac{y}{\|y\|} = \dfrac{x^T y}{\|y\|^2}\vec{y}$



$$\alpha = \|\mathbf{x}\|\cos\theta = \frac{\|\mathbf{x}\|\,\|\mathbf{y}\|\cos\theta}{\|\mathbf{y}\|} = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|}$$

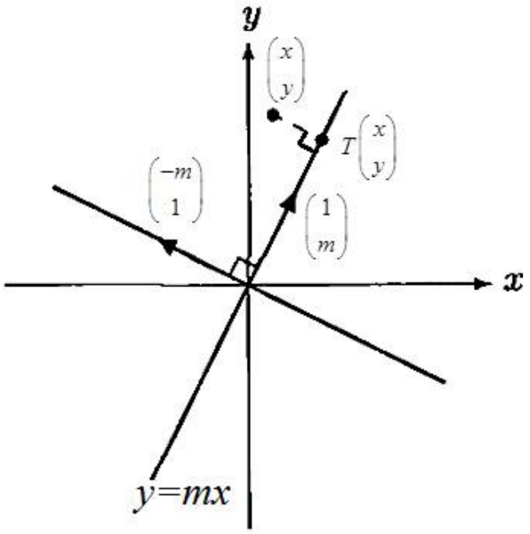
Example 5.1.6

Find the vector projection of $\begin{pmatrix} x \\ y \end{pmatrix}$ onto $\begin{pmatrix} 1 \\ m \end{pmatrix}$.

Solution

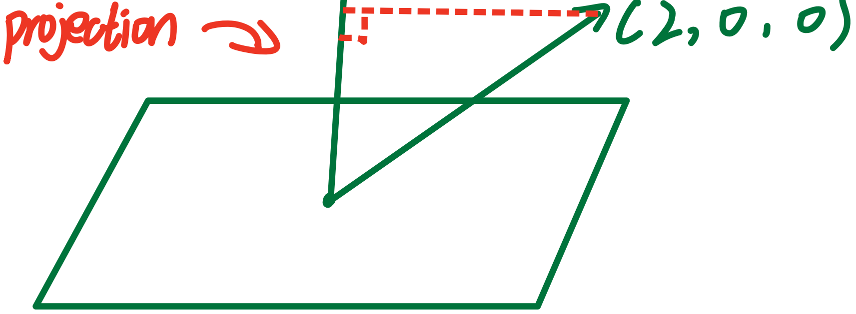
$$\boldsymbol{p} = \frac{(x,y)\begin{pmatrix} 1 \\ m \end{pmatrix}}{(1,m)\begin{pmatrix} 1 \\ m \end{pmatrix}}\begin{pmatrix} 1 \\ m \end{pmatrix} = \frac{x+my}{1+m^2}\begin{pmatrix} 1 \\ m \end{pmatrix} = \begin{pmatrix} \dfrac{1}{1+m^2} & \dfrac{m}{1+m^2} \\ \dfrac{m}{1+m^2} & \dfrac{m^2}{1+m^2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

which is the same as $T\begin{pmatrix} x \\ y \end{pmatrix}$ in Example 4.2.9



(1, 2, 2)

Simply as projection ⇒



Example 5.1.7

Find the distance from the point $(2,0,0)^T$ to the plane $x+2y+2z=0$.

Solution

The vector $\mathbf{N}=(1,2,2)^T$ is normal to the plane and the plane passes through the origin. Let $\boldsymbol{v}=(2,0,0)^T$. The distance d from $(2,0,0)^T$ to the plane is simply the absolute value of the scalar projection of $\boldsymbol{v}$ onto $\mathbf{N}$. Thus,

$$d=\frac{|\boldsymbol{v}^T\mathbf{N}|}{\|\mathbf{N}\|}=\frac{2}{3}.$$

Definition 5.1.8 orthogonal subspace

Two subspaces X and Y of R^n are said to be **orthogonal** if $\boldsymbol{x}^T\boldsymbol{y}=0$ for every $\boldsymbol{x}\in X$ and $\boldsymbol{y}\in Y$. If X and Y are orthogonal, we write $X\perp Y$.

Definition 5.1.9 orthogonal complement

Let Y be a subspace of R^n. The set of all vectors in R^n that are orthogonal to every vector in Y will be denoted $Y^\perp$. Thus Y所有能正交的向量 ⇒ 集

$$Y^\perp=\{\boldsymbol{x}\in R^n \mid \boldsymbol{x}^T\boldsymbol{y}=0 \text{ for every } \boldsymbol{y}\in Y\}$$

The set $Y^\perp$ is called the **orthogonal complement** of Y.

Example 5.1.10

Let $X=span\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix}\right\}$ and $Y=span\left\{\begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$ be subspace of R^3. For any $\begin{pmatrix}x_1\\x_2\\0\end{pmatrix}\in X$ and $\begin{pmatrix}0\\0\\y_3\end{pmatrix}\in Y$, $(x_1,x_2,0)\begin{pmatrix}0\\0\\y_3\end{pmatrix}=0$. Thus $X\perp Y$, i.e., $X=Y^\perp$.

$\dim X = 2$, $\dim Y = 1$

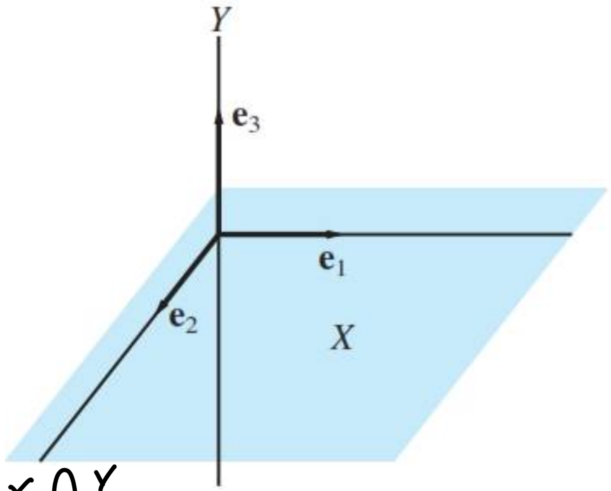
$\dim X + \dim Y = 3$

X不一定等于 $Y^\perp$

若 $x = span\begin{bmatrix}1\\0\\0\end{bmatrix}$

$y = span\begin{bmatrix}0\\0\\1\end{bmatrix}$

$X \perp Y$ 但 $X \neq Y^\perp$ 因为没有找全



suppose $\vec{z} \in X \cap Y$

$\vec{z} \in X$ $\vec{z} \in Y$ and $X \perp Y$

Remark 5.1.11

$\Rightarrow \vec{z} \cdot \vec{z} = 0 \Rightarrow \|\vec{z}\| = 0 \Rightarrow \vec{z} = \vec{0}$

(i) If X and Y are orthogonal subspaces of R^n, then $X \cap Y = \{\mathbf{0}\}$.

(ii) If Y is a subspace of R^n, $Y^\perp$ is also a subspace of R^n.

Proof of subspace:
① $Y^\perp \subseteq R^n$
② $\vec{0} \in Y^\perp$
③ $Y^\perp$ closed under "+" and "·"

Theorem 5.1.12

If W is a subspace of R^n, then $\dim W + \dim W^\perp = n$. Furthermore, if $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r\}$ is a basis for W and $\{\boldsymbol{x}_{r+1}, \boldsymbol{x}_{r+2}, \cdots, \boldsymbol{x}_n\}$ is a basis for $W^\perp$, then $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_r, \boldsymbol{x}_{r+1}, \cdots, \boldsymbol{x}_n\}$ is a basis for R^n.

$W \sqcup W^\perp$ = whole $\quad W \cap W^\perp = \{0\}$

Theorem 5.1.13

If A is an $m \times n$ matrix, then $N(A)^\perp = \mathrm{Col}\left(A^T\right)$ and $N\left(A^T\right) = \mathrm{Col}(A)^\perp$.

$N(A)^\perp = \mathrm{Col}(A^T) = \mathrm{Row}(A)$

$N(A^T)^\perp = \mathrm{Col}(A) = \mathrm{Row}(A^T)$

Example 5.1.14

Let

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{pmatrix}$$

Find the bases for $N(A), \mathrm{Col}(A^T), N(A^T), \mathrm{Col}(A)$ and verify Theorem 5.1.13.

Solution

The reduced row echelon form of A is

$$\begin{pmatrix}1&1&2\\0&1&1\\1&3&4\end{pmatrix}\xrightarrow{R_3-R_1\to R_3}\begin{pmatrix}1&1&2\\0&1&1\\0&2&2\end{pmatrix}\xrightarrow{R_3-2R_2\to R_3}\begin{pmatrix}1&1&2\\0&1&1\\0&0&0\end{pmatrix}\xrightarrow{R_1-R_2\to R_1}\begin{pmatrix}1&0&1\\0&1&1\\0&0&0\end{pmatrix}$$

Since $(1,0,1)$ and $(0,1,1)$ form a basis of Row(A), $\left\{\begin{pmatrix}1\\0\\1\end{pmatrix},\begin{pmatrix}0\\1\\1\end{pmatrix}\right\}$ forms a basis of $\mathrm{Col}\left(A^T\right)$.

Let $\boldsymbol{x}\in N(A)$. We have $x_1+x_3=0$ and $x_2+x_3=0$. Thus, $x_1=x_2=-x_3$.

$N(A)=\left\{\begin{pmatrix}\alpha\\\alpha\\-\alpha\end{pmatrix}\middle|\alpha\in R\right\}$ and $\begin{pmatrix}1\\1\\-1\end{pmatrix}$ forms a basis of $N(A)$.

Clearly, $\mathrm{Col}\left(A^T\right)=N(A)^{\perp}$.

The reduced row echelon form of A^T is

$$\begin{pmatrix}1&0&1\\1&1&3\\2&1&4\end{pmatrix}\xrightarrow{R_2-R_1\to R_2}\begin{pmatrix}1&0&1\\0&1&2\\2&1&4\end{pmatrix}\xrightarrow{R_3-2R_1\to R_3}\begin{pmatrix}1&0&1\\0&1&2\\0&1&2\end{pmatrix}\xrightarrow{R_3-R_2\to R_3}\begin{pmatrix}1&0&1\\0&1&2\\0&0&0\end{pmatrix}$$

Since $(1,0,1)$ and $(0,1,2)$ form a basis of $\mathrm{Row}\left(A^T\right)$, $\left\{\begin{pmatrix}1\\0\\1\end{pmatrix},\begin{pmatrix}0\\1\\2\end{pmatrix}\right\}$ forms a basis of $\mathrm{Col}(A)$.

Let $\boldsymbol{x}\in N(A^T)$. We have $x_1+x_3=0$ and $x_2+2x_3=0$. Thus, $x_1=-x_3, x_2=-2x_3$.

$N(A^T)=\left\{\begin{pmatrix}\alpha\\2\alpha\\-\alpha\end{pmatrix}\middle|\alpha\in R\right\}$ and $\begin{pmatrix}1\\2\\-1\end{pmatrix}$ forms a basis for $N(A^T)$.

Clearly, $N\left(A^T\right)=\mathrm{Col}(A)^{\perp}$.

$\vec{x}\cdot\vec{y}=0 \Rightarrow \vec{x}\perp\vec{y}$

$\vec{x}\cdot\vec{y}=0,\ \forall\vec{y}\in Y \Rightarrow \vec{x}\perp Y$

$\vec{x}\cdot\vec{y}=0,\ \forall\vec{x}\in X,\ \forall\vec{y}\in Y \Rightarrow X\perp Y$

$Y^{\perp}=\{\vec{x}\in R^n \mid \vec{x}\perp Y\}$

$V=\mathrm{span}\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_r\}$

$A=\begin{bmatrix}\vec{x}_1^T\\ \vec{x}_2^T\\ \vdots\\ \vec{x}_r^T\end{bmatrix}\Rightarrow A^T=[\vec{x}_1, \vec{x}_2 \dots \vec{x}_r$

$Col(A^T) = Span\{\vec{x_1}, \vec{x_2} .. \vec{x_r}\} = V$

$= N(A)^{\perp}$.

Section 5.2 Least Squares Problems

A least squares problem can generally be formulated as an overdetermined linear system of equations (more equations than unknowns). Such systems are usually inconsistent. Thus given an $m \times n$ system $A\boldsymbol{x} = \boldsymbol{b}$ with $m > n$, we cannot expect in general to find a vector $\boldsymbol{x} \in R^n$ for which $A\boldsymbol{x} = \boldsymbol{b}$. Instead, we can look for a vector $\hat{\boldsymbol{x}} \in R^n$ for which $A\hat{\boldsymbol{x}}$ is closest to $\boldsymbol{b}$.

Definition 5.2.1

If A is $m \times n$ and $\boldsymbol{b}$ is in R^m, a **least-squares solution** of $A\boldsymbol{x} = \boldsymbol{b}$ is an $\hat{\boldsymbol{x}}$ in R^n such that

$$\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\| \le \|\boldsymbol{b} - A\boldsymbol{x}\|$$

for all $\boldsymbol{x}$ in R^n.

① $A\vec{x} = \vec{b}$ inconsistent
$\min \|A\vec{x} - \vec{b}\|$ $x \in R^n$
⇒ 最优直线

② $A\vec{x} \in col(A) = \{c_1\vec{x_1} + c_2\vec{x_2} + \cdots c_n\vec{x_n}$
$A\vec{x} = \vec{b}$ 当 b in the plane
$A\vec{x} = \vec{b}$ consistent

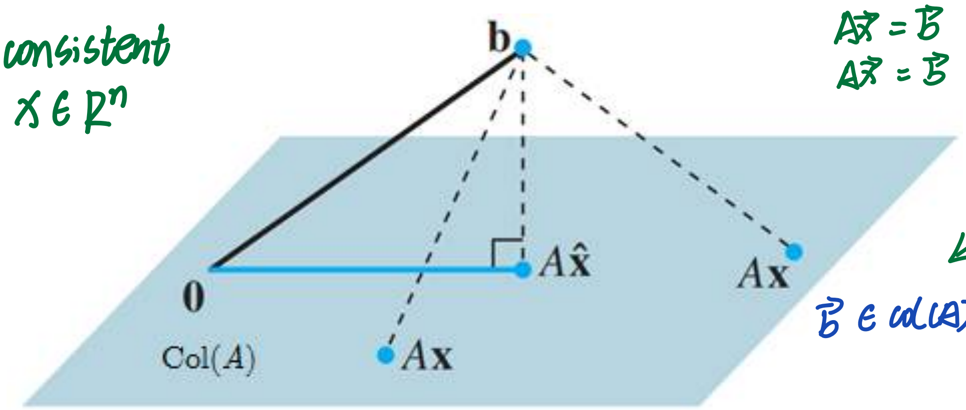
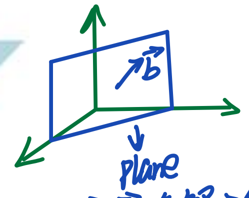


FIGURE 1 The vector **b** is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other **x**.



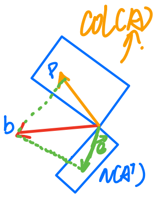
$\vec{b} \in col(A)$ $A\vec{x} = \vec{b}$ 有解 ⇒ 集合论

Theorem 5.2.2

Let A be a $m \times n$ matrix with $m > n$. For each $\boldsymbol{b} \in R^m$, there is a unique vector $\boldsymbol{p} = A(\hat{\boldsymbol{x}}) \in \text{Col}(A)$ such that

$$\|\boldsymbol{b} - \boldsymbol{v}\| > \|\boldsymbol{b} - A\hat{\boldsymbol{x}}\|$$

for any $\boldsymbol{v} \ne A\hat{\boldsymbol{x}}$ in $\text{Col}(A)$. Furthermore, $\boldsymbol{b} - A\hat{\boldsymbol{x}} \in \text{Col}(A)^{\perp} = N\left(A^T\right)$.



↗ projection matrix
Pb
↑
$\vec{p} + \vec{e} = \vec{b}$ → $(I - P)\, b$
↓
projection matrix

$\vec{b}$ project onto two space $\begin{cases} col(A) \\ N(A^T) \end{cases}$

Find projection of $\vec{b}$ onto col(A) $\Rightarrow A\hat{x}$ { 代数视角：看相判
几何视角：$\vec{b}$ 落于以投影

$\Rightarrow b - A\hat{x} = \vec{e} \in N(A^T)$

$A^T(b - A\hat{x}) = 0$

$A^T b = A^T A\hat{x}$

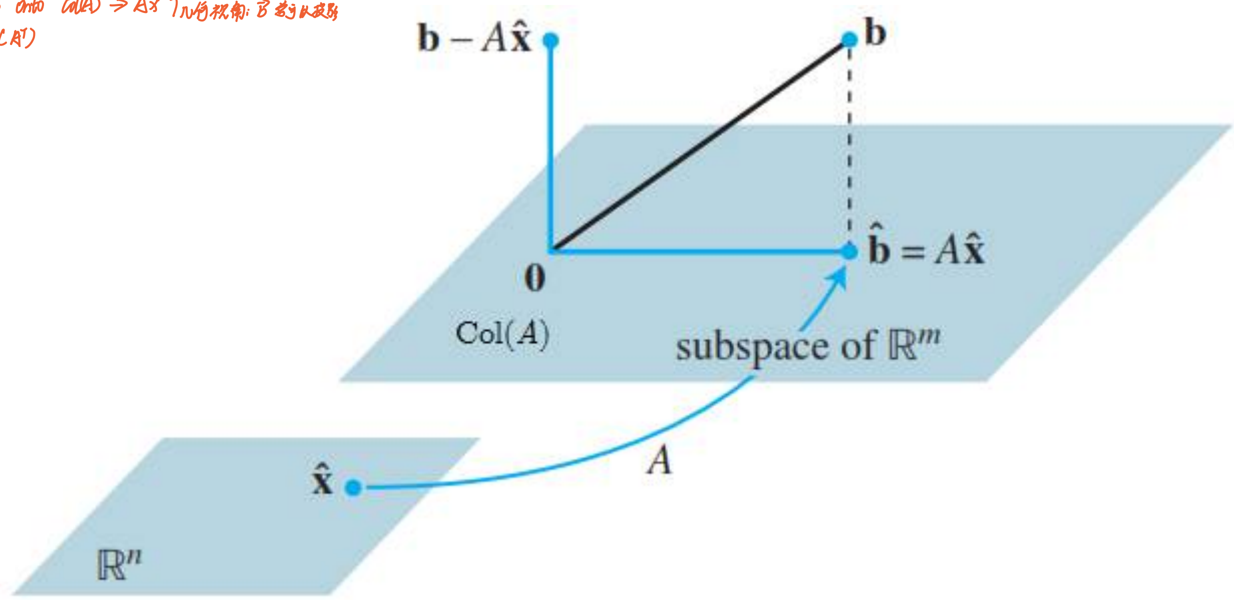


FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

Theorem 5.2.3

If A is an $m \times n$ matrix of rank n, $\hat{\boldsymbol{x}} = \left(A^T A\right)^{-1} A^T \boldsymbol{b}$ is the unique solution of the system $A\boldsymbol{x} = \boldsymbol{b}$.

Outline of the proof

Since rank of $A = n$, $A^T A$ is invertible. By Theorem 5.2.2, we have

$$A^T(\boldsymbol{b} - A\hat{\boldsymbol{x}}) = \mathbf{0}$$
$$A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$$
$$\hat{\boldsymbol{x}} = \left(A^T A\right)^{-1} A^T\boldsymbol{b}.$$

Example 5.2.4

Find a least-squares solution of the inconsistent system $A\boldsymbol{x} = \boldsymbol{b}$ for

$$A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix}.$$

Determine the least-squares error in the least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$.

Prove : $A_{m\times n}$ has rank $= n$, then $A^T A$ is non-singular

$\Rightarrow$ non-singular ① $\det(A^T A)$ ② $A^T A\hat{x} = \vec{b}$ unique solution

③ $A^T A\vec{x} = 0$ has only trivial solution

框架. 理论↘

$U \perp W$, $U \cap W = \{0\}$: $\begin{cases} A\vec{x} \in \text{col}(A) = N(A^T)^{\perp} \\ A\vec{x} \in N(A^T) \end{cases}$ $\Rightarrow A\vec{x} = 0$

$\Rightarrow x_1\vec{a_1} + x_2\vec{a_2} + \cdots x_n\vec{a_n} = 0$ 线性表示

∠a 有线性无关列向量>

$x_1 = x_2 \cdots = x_n = 0 \Rightarrow \vec{x} = \vec{0}$

$A^T A$ is non-singular

Solution

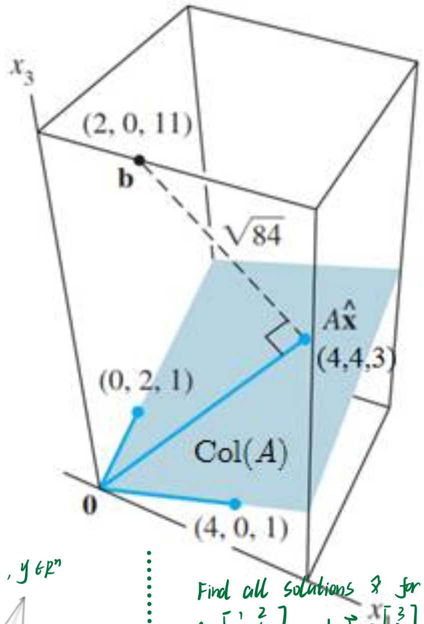
We have

$$A^T A = \begin{pmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{pmatrix}\begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix}, \quad A^T \boldsymbol{b} = \begin{pmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix} = \begin{pmatrix} 19 \\ 11 \end{pmatrix}.$$

$$\begin{aligned} \hat{\boldsymbol{x}} &= \left(A^T A\right)^{-1} A^T \boldsymbol{b} \\ &= \begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix}^{-1}\begin{pmatrix} 19 \\ 11 \end{pmatrix} \\ &= \frac{1}{84}\begin{pmatrix} 5 & -1 \\ -1 & 17 \end{pmatrix}\begin{pmatrix} 19 \\ 11 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 2 \end{pmatrix} \end{aligned}$$

Clearly, $A\hat{\boldsymbol{x}} = \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \\ 3 \end{pmatrix}$. The least-squares error is

$$\left\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\right\| = \left\|\begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix} - \begin{pmatrix} 4 \\ 4 \\ 3 \end{pmatrix}\right\| = \left\|\begin{pmatrix} -2 \\ -4 \\ 8 \end{pmatrix}\right\| = \sqrt{84}.$$

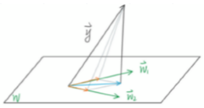


let W be a subspace of R^n with an $\boxed{\text{orthogonal basis}}$ $\{\vec{w_1}, \vec{w_2}\}$, $y \in R^n$

orthogonal projection of $\vec{y}$ onto W in R^n?

$\hat{y} = \text{proj}_W \vec{y} = c_1 \vec{w_1} + c_2 \vec{w} = \text{proj}_{w_1} \vec{y} + \text{proj}_{w_2} \vec{y}$

$= \frac{w_1^T y}{\|\vec{w_1}\|^2} w_1 + \frac{w_2^T y}{\|w_2\|^2} w_2$



eg. Find the closet point to $\vec{y}$ in $W = \text{span}\{\vec{u_1}, \vec{u_2}\}$ // shortest distance from $\vec{y}$ to w

$\vec{u_1} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ $\vec{u_2} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ $\vec{y} = \begin{bmatrix} 2 \\ 4 \\ 0 \\ 2 \end{bmatrix}$ $\vec{u_1} \perp \vec{u_2}$, $\{\vec{u_1}, \vec{u_2}\}$ from an orthogonal

$\text{Prg}_W Y = \text{Proj}_{\vec{u_1}} \vec{y} + \text{Brg}_{\vec{u_2}} \vec{y} = \frac{\vec{y}^T \vec{u_1}}{\|\vec{u_1}\|^2} \vec{u_1} + \frac{\vec{y}^T \vec{u_2}}{\|\vec{u_2}\|^2} \vec{u_2} = \begin{bmatrix} 3 \\ 3 \\ 1 \\ 1 \end{bmatrix}$

least-square error: $\|\vec{y} - \text{Prg}_W \vec{y}\| = \left\|\begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}\right\| = 2$ which is shortest distand from y to W

Find all solutions $\hat{x}$ for the following $A\vec{x} = \vec{b}$

$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ -1 & -2 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$

Remark: The least-square solution is not unique now.

① 将 $\vec{b}$ project 到 A 的 basis 中去 得到 $\hat{b}$

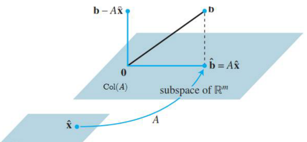
再求 $A\hat{x} = \hat{b}$ （类似右边的题）

$\text{Col}(A) = \text{Span}\left\{\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}\right\}$ $\hat{b} = \frac{\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}^T \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}}{\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}^T \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}} \cdot \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$

$A\hat{x} = \hat{b} \Rightarrow \hat{x} = \begin{bmatrix} 1-2x_2 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c\begin{bmatrix} -2 \\ 1 \end{bmatrix}$

$\hat{x} = \text{span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \end{bmatrix}\right\}$

利用 $A^TA\hat{x} = A^Tb \Rightarrow$ Col(A) linearly independent



基本思路：将 $\vec{b}$ project 到 col(A) 上 得到 $\hat{b}$（取出 A 中列向量的 basis 然后 $\vec{b}$ project 到 一个一个 basis 上 最后相加得到 $\hat{b}$

$y = e^{kx+b} \Rightarrow \ln y = z = kx + b$

→变成线性去做

Given a table of data

x	x_1	x_2	$\cdots$	x_m
y	y_1	y_2	$\cdots$	y_m

Consider a line $y = \beta_0 + \beta_1 x$. We call y_j the j-th *observed value* of y and $\beta_0 + \beta_1 x_j$ the j-th *predicted value*. The difference between an observed value and a predicted value is called a *residual*.

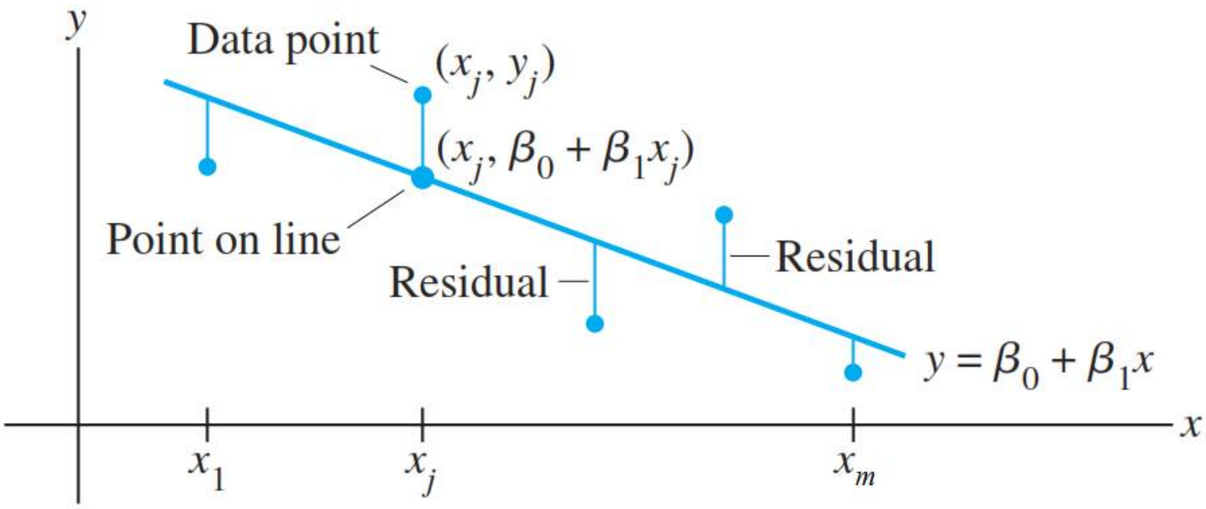


FIGURE 1 Fitting a line to experimental data.

We wish to find a linear function $y = \beta_0 + \beta_1 x$ that minimizes the sum of squares of the residuals. That means we want to find a least-squares solution of the inconsistent system of linear equations

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

Let $A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}$, $\boldsymbol{x} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$, $\boldsymbol{b} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$. Clearly,

Rank A = Rank A^TA ⇒ conclusion

10. Let A be an 8×5 matrix of rank 3, and let **b** be a nonzero vector in $N(A^T)$.

(a) Show that the system $A\mathbf{x} = \mathbf{b}$ must be inconsistent.

(b) How many least squares solutions will the system $A\mathbf{x} = \mathbf{b}$ have? Explain.

b> If A has rank 3 Then A^TA has rank 3 The normal equations are always consistent and in this case there will be 2 free variable. So the least square problem will have infinitely solutions

a> by consistency Theorem $A\vec{x} = \vec{b}$ is consistent if and only if b is in col(A). we given $\vec{b}$ is in $N(A^T)$. So if the system is consistent then $\vec{b}$ would be $col(A) \cap N(A^T) = \{0\}$ Since $\vec{b} \neq 0$, the system must be consistent

7. Given a collection of points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, let

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)^T \qquad \mathbf{y} = (y_1, y_2, \ldots, y_n)^T$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

and let $y = c_0 + c_1 x$ be the linear function that gives the best least squares fit to the points. Show that if $\bar{x} = 0$, then

$$c_0 = \bar{y} \quad \text{and} \quad c_1 = \frac{\mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{x}}$$

$$\Rightarrow \bar{x} = 0 \Rightarrow \sum_{i=1}^{n} x_i = 0$$

$$A^TA = \begin{bmatrix} n & \sum_{j=1}^{n} x_j \\ \sum_{j=1}^{n} x_j & \sum_{j=1}^{n} x_j^2 \end{bmatrix} = \begin{bmatrix} n & 0 \\ 0 & x^Tx \end{bmatrix} \qquad A^Tb = \begin{bmatrix} \sum_{j=1}^{n} y_i \\ \sum_{j=1}^{n} x_jy_j \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ x^Ty \end{bmatrix}$$

ai WONG

clearly: $c_0 = \bar{y} \quad c_1 = \dfrac{x^Ty}{x^Tx}$

$$A^TA = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_m \end{pmatrix}\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix} = \begin{pmatrix} m & \sum_{j=1}^{m} x_j \\ \sum_{j=1}^{m} x_j & \sum_{j=1}^{m} x_j^2 \end{pmatrix}, \quad A^T\boldsymbol{b} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_m \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{m} y_j \\ \sum_{j=1}^{m} x_jy_j \end{pmatrix}.$$

$$\hat{\boldsymbol{x}} = \left(A^TA\right)^{-1} A^T\boldsymbol{b}$$

$$= \begin{pmatrix} m & \sum_{j=1}^{m} x_j \\ \sum_{j=1}^{m} x_j & \sum_{j=1}^{m} x_j^2 \end{pmatrix}^{-1}\begin{pmatrix} \sum_{j=1}^{m} y_j \\ \sum_{j=1}^{m} x_jy_j \end{pmatrix}.$$

Example 5.2.5

Given the data

$y = \beta_0 + \beta_1 x$

x	0	3	6
y	1	4	5

Find the best least squares fit by a linear function. Determine the least-squares error.

Solution

Since $m = 3 \quad \sum_{j=1}^{3} x_j = 9 \quad \sum_{j=1}^{3} x_j^2 = 3^2 + 6^2 = 45 \quad \sum_{j=1}^{3} y_j = 10 \quad \sum_{j=1}^{3} x_jy_j = 12 + 30 = 42,$

$$\hat{\boldsymbol{x}} = \left(A^TA\right)^{-1} A^T\boldsymbol{b} = \begin{pmatrix} 3 & 9 \\ 9 & 45 \end{pmatrix}^{-1}\begin{pmatrix} 10 \\ 42 \end{pmatrix} = \frac{1}{54}\begin{pmatrix} 72 \\ 36 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 4 \\ 2 \end{pmatrix}$$

the best linear least-square fit:

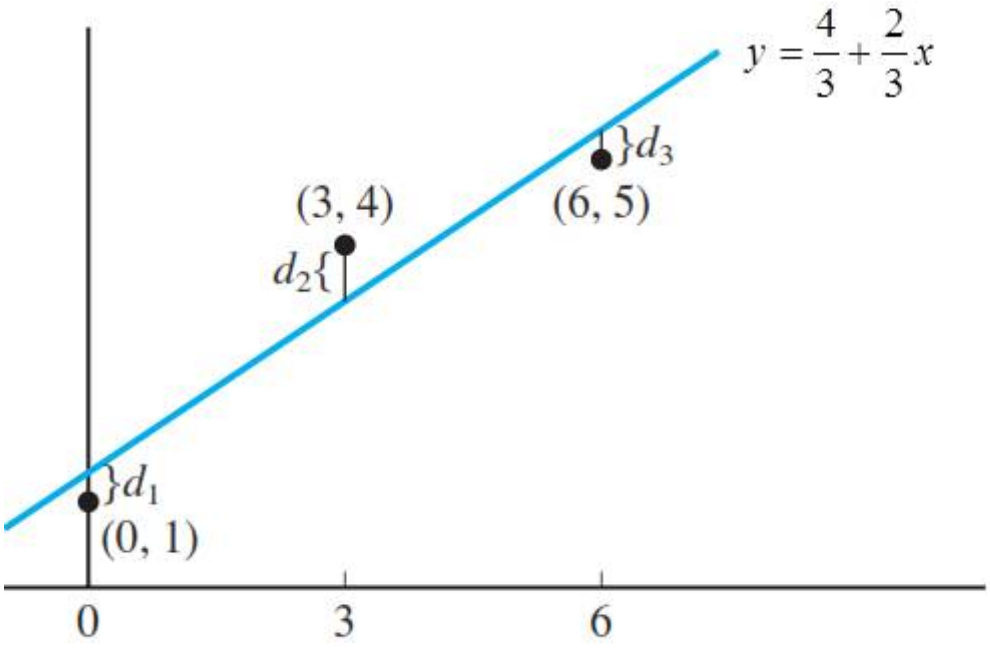
$y = \frac{4}{3} + \frac{2}{3}x$

perdition of y_i:

$\hat{y}_1 = \frac{4}{3} + \frac{2}{3} \cdot 0 = \frac{2}{3} < \frac{3}{3}$

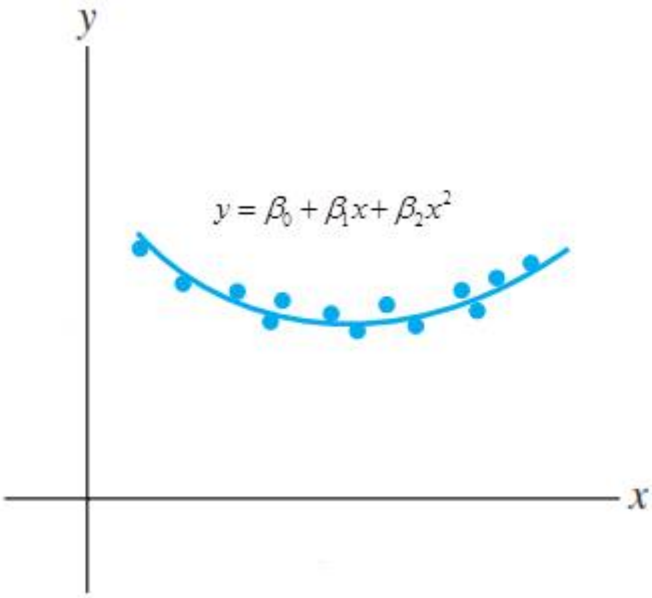
The best least squares fit is $y = \dfrac{4}{3} + \dfrac{2}{3}x.$

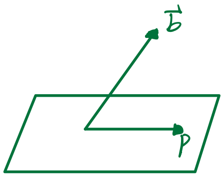
The least-squares error is $\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\| = \left\|\begin{pmatrix} 1 \\ 4 \\ 5 \end{pmatrix} - \frac{1}{3}\begin{pmatrix} 1 & 0 \\ 1 & 3 \\ 1 & 6 \end{pmatrix}\begin{pmatrix} 4 \\ 2 \end{pmatrix}\right\| = \left\|\begin{pmatrix} -\frac{1}{3} \\ \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}\right\| = \sqrt{\frac{2}{3}}.$



Similarly, we can find the best quadratic least squares fit $y = \beta_0 + \beta_1 x + \beta_2 x^2$ by considering the inconsistent system of linear equations

$$\begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$





vector $\vec{b}$ project onto col(A) [平面] How to do it

$A\vec{x} = \vec{b} \Rightarrow A^TA\hat{x} = A^T\vec{b} \Rightarrow \hat{x} = (A^TA)^{-1} \cdot A^Tb$

$\underset{\in col(A)}{A\hat{x}} = A(A^TA)^{-1}A^Tb = \underset{\text{Proj matrix}}{\underset{\downarrow}{P}}b \Rightarrow$ Projection matrix: $P = A(A^TA)^{-1}A^T$

9. Let A be an $m \times n$ matrix of rank n and let $P = A(A^TA)^{-1}A^T$. $R(A) = \text{Range}(A) = \text{col}(A)$

(a) Show that $P\mathbf{b} = \mathbf{b}$ for every $\mathbf{b} \in R(A)$. Explain this property in terms of projections.

(b) If $\mathbf{b} \in R(A)^\perp$, show that $P\mathbf{b} = \mathbf{0}$.

(c) Give a geometric illustration of parts (a) and (b) if $R(A)$ is a plane through the origin in $\mathbb{R}^3$.

$\vec{b} \in R(A) \Rightarrow \vec{b} \in \text{col}(A) \Rightarrow \vec{b} = A\vec{x}$

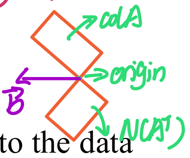
$P\vec{b} = A(A^TA)^{-1} \cdot A^T\vec{b} = A\underbrace{(A^TA)^{-1}A^TA}_{I\text{(单位矩阵)}}\vec{x} = A\vec{x} = \vec{b}$

$A^T\vec{b} = 0$

b> $\vec{b} \in R(A)^\perp \Rightarrow \vec{b} \in \text{col}(A)^\perp \Rightarrow \vec{b} \in N(A^T)$

$P\vec{b} = A(A^TA)^{-1} \cdot \underbrace{A^T\vec{b}}_{\vec{0}} = A(A^TA)^{-1} \cdot \vec{0} = \vec{0}$

c>



任何 $\vec{b}$ 都可以分解成 一个 $\vec{p} \in \text{col}(A)$，一个 $\vec{e} \in N(A^T)$ $\Rightarrow \vec{p} + \vec{e} = \vec{b}$

$\vec{p} \perp \vec{e}$

Example 5.2.6

Find the best quadratic least squares fit to the data

x	0	1	2	3
y	3	2	4	4

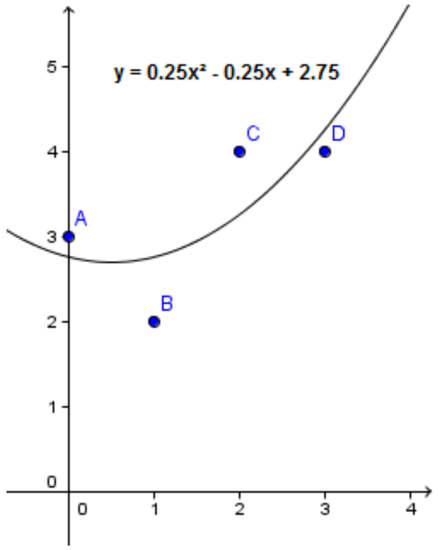
Determine the least-squares error.

Solution

$$\text{Let } A = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 4 \\ 4 \end{pmatrix}.$$

$$A^TA = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} = \begin{pmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{pmatrix}, \quad A^T\boldsymbol{b} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{pmatrix}\begin{pmatrix} 3 \\ 2 \\ 4 \\ 4 \end{pmatrix} = \begin{pmatrix} 13 \\ 22 \\ 54 \end{pmatrix}.$$

$$\begin{aligned} \hat{x} &= \left(A^TA\right)^{-1}A^Tb \\ &= \begin{pmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{pmatrix}^{-1}\begin{pmatrix} 13 \\ 22 \\ 54 \end{pmatrix} \\ &= \frac{1}{20}\begin{pmatrix} 19 & -21 & 5 \\ -21 & 49 & -15 \\ 5 & -15 & 5 \end{pmatrix}\begin{pmatrix} 13 \\ 22 \\ 54 \end{pmatrix} \\ &= \frac{1}{20}\begin{pmatrix} 55 \\ -5 \\ 5 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 11 \\ -1 \\ 1 \end{pmatrix} \end{aligned}$$



The best quadratic least squares fit is $y = 2.75 - 0.25x + 0.25x^2$.

$$\text{The least-squares error is } \|\boldsymbol{b} - A\hat{\mathbf{x}}\| = \left\|\begin{pmatrix} 3 \\ 2 \\ 4 \\ 4 \end{pmatrix} - \frac{1}{4}\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}\begin{pmatrix} 11 \\ -1 \\ 1 \end{pmatrix}\right\| = \left\|\begin{pmatrix} 0.25 \\ -0.75 \\ 0.75 \\ -0.25 \end{pmatrix}\right\| = \sqrt{\frac{5}{4}}.$$

b> For any vector z in $\mathbb{R}^n$

$Az = xy^Tz + yx^Tz = c_1x + c_2y$

$\begin{cases} c_1 = y^Tz & c_2 = x^Tz \end{cases}$

If z is in $N(A)$, then

$0 = Az = c_1x + c_2y \Rightarrow$ since x, y are linearly independent

$\Rightarrow y^Tz = c_1 = 0$ and $x^Tz = c_2 = 0$

$\Rightarrow z$ is orthogonal to both $\vec{x}$ and $\vec{y}$

Since x, y span S it follows that $z \in S^\perp$

conversely:

if z is in $S^\perp$ then z is orthogonal to both x and y

$\Rightarrow Az = c_1x + c_2y = 0$

Since $c_1 = y^Tz$ $c_2 = x^Tz = 0$

$\Rightarrow z$ is in $N(A) \Rightarrow N(A) = S^\perp$.

Clearly: $\dim S = 2$ and by theorem 5.2.2 $\dim S + \dim S^\perp = n$. use the result of A we have: $\dim N(A) = \dim S^\perp = n-2$

So A has nullity $n-2$ it follows from the Rank-Nullity Theorem that the rank of A must be 2.

10. (Optional) Let $\mathbf{x}$ and $\mathbf{y}$ be linearly independent vectors in $\mathbb{R}^n$ and let $S = \operatorname{span}(\mathbf{x}, \mathbf{y})$. We can use $\mathbf{x}$ and $\mathbf{y}$ to define a matrix A by setting

$$A = \mathbf{x}\mathbf{y}^T + \mathbf{y}\mathbf{x}^T$$

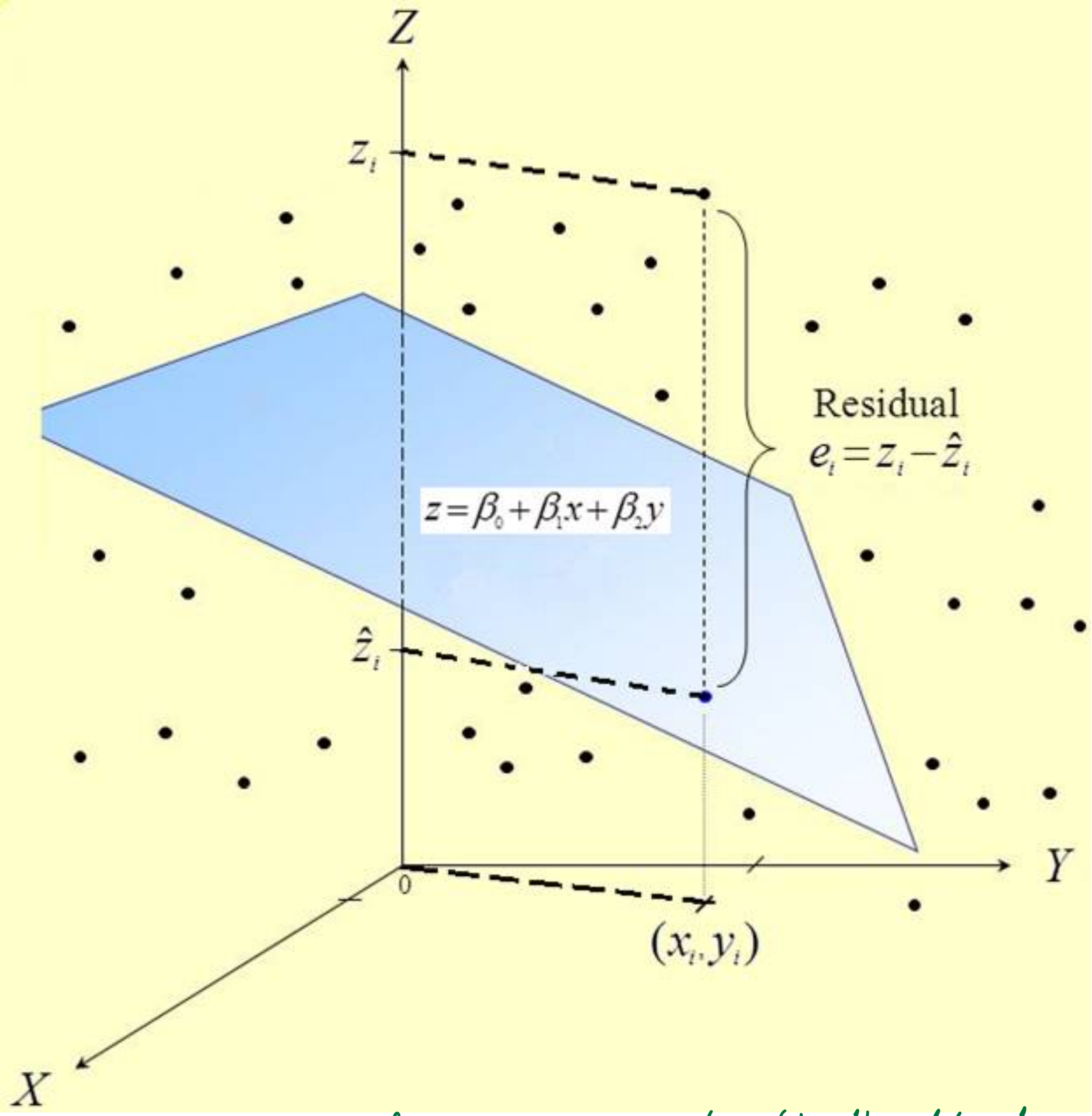
(a) Show that A is symmetric.

(b) Show that $N(A) = S^{\perp}$.

(c) Show that the rank of A must be 2.

We can also find the best fitted plane $z = \beta_0 + \beta_1 x + \beta_2 y$ by considering the inconsistent system of linear equations

$$\begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & y_m \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix}$$



Remark: Other than linear function, we can also fit the data to polynomials, trigonometric exponential function and etc

① Trigonometric Functions:

$$A = \begin{bmatrix} 1 & \sin x_1 & \cos x_1 \\ 1 & \sin x_2 & \cos x_2 \\ \vdots & \vdots & \vdots \\ 1 & \sin x_m & \cos x_m \end{bmatrix} \quad \vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad \vec{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \Leftrightarrow \left\{ \begin{array}{l} y_1 = \beta_0 + \beta_1 \sin x_1 + \beta_2 \cos x_1 \\ y_2 = \beta_0 + \beta_1 \sin x_2 + \beta_2 \cos x_2 \\ \vdots \\ y_m = \beta_0 + \beta_1 \sin x_m + \beta_2 \cos x_2 \end{array} \right.$$

3-D plane function:

$$A = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & y_m \end{bmatrix} \quad \vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad \vec{b} = \begin{bmatrix} \vec{z}_1 \\ \vec{z}_2 \\ \vdots \\ \vec{z}_n \end{bmatrix} \iff z = \beta_0 + \beta_1 x + \beta_2 y$$

quadratic polynomial (C = [illegible])

$$\left\{ y = \beta_0 + \beta_1 t + \beta_2 t^2 \ \ldots \right.$$

Example 5.2.7

Find the best fitted plane to the data

(x, y)	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
z	2	3	5	7

Determine the least-squares error.

Solution

Let $A = \begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \\ 1 & x_4 & y_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$, $\boldsymbol{x} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$, $\boldsymbol{b} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix}$.

$$A^T A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}, \quad A^T \boldsymbol{b} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix} = \begin{pmatrix} 17 \\ 10 \\ 12 \end{pmatrix}.$$

$$\begin{aligned} \hat{\boldsymbol{x}} &= \left(A^T A\right)^{-1} A^T \boldsymbol{b} \\ &= \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 17 \\ 10 \\ 12 \end{pmatrix} \\ &= \frac{1}{4} \begin{pmatrix} 3 & -2 & -2 \\ -2 & 4 & 0 \\ -2 & 0 & 4 \end{pmatrix} \begin{pmatrix} 17 \\ 10 \\ 12 \end{pmatrix} \\ &= \frac{1}{4} \begin{pmatrix} 7 \\ 6 \\ 14 \end{pmatrix} \end{aligned}$$

The best fitted plane is $z = 1.75 + 1.5x + 3.5y$

The least-squares error is $\left\| \boldsymbol{b} - A\hat{\boldsymbol{x}} \right\| = \left\| \begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix} - \frac{1}{4} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 7 \\ 6 \\ 14 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 0.25 \\ -0.25 \\ -0.25 \\ 0.25 \end{pmatrix} \right\| = \frac{1}{2}$.

V: vector space ⇒ set and operation $\{ +, \cdot$

$\begin{cases} \text{closed under two operation} \\ \text{laws} + \text{span } \{\vec{v}_1, \vec{v}_2, \cdots \vec{v}_k\} \end{cases}$

Section 5.3 Inner product space

Definition 5.3.1

An **inner product** on a vector space V is an operation on V that assigns, to each pair vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in V, a real number $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ satisfying the following conditions:

三大条件

Positive

I. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$ with equality if and only if $\boldsymbol{x} = \boldsymbol{0}$.

Symmetric

II. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in V.

$\langle \alpha \vec{x} + \beta \vec{y}, \vec{z} \rangle = \alpha \langle \vec{x}, \vec{z} \rangle + \beta \langle \vec{y}, \vec{z} \rangle \Leftrightarrow \begin{cases} \langle \alpha x, \vec{y} \rangle = \alpha \langle \vec{x}, \vec{y} \rangle \\ \langle \vec{x} + \vec{y}, \vec{z} \rangle = \langle \vec{x} + \vec{z} \rangle + \langle \vec{y} + \vec{z} \rangle \end{cases}$

Bilinear

III. $\langle \alpha \boldsymbol{x} + \beta \boldsymbol{y}, \boldsymbol{z} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \beta \langle \boldsymbol{y}, \boldsymbol{z} \rangle$ for all $\boldsymbol{x}, \boldsymbol{y}$ and $\boldsymbol{z}$ in V and all scalars α and β.

A vector space V with an inner product is called an **inner product space**.

orthogonal knowledge:

1. $\vec{x} \cdot \vec{y} = 0 \Rightarrow \vec{x} \perp \vec{y}$ (vector)
2. $\vec{x} \cdot \vec{y} = 0, \forall \vec{y} \in Y \Rightarrow \vec{x} \perp Y$ (vector and set)
3. $\vec{x} \cdot \vec{y} = 0, \forall \vec{x} \in X, \forall \vec{y} \in Y \Rightarrow X \perp Y$ (set and set)

Definition 5.3.2

The **standard inner product** for R^n is the scalar product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^T \boldsymbol{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$. Given a vector $\boldsymbol{w} = (w_1, \cdots, w_n)$ with positive entries, we could also define an inner product on R^n by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = w_1 x_1 y_1 + w_2 x_2 y_2 + \cdots + w_n x_n y_n.$$

The entries w_i are referred to as **weights**.

证明是 Inner product:

1. $\langle f, f \rangle = \int_a^b f^2(x)\,dx \Rightarrow \langle f, f \rangle = 0 \Rightarrow$ only $f = 0$

II. $\langle f, g \rangle = \int_a^b f(x) g(x) dx = \int_a^b g(x) f(x) dx = \langle g, f \rangle$

III. $\langle \alpha f + \beta g, h \rangle = \int_a^b (\alpha f + \beta g) h\, dx = \alpha \int_a^b f\, h\, dx + \beta \int_a^b g \cdot h \cdot dx$

Definition 5.3.3

We may define an inner product on $C[a,b]$, vector space of continuous functions on $[a,b]$, by

$$\langle f, g \rangle = \int_a^b f(x) g(x) dx$$

If $w(x) > 0$ and continuous on $[a,b]$, then

$$\langle f, g \rangle = \int_a^b w(x) f(x) g(x) dx$$

also defines an inner product on $C[a,b]$. The function $w(x)$ is called a **weight function**.

consider they are orthogonal or not

$\cos x \in [-\pi, \pi]$ $\sin x \in [-\pi, \pi]$

$\langle \cos x, \sin x \rangle = \int_{-\pi}^{\pi} \cos x \sin x\, dx$ (奇函数 + 区间对称 = 积分为0)

$\Rightarrow \cos x, \sin x$ are orthogonal in $[-\pi, \pi]$

Definition 5.3.4

Two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in an inner product space V are said to be **orthogonal** if $\langle \boldsymbol{u},\boldsymbol{v}\rangle = 0$. Two subspaces W_1 and W_2 of V are said to be **orthogonal** if $\langle \boldsymbol{u},\boldsymbol{v}\rangle = 0$ for every $\boldsymbol{u}\in W_1$ and $\boldsymbol{v}\in W_2$. If W_1 and W_2 are orthogonal, we write $W_1 \perp W_2$.

Theorem 5.3.5

For the vector space $C[-L,L]$, define an inner product

$$\langle f,g\rangle = \frac{1}{L}\int_{-L}^{L} f(x)g(x)dx.$$

奇函数

Let $W_1 = \{f \in C[-L,L] : f(-x) = -f(x) \text{ for all } x\in[-L,L]\}$ be subspace of odd functions and $W_2 = \{g \in C[-L,L] : g(x) = g(-x) \text{ for all } x\in[-L,L]\}$ be subspace of even functions, then $W_1 \perp W_2$. That means $\langle f,g\rangle = 0$ for any $f\in W_1$ and $g \in W_2$. 奇函数空间 ⊥ 偶函数空间.

偶函数

Proof

Let $f\in W_1$ and $g\in W_2$. We have

$$\begin{aligned}\langle f,g\rangle &= \frac{1}{L}\int_{-L}^{L} f(x)g(x)dx \quad [x=-y]\\ &= \frac{1}{L}\int_{-y=-L}^{-y=L} f(-y)g(-y)d(-y)\\ &= \frac{1}{L}\int_{y=L}^{y=-L} -f(y)g(y)(-dy)\\ &= \frac{-1}{L}\int_{-L}^{L} f(y)g(y)dy\\ &= -\langle f,g\rangle\end{aligned}$$

一换 ⇒ 全部跟着换.

$\langle f,g\rangle = 0.$

Definition 5.3.6

The **trace** of an $n\times n$ matrix M, denoted $Tr(M)$, is the sum of the diagonal entries of M; i.e.,

$$Tr(M) = M_{11} + M_{22} + \cdots + M_{nn}.$$

$$A=\begin{bmatrix} - & \vec{r_1} & - \\ - & \vec{r_2} & - \\ & \vdots & \\ - & \vec{r_n} & - \end{bmatrix}_{n\times m} \quad B=\begin{bmatrix} - & \vec{c_1} & - \\ - & \vec{c_2} & - \\ & \vdots & \\ - & \vec{c_n} & - \end{bmatrix}_{n\times m} \quad B^T=\begin{bmatrix} | & | & & | \\ \vec{c_1} & \vec{c_2} & \cdots & \vec{c_n} \\ | & | & & | \end{bmatrix}_{m\times n}$$

$$AB^T=\begin{bmatrix} \vec{r_1}\vec{c_1} & \vec{r_1}\vec{c_2} & \cdots & \vec{r_1}\vec{c_n} \\ \cdots & \vec{r_2}\vec{c_2} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \vec{r_n}\vec{c_n} \end{bmatrix}_{n\times n} \qquad Tr(AB^T) = \sum_{i=1}^{n} \vec{r_i}\vec{c_i}$$

Exercise:

$$A=\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad B=\begin{bmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}$$

$$Tr(AB^T) = 1\times 7 + 2\times 8 + \cdots\cdots 6\times 12.$$

Definition 5.3.7

Given $A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, B=\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} \in M_{m\times n}$, we can define an inner product by

$$\langle A,B\rangle=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}b_{ij}=Tr(AB^T).$$

$<A.B> = Tr(AB^T)$

1. $<A.A> = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2 \geqslant 0 \Rightarrow$ only A is 0 matrix

II. $<A.B> = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\cdot b_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} b_{ij}a_{ij} = <B.A>$

III. $<\alpha A+\beta B, C> = \sum_{i=1}^{m}\sum_{j=1}^{n}(\alpha a_{ij}+\beta b_{ij})c_{ij} = \alpha<A.C) + \beta<B,C>$

Definition 5.3.8

Let $x_1,\cdots,x_n$ be distinct real numbers. We may define an inner on P_n by

$\{x_i\}_{i=1}^{n}$ > distinct number.

$$\langle p,q\rangle=\sum_{i=1}^{n}p(x_i)q(x_i).$$

Given $x_1=0$ $x_2=1$ $x_3=-1$ Definite an inner product on P_3 as

$<P,q> = \sum_{i=1}^{3} P(x_i)q(x_i)$ $\forall P.q \in P_3$

Take $P(x)=x^2+2$ $q(x)=x+1$

Find $<P,q>$

$<P,q> = P(x_1)q(x_1) + P(x_2)q(x_2) + P(x_3)q(x_3)$

$=8.$

If $w(x)$ is a positive function, then

$$\langle p,q\rangle=\sum_{i=1}^{n}w(x_i)p(x_i)q(x_i)$$

also defines an inner product on P_n.

Definition 5.3.9

If $\boldsymbol{v}$ is a vector in an inner product space V, the **length** or **norm**, of $\boldsymbol{v}$ is given by

$$\|\boldsymbol{v}\|=\sqrt{\langle \boldsymbol{v},\boldsymbol{v}\rangle}.$$

Theorem 5.3.10 (Pythagorean Law)

If $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal vectors in an inner product space V, then $\|\boldsymbol{u}+\boldsymbol{v}\|^2=\|\boldsymbol{u}\|^2+\|\boldsymbol{v}\|^2$.

Example 5.3.11

For the vector space $C[-L,L]$, define an inner product as in Theorem 5.3.5. We obtain

		Projection	
space	Inner product	Scalar	vector
R^n	x^Tx	$\alpha=\frac{<\vec{u},\vec{v}>}{\|\vec{v}\|}$	$\vec{P}=\alpha\frac{\vec{v}}{\|\vec{v}\|}=\frac{<\vec{u},\vec{v}>}{<\vec{v},\vec{v}>}\vec{v}$
Function	$<f,g>=\int_{-L}^{L}f(x)g(x)dx$	$\alpha=\frac{<f,g>}{\|g\|}$	$\vec{P}=\alpha\cdot\frac{g}{\|g\|}$
matrix	$<A,B>=tr(AB^T)$	$\alpha=\frac{<A.B>}{\|B\|}$	$\vec{P}=\alpha\cdot\frac{B}{\|B\|}$

$\vec{P}$ 与 $\vec{v}$ 同东西

一般空间的推广

consider : $\{ \langle \frac{1}{\sqrt{2}}, \cos\frac{m\pi x}{L}, \sin\frac{n\pi x}{L} \}$ ......

$$\left\langle \frac{1}{\sqrt{2}}, \cos\frac{m\pi x}{L} \right\rangle = \frac{1}{\sqrt{2}L}\int_{-L}^{L} \cos\frac{m\pi x}{L}\,dx = \frac{1}{m\sqrt{2}\pi}\left[\sin\frac{m\pi x}{L}\right]_{-L}^{L} = 0$$

$$\left\langle \frac{1}{\sqrt{2}}, \sin\frac{m\pi x}{L} \right\rangle = \frac{1}{\sqrt{2}L}\int_{-L}^{L} \sin\frac{m\pi x}{L}\,dx = \frac{-1}{m\sqrt{2}\pi}\left[\cos\frac{m\pi x}{L}\right]_{-L}^{L} = 0$$

$$\left\langle \sin\frac{m\pi x}{L}, \cos\frac{n\pi x}{L} \right\rangle = 0 \quad \text{by Theorem 5.3.5}$$

For any distinct positive integers m and n,

very important !!!

$$\begin{aligned}\left\langle \cos\frac{m\pi x}{L}, \cos\frac{n\pi x}{L} \right\rangle &= \frac{1}{L}\int_{-L}^{L} \cos\frac{m\pi x}{L}\cos\frac{n\pi x}{L}\,dx \\ &= \frac{1}{2L}\int_{-L}^{L} \cos\frac{(m-n)\pi x}{L} + \cos\frac{(m+n)\pi x}{L}\,dx \\ &= \frac{1}{2(m-n)\pi}\left[\sin\frac{(m-n)\pi x}{L}\right]_{-L}^{L} + \frac{1}{2(m+n)\pi}\left[\sin\frac{(m+n)\pi x}{L}\right]_{-L}^{L} \\ &= 0\end{aligned}$$

$$\begin{aligned}\left\langle \sin\frac{m\pi x}{L}, \sin\frac{n\pi x}{L} \right\rangle &= \frac{1}{L}\int_{-L}^{L} \sin\frac{m\pi x}{L}\sin\frac{n\pi x}{L}\,dx \\ &= \frac{1}{2L}\int_{-L}^{L} \cos\frac{(m-n)\pi x}{L} - \cos\frac{(m+n)\pi x}{L}\,dx \\ &= \frac{1}{2(m-n)\pi}\left[\sin\frac{(m-n)\pi x}{L}\right]_{-L}^{L} + \frac{1}{2(m+n)\pi}\left[\sin\frac{(m+n)\pi x}{L}\right]_{-L}^{L} \\ &= 0\end{aligned}$$

Example 5.3.12 $\quad \|x\| = \sqrt{\langle x,x\rangle} = \sqrt{x_i^2}$

In P_5, define an inner product by as in Definition 5.3.8 with $x_i = i$ for i = 1,….,5.

$p(x) = x^2 \Rightarrow$

$$\|x\| = \sqrt{1+4+9+16+25} = \sqrt{55}.$$

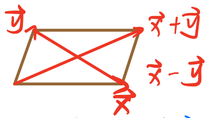
Definition 5.3.13 $\quad \{x_i = i\}_{i=1}^{5}$

Let $\boldsymbol{v} \neq \mathbf{0}$. Define scalar projection of $\boldsymbol{u}$ onto $\boldsymbol{v}$ by

$$\alpha = \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\|\boldsymbol{v}\|}$$

and vector projection of $\boldsymbol{u}$ onto $\boldsymbol{v}$ by

$$\mathbf{p} = \frac{\alpha}{\|\boldsymbol{v}\|}\boldsymbol{v} = \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\langle \boldsymbol{v}, \boldsymbol{v}\rangle}\boldsymbol{v}$$



$\Rightarrow$ 在 L_1-norm 里成立.

原理：原因满足平行四边形法则

$2\|\vec{x}\|^2 + 2\|\vec{y}\|^2 = \|\vec{x}+\vec{y}\|^2 + \|\vec{x}-\vec{y}\|^2$

$\Downarrow$

$\langle x+y, x+y\rangle + \langle x-y, x-y\rangle$

$\Downarrow$ $2\|\vec{x}\|^2 + 2\|\vec{y}\|^2$ $\quad$ L_2-norm correct

$\Rightarrow$ For L_1-norm space 里: $\vec{x} = (x_1, y_1)$ $\quad$ $\vec{y} = (x_2, y_2)$

$\|x\|_1 = |x_1| + |x_2|$

$2\|\vec{x}\|^2 + 2\|\vec{y}\|^2 = 2(|x_1|+|x_2|)^2 + 2(|x_2|+|y_2|)^2$

$\neq \|\vec{x}+\vec{y}\|^2 + \|\vec{x}-\vec{y}\|^2$

$\Uparrow$ 上面例子得到：通常 norm 不是由
inner space 诱导出来的，不满足平
行四边形法则

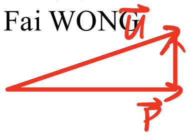
$\langle \vec{u}-\vec{p}, \vec{p} \rangle = \langle \vec{u} - \vec{p} \rangle - \langle \vec{p}, \vec{p} \rangle$

Prove

代数 $= \langle \vec{u}, \alpha \frac{\vec{v}}{\|\vec{v}\|} \rangle - \alpha^2 = \langle \vec{u}, \frac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{v}\|} \cdot \frac{\vec{v}}{\|\vec{v}\|} \rangle - \alpha^2$

$= \frac{\langle \vec{u}, \vec{v} \rangle^2}{\langle \vec{v}, \vec{v} \rangle} - \alpha^2 = \alpha^2 - \alpha^2 = 0$

: 几何:



$\vec{u} - \vec{p}$ clearly orthogonal

Remark 5.3.14

If $\boldsymbol{v} \neq \mathbf{0}$ and $\mathbf{p}$ is the vector projection of $\boldsymbol{u}$ onto $\boldsymbol{v}$, then

(i) $\boldsymbol{u} - \mathbf{p}$ and $\mathbf{p}$ are orthogonal.

$\vec{u} - \vec{p} = \vec{0}$

(ii) $\mathbf{u} = \mathbf{p}$ if and only if $\mathbf{u}$ is a scalar multiple of $\mathbf{v}$.

prove: If $\vec{u} = \beta \vec{v}$

$\vec{p} = \frac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{v}\|} \cdot \frac{\vec{v}}{\|\vec{v}\|} = \frac{\beta \langle \vec{v}, \vec{v} \rangle}{\|\vec{v}\|} \cdot \frac{\vec{v}}{\|\vec{v}\|} = \beta \vec{v} = \vec{u}$

Theorem 5.3.15

If $\boldsymbol{u}$ and $\boldsymbol{v}$ are any two vectors in an inner product space V, then

$|\alpha| \Rightarrow$ 作用在 constant 上

$\|\alpha\| \Rightarrow$ 作用在 vector 上

$$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\| \|\boldsymbol{v}\|$$

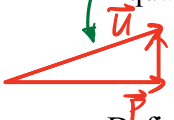
$\langle \sin x, \cos x \rangle$ angle

$\cos\theta = \frac{\langle \sin x, \cos x \rangle}{\|\sin x\| \|\cos x\|} = 0 \Rightarrow \frac{\pi}{2} = \theta$

$\langle \cos x, \cos x \rangle$ angle

$\cos\theta = \frac{\langle \cos x, \cos x \rangle}{\|\cos x\| \|\cos x\|} = 1 \Rightarrow 0 = \theta$

Equality holds if and only if $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly dependent.



$\vec{u} - \vec{p} \Rightarrow \|\vec{p}\|^2 = \frac{|\langle \vec{u}, \vec{v} \rangle|^2}{\|\vec{v}\|^2} = \|\vec{u}\|^2 - \|\vec{u} - \vec{p}\|^2$

$\Rightarrow |\langle \vec{u}, \vec{v} \rangle|^2 = \|\vec{u}\|^2 \|\vec{v}\|^2 - \|\vec{v}\|^2 \|\vec{u} - \vec{p}\|^2 \leq \|\vec{u}\|^2 \|\vec{v}\|^2$. $\Rightarrow \cos\theta = \frac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{u}\| \|\vec{v}\|}$

Definition 5.3.16

$\Rightarrow |\langle \vec{u}, \vec{v} \rangle| \leq \|\vec{u}\| \|\vec{v}\|$

A vector space V is said to be a **normed linear space** if, to each vector $\boldsymbol{v} \in V$, there is associated a real number $\|\boldsymbol{v}\|$, called the norm of $\boldsymbol{v}$, satisfying

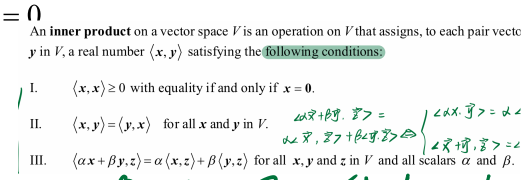
满足条件即为 norm (全部)

(i) $\|\boldsymbol{v}\| \geq 0$ with equality if and only if $\boldsymbol{v} = 0$

(ii) $\|\alpha \boldsymbol{v}\| = |\alpha| \|\boldsymbol{v}\|$ for any scalar α.

(iii) $\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$ for all $\boldsymbol{v}, \boldsymbol{w} \in V$.

不是 linearly independent



An **inner product** on a vector space V is an operation on V that assigns, to each pair vecto

$\boldsymbol{y}$ in V, a real number $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ satisfying the following conditions:

I. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$ with equality if and only if $\boldsymbol{x} = \mathbf{0}$.

II. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in V.

III. $\langle \alpha \boldsymbol{x} + \beta \boldsymbol{y}, \boldsymbol{z} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \beta \langle \boldsymbol{y}, \boldsymbol{z} \rangle$ for all $\boldsymbol{x}, \boldsymbol{y}$ and $\boldsymbol{z}$ in V and all scalars α and β.

$\langle \alpha \vec{x} + \beta \vec{y}, \vec{z} \rangle = \alpha \langle \vec{x}, \vec{z} \rangle + \beta \langle \vec{y}, \vec{z} \rangle \Leftrightarrow$ $\langle \alpha x, \vec{y} \rangle = \alpha \langle$; $\langle \vec{x} + \vec{y}, \vec{z} \rangle = \langle$

↑ 本质: 是一种 linearly independent.

The 3rd condition is called triangle inequality.

1. Any inner product space naturally has a norm.

$V \langle , \rangle$ $\forall \vec{v} \in V$, $\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle}$

$\vec{v} = 0$

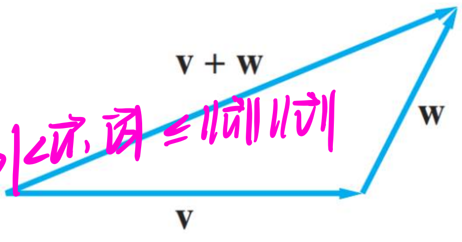
i. $\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle} \geq 0$ and $\|\vec{v}\| = 0 \Leftrightarrow \langle \vec{v}, \vec{v} \rangle = 0$

ii. $\|\alpha \vec{v}\| = \sqrt{\langle \alpha \vec{v}, \alpha \vec{v} \rangle} = \sqrt{\alpha^2 \langle \vec{v}, \vec{v} \rangle} = |\alpha| \sqrt{\langle \vec{v}, \vec{v} \rangle} = |\alpha| \|\vec{v}\|$

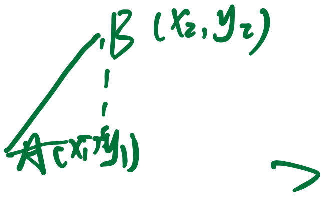
iii. $\|\vec{u} + \vec{v}\| = \sqrt{\langle \vec{u} + \vec{v}, \vec{u} + \vec{v} \rangle} = \sqrt{\langle \vec{u}, \vec{u} \rangle + \langle \vec{v}, \vec{v} \rangle + 2\langle \vec{u}, \vec{v} \rangle}$

$\leq \sqrt{\|\vec{u}\|^2 + \|\vec{v}\|^2 + 2\|\vec{u}\|\|\vec{v}\|}$

$= \|\vec{u}\| + \|\vec{v}\|$



$\Rightarrow |\langle \vec{u}, \vec{v} \rangle| \leq \|\vec{u}\| \|\vec{v}\|$



Definition 5.3.17

↗ 常给勾股定理.

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be vectors in a normed space. The **distance** between $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined to be the number $\|\boldsymbol{y} - \boldsymbol{x}\|$.

两个向量 independent 即不共线.

Tutorial:

a> 若 x and y are unit vector in $\mathbb{R}^n$ $|x^T y| = 1$ then x, y are linearly independent?

$|x^T y| = 1 \Rightarrow |\langle x, y \rangle| = 1$

$\cos\theta = \frac{\langle x, y \rangle}{\|x\| \|y\|} = \pm 1$

$\Rightarrow \theta = 0 / \pi \Rightarrow$ dependent.

↗ Algebra: 平行四边形法则

b> 若 u and v are vectors in an inner product space V then

$|\|\vec{u}\| - \|\vec{v}\|| \le \|\vec{u} \pm \vec{v}\|$ ⇒ square: $|\|\vec{u}\| - \|\vec{v}\||^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\|\vec{u}\|\|\vec{v}\| \le \|\vec{u}\|^2 + \|\vec{v}\|^2 \pm 2\vec{u}\vec{v}$

⇒ $-2\|\vec{u}\|\|\vec{v}\| \le \pm 2\vec{u}\vec{v}$ ⇒ $1 \le \cos\theta \le 1$ correct.

Theorem 5.3.18

If V is an inner product space, then

$$\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle} \text{ for all } \boldsymbol{v} \in V$$

defines a norm on V.

Example 5.3.19

For the vector space $C[-L, L]$, define an inner product as in Theorem 5.3.5. Define a norm

$$\|f\| = \sqrt{\frac{1}{L}\int_{-L}^{L} f(x)^2\, dx}.$$

Then for any positive integer k,

$$\left\|\frac{1}{\sqrt{2}}\right\| = \sqrt{\frac{1}{L}\int_{-L}^{L}\frac{1}{2}\,dx} = \sqrt{\frac{2L}{2L}} = 1$$

$$\left\|\cos\frac{n\pi x}{L}\right\| = \sqrt{\frac{1}{L}\int_{-L}^{L}\cos^2\frac{n\pi x}{L}\,dx} = \sqrt{\frac{1}{2L}\int_{-L}^{L}\left(1+\cos\frac{2n\pi x}{L}\right)dx} = \sqrt{\frac{2L}{2L}+\frac{1}{4n\pi}\left[\sin\frac{2n\pi x}{L}\right]_{-L}^{L}} = 1$$

$$\left\|\sin\frac{n\pi x}{L}\right\| = \sqrt{\frac{1}{L}\int_{-L}^{L}\sin^2\frac{n\pi x}{L}\,dx} = \sqrt{\frac{1}{2L}\int_{-L}^{L}\left(1-\cos\frac{2n\pi x}{L}\right)dx} = \sqrt{\frac{2L}{2L}-\frac{1}{4n\pi}\left[\sin\frac{2n\pi x}{L}\right]_{-L}^{L}} = 1$$

Definition 5.3.20

函有绝对值.

Let $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T \in R^n$. Define

$\|\vec{x}\| = \sqrt{\sum_{i=1}^{\infty} x_i^2}$ → L_2-norm

→ Lp norm

$$\|\boldsymbol{x}\|_p = \begin{cases} \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} & \text{if } 1 \le p < \infty \\ \max_{1 \le i \le n} |x_i| & \text{if } p = \infty \end{cases}$$

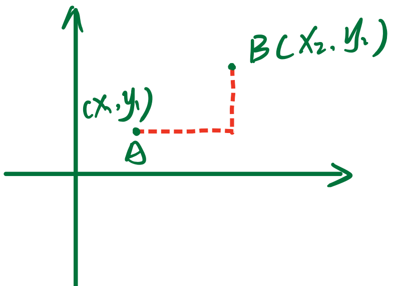
$\|\cdot\|_p$ can be shown as the norm in R^n. $\|\cdot\|_2$ is derived from the inner product in Definition 5.3.2. When $p = \infty$, $\|\cdot\|_\infty$ is called the **uniform norm** or **infinity norm**.

Example 5.3.21

Let $\boldsymbol{x} = (4, -5, 3)^T \in R^3$. Compute $\|\boldsymbol{x}\|_1$, $\|\boldsymbol{x}\|_2$, and $\|\boldsymbol{x}\|_\infty$.

Solution

$$\|\boldsymbol{x}\|_1 = |4| + |-5| + |3| = 12, \quad \|\boldsymbol{x}\|_2 = \sqrt{16+25+9} = 5\sqrt{2}, \quad \|\boldsymbol{x}\|_\infty = \max(|4|, |-5|, |3|) = 5.$$



$\|AB\|_2 = \sqrt{(x_2-x_0)^2 + (x_1-x_0)^2}$ → L_1-norm

$\|AB\|_1 = |x_2 - x_1| + |y_2 - y_1|$ → L_2-norm why should we need to use orthogonal sets?

Example: $R^3 \Rightarrow$ basis: $\begin{bmatrix}1\\0\\0\end{bmatrix} \begin{bmatrix}0\\1\\0\end{bmatrix} \begin{bmatrix}0\\0\\1\end{bmatrix} \Rightarrow \begin{bmatrix}1\\4\\5\end{bmatrix} = 1\begin{bmatrix}1\\0\\0\end{bmatrix} + 4\begin{bmatrix}0\\1\\0\end{bmatrix} + 5\begin{bmatrix}0\\0\\1\end{bmatrix}$ 很快得到系数.

$R^3 \Rightarrow$ basis: $\begin{bmatrix}1\\0\\0\end{bmatrix} \begin{bmatrix}1\\1\\0\end{bmatrix} \begin{bmatrix}1\\1\\1\end{bmatrix} \Rightarrow \begin{bmatrix}1\\4\\5\end{bmatrix} = -3\begin{bmatrix}1\\0\\0\end{bmatrix} + (-1)\begin{bmatrix}1\\1\\0\end{bmatrix} + 5\begin{bmatrix}1\\1\\1\end{bmatrix}$ 需要花时间

Section 5.4 Orthonormal Sets

Definition 5.4.1

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ be nonzero vectors in an inner product space V. If $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle = 0$ whenever $i \neq j$, then $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ is said to be an **orthogonal set** of vectors.

Example 5.4.2

The set $\{(1,1,1)^T, (2,1,-3)^T, (4,-5,1)^T\}$ is an orthogonal set in R^3 since

$$\langle (1,1,1)^T, (2,1,-3)^T \rangle = 0$$
$$\langle (1,1,1)^T, (4,-5,1)^T \rangle = 0$$
$$\langle (2,1,-3)^T, (4,-5,1)^T \rangle = 0$$

Theorem 5.4.3

If $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ is an orthogonal set of nonzero vectors in an inner product space V, then $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ are linearly independent.

Definition 5.4.4

An **orthonormal set** of vectors is an orthogonal set of unit vectors.

1. 相互垂直.
2. 单位向量.

The set $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n\}$ is orthonormal if and only if $\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = \delta_{ij}$ where $\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

Given any orthogonal set of nonzero vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$, it is possible to form an orthonormal set by defining

$$\boldsymbol{u}_i = \left(\frac{1}{\|\boldsymbol{v}_i\|}\right)\boldsymbol{v}_i \quad \text{for } i = 1, 2, \cdots, n$$

orthogonal set $\xrightarrow[\|norm\|]{devide}$ orthonormal set

Example 5.4.5

In Example 5.4.2, $\left\{(1,1,1)^T, (2,1,-3)^T, (4,-5,1)^T\right\}$ is an orthogonal set. To form an orthonormal set, let

$$\boldsymbol{u}_1 = \left(\frac{1}{\|\boldsymbol{v}_1\|}\right)\boldsymbol{v}_1 = \frac{1}{\sqrt{3}}(1,1,1)^T \quad \boldsymbol{u}_2 = \left(\frac{1}{\|\boldsymbol{v}_2\|}\right)\boldsymbol{v}_2 = \frac{1}{\sqrt{14}}(2,1,-3)^T \quad \boldsymbol{u}_3 = \left(\frac{1}{\|\boldsymbol{v}_3\|}\right)\boldsymbol{v}_3 = \frac{1}{\sqrt{42}}(4,-5,1)^T$$

$\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ forms an orthonormal set.

prove:

$\forall \vec{u} \in V$, $\vec{u} = c_1\vec{u_1} + c_2\vec{u_2} + \cdots c_n\vec{u_n}$

$\vec{v} = d_1\vec{v_1} + d_2\vec{v_2} + \cdots d_n\vec{v_n}$

$\langle \vec{u}, \vec{u_1} \rangle = \sum_{i=1}^{n} c_i \langle u_i, u_1 \rangle \Rightarrow$ orthogonal set

$= c_1 \cdot \|\vec{u_1}\|^2 = c_1$

conclusion: $\langle \vec{u}, \vec{u_k} \rangle = c_k$

$\langle \vec{u}, \vec{v} \rangle = \sum_{\substack{i=1 \\ j=1}}^{n} c_i d_j \langle \vec{u_i}, \vec{v_j} \rangle = \sum_{\substack{i=1 \\ j=1}}^{n} c_i d_j$

Theorem 5.4.6

Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n\}$ be orthonormal basis for an inner product space V. If $\boldsymbol{u} = \sum_{i=1}^{n} a_i\boldsymbol{u}_i$ and $\boldsymbol{v} = \sum_{i=1}^{n} b_i\boldsymbol{u}_i$, then $a_i = \langle \boldsymbol{u}, \boldsymbol{u}_i \rangle$, $b_i = \langle \boldsymbol{v}, \boldsymbol{u}_i \rangle$, $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \sum_{i=1}^{n} a_i b_i$ and $\|\boldsymbol{u}\|^2 = \sum_{i=1}^{n} a_i^2$.

类似 projection

☆ $\|\vec{u}\|^2 = \langle \vec{u}, \vec{u} \rangle = \sum_{\substack{i=1 \\ j=1}}^{n} c_i c_j \langle \vec{u_i}, \vec{u_j} \rangle = \sum_{i=1}^{n} c_i^2$

Example 5.4.7

Express $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ as linear combination of $\boldsymbol{u}_1 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $\boldsymbol{u}_2 = \frac{1}{\sqrt{14}}\begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}$, $\boldsymbol{u}_3 = \frac{1}{\sqrt{42}}\begin{pmatrix} 4 \\ -5 \\ 1 \end{pmatrix}$.

Solution

Since $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \in R^3 = span\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$, $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = a_1\boldsymbol{u}_1 + a_2\boldsymbol{u}_2 + a_3\boldsymbol{u}_3$. We have

$$a_1 = \left\langle (1,2,3)^T, \boldsymbol{u}_1 \right\rangle = \frac{6}{\sqrt{3}} \quad a_2 = \left\langle (1,2,3)^T, \boldsymbol{u}_2 \right\rangle = \frac{-5}{\sqrt{14}} \quad a_3 = \left\langle (1,2,3)^T, \boldsymbol{u}_3 \right\rangle = \frac{-3}{\sqrt{42}}$$

Hence $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{6}{\sqrt{3}}\boldsymbol{u}_1 - \frac{5}{\sqrt{14}}\boldsymbol{u}_2 - \frac{3}{\sqrt{42}}\boldsymbol{u}_3$.

利用 orthogonal 快速求得比例系数

$\frac{c_0}{\sqrt{2}} \cdot 1$

$$f(x) = C_0 \cdot \frac{1}{\sqrt{2}} + \sum_{n=1}^{\infty} C_n \frac{\cos n\pi x}{2} + \sum_{n=1}^{\infty} d_n \frac{\sin n\pi x}{2}$$

orthonormal set

Theorem 5.4.8 (Fourier Series)

In $C[-L,L]$, define an inner product as in Theorem 5.3.5. By Example 5.3.11 and 5.3.19, the set $\left\{\frac{1}{\sqrt{2}}, \cos\frac{\pi x}{L}, \sin\frac{\pi x}{L}, \cdots, \cos\frac{n\pi x}{L}, \sin\frac{n\pi x}{L}, \cdots\right\}$ is an orthonormal set of vectors. Let f be a function on $[-L,L]$. Suppose

vector

$$f(x) = \frac{a_0}{2} + a_1\cos\frac{\pi x}{L} + b_1\sin\frac{\pi x}{L} + \cdots + a_n\cos\frac{n\pi x}{L} + b_n\sin\frac{n\pi x}{L} + \cdots$$

Then the same as before.

无论 1与 $\frac{1}{\sqrt{2}}$ 都本质没有区别

$$a_0 = \frac{1}{L}\int_{-L}^{L}\frac{a_0}{2}\,dx = \langle f, 1\rangle = \frac{1}{L}\int_{-L}^{L} f(x)\,dx$$

$$a_n = \left\langle f, \cos\frac{n\pi x}{L}\right\rangle = \frac{1}{L}\int_{-L}^{L} f(x)\cos\frac{n\pi x}{L}\,dx \quad \text{for } n = 1, 2, \cdots$$

$$b_n = \left\langle f, \sin\frac{n\pi x}{L}\right\rangle = \frac{1}{L}\int_{-L}^{L} f(x)\sin\frac{n\pi x}{L}\,dx \quad \text{for } n = 1, 2, \cdots$$

Definition 5.4.9

$$\langle f, \tfrac{1}{\sqrt{2}}\rangle = C_0 \cdot 1$$

The coefficient a_n and b_n are called the Fourier coefficients of f.

$f = C_0 \cdot \frac{1}{\sqrt{2}} + \sum_{n=1}^{\infty} C_n \cos\frac{n\pi x}{L} + \sum_{n=1}^{\infty} d_n \sin\frac{n\pi x}{L}$

$\Rightarrow C_0 = \langle f, \frac{1}{\sqrt{2}}\rangle \quad C_n = \langle f, \cos\frac{n\pi x}{L}\rangle \quad d_n = \langle f, \sin\frac{n\pi x}{L}\rangle$

对 f is odd function. $C_0 = 0 \quad C_n = 0 \ n \geq 1$

$d_n = \langle f, \sin\frac{n\pi x}{L}\rangle = \frac{2}{L}\int_0^L f \cdot \sin\frac{n\pi x}{L}\,dx$

对 f is even function. $C_0 = \langle f, \frac{1}{\sqrt{2}}\rangle = \frac{2}{L}\int_0^L f \cdot \frac{1}{\sqrt{2}}\,dx$

$C_n = \langle f, \cos\frac{n\pi x}{L}\rangle = \frac{2}{L}\int_0^L f \cdot \cos\frac{n\pi x}{L}\,dx$

$d_n = 0 \quad n \geq 1$

Theorem 5.4.10

If $f \in C[-L,L]$ is an odd function then $a_n = 0$ for $n = 0, 1, 2, \cdots$. We have

$$f(x) = b_1\sin\frac{\pi x}{L} + \cdots + b_n\sin\frac{n\pi x}{L} + \cdots$$

If $f \in C[-L,L]$ is an even function then $b_n = 0$ for $n = 1, 2, \cdots$. We have

$$f(x) = \frac{a_0}{2} + a_1\cos\frac{\pi x}{L} + \cdots + a_n\cos\frac{n\pi x}{L} + \cdots$$

$[-\pi, \pi]$

Example 5.4.11

$L = \pi$

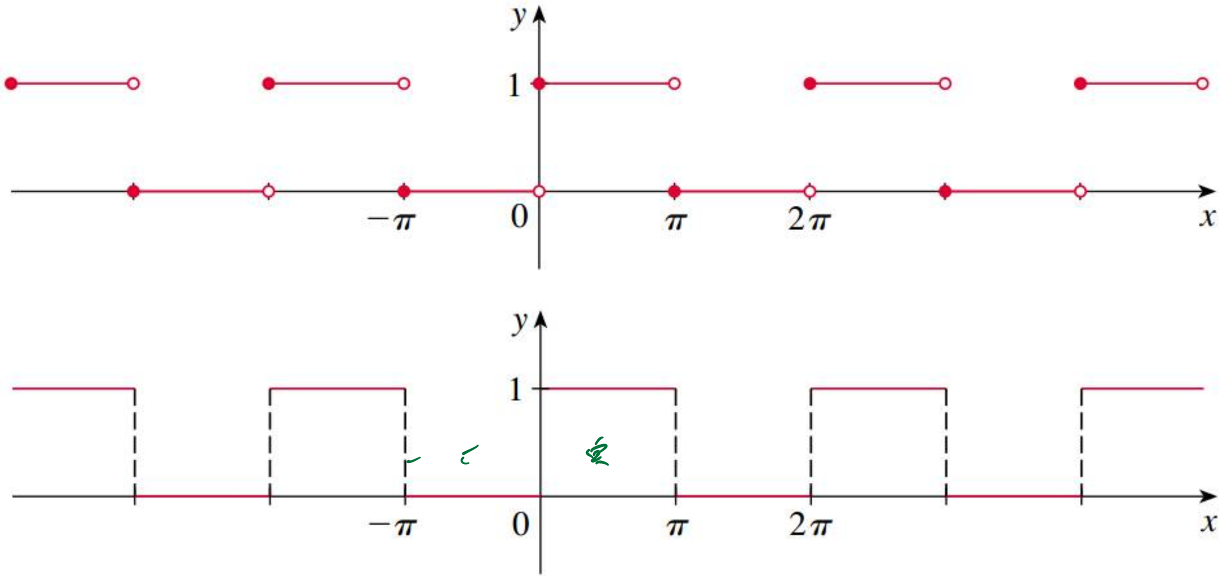
Find the Fourier coefficients and Fourier series of the square-wave function f defined by

代入 $L=\pi$

$$f(x) = \begin{cases} 0 & \text{if } -\pi \le x < 0 \\ 1 & \text{if } 0 \le x < \pi \end{cases} \quad \text{and} \quad f(x + 2\pi) = f(x)$$

$\Rightarrow$ discontinuous function

$$\begin{cases} \cos\frac{n\pi x}{L} = \cos\frac{n\pi x}{\pi} = \cos nx \\ \sin\frac{n\pi x}{L} = \sin\frac{n\pi x}{\pi} = \sin nx \end{cases}$$



Solution

$$a_0=\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)dx=\frac{1}{\pi}\int_{-\pi}^{0} f(x)dx+\frac{1}{\pi}\int_{0}^{\pi} f(x)dx=\frac{1}{\pi}\int_{-\pi}^{0} 0dx+\frac{1}{\pi}\int_{0}^{\pi} 1dx=1.$$

For $n\ge 1$, $\cos\frac{n\pi x}{L}$ (这里 $L=\pi$) $\uparrow$ $\searrow$ $\cos nx$

证明为什么可以这么用 $\Downarrow$ orthogonal set 缘故

$$a_n=\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nxdx$$

$$=\frac{1}{\pi}\int_{-\pi}^{0} f(x)\cos nxdx+\frac{1}{\pi}\int_{0}^{\pi} f(x)\cos nxdx$$

$$=\frac{1}{\pi}\int_{-\pi}^{0} 0dx+\frac{1}{\pi}\int_{0}^{\pi}\cos nxdx$$

$$=\frac{\sin nx}{n\pi}\Big|_0^{\pi}=0$$

$$\langle f(x),1\rangle=\langle\frac{a_0}{2},1\rangle=\frac{1}{L}\int_{-L}^{L}\frac{a_0}{2}\cdot 1\,dx=a_0$$

$$\langle f(x),\cos\frac{n\pi x}{L}\rangle=\langle a_n\cos\frac{n\pi x}{L},\cos\frac{n\pi x}{L}\rangle=a_n$$

$$\langle f(x),\sin\frac{n\pi x}{L}\rangle=\langle b_n\sin\frac{n\pi x}{L},\sin\frac{n\pi x}{L}\rangle=b_n$$

$$b_n=\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nxdx$$

$$=\frac{1}{\pi}\int_{-\pi}^{0} f(x)\sin nxdx+\frac{1}{\pi}\int_{0}^{\pi} f(x)\sin nxdx$$

$$=\frac{1}{\pi}\int_{-\pi}^{0} 0dx+\frac{1}{\pi}\int_{0}^{\pi}\sin nxdx$$

$$=-\frac{\cos nx}{n\pi}\Big|_0^{\pi}$$

Δ分类讨论

$$=\begin{cases}0 & \text{if } n \text{ is even}\\ \frac{2}{n\pi} & \text{if } n \text{ is odd}\end{cases}$$

1. orthogonal set
2. orthonormal set
3. orthogonal matrix

(列向量都是 column vector.)

a. $\vec{u}=\sum_{i=1}^{\infty} c_i\vec{u}_i$

$c_i=\langle\vec{u},\vec{u}_i\rangle$

b. = Fourier series

$\vec{u}_1=\begin{pmatrix}2\\5\\1\end{pmatrix}$ $\vec{u}_2=\begin{pmatrix}-2\\1\\1\end{pmatrix}$ $\vec{y}=\begin{pmatrix}1\\2\\3\end{pmatrix}$

$W=\operatorname{span}\{\vec{u}_1,\vec{u}_2\}$ the vector in W which is closet to $\vec{y}$

$\vec{p}=\frac{\langle\vec{y},\vec{u}_1\rangle}{\langle\vec{u}_1,\vec{u}_1\rangle}\cdot\vec{u}_1+\frac{\langle\vec{y},\vec{u}_2\rangle}{\langle\vec{u}_2,\vec{u}_2\rangle}\cdot\vec{u}_2=\frac{9}{30}\vec{u}_1+\frac{3}{6}\vec{u}_2$

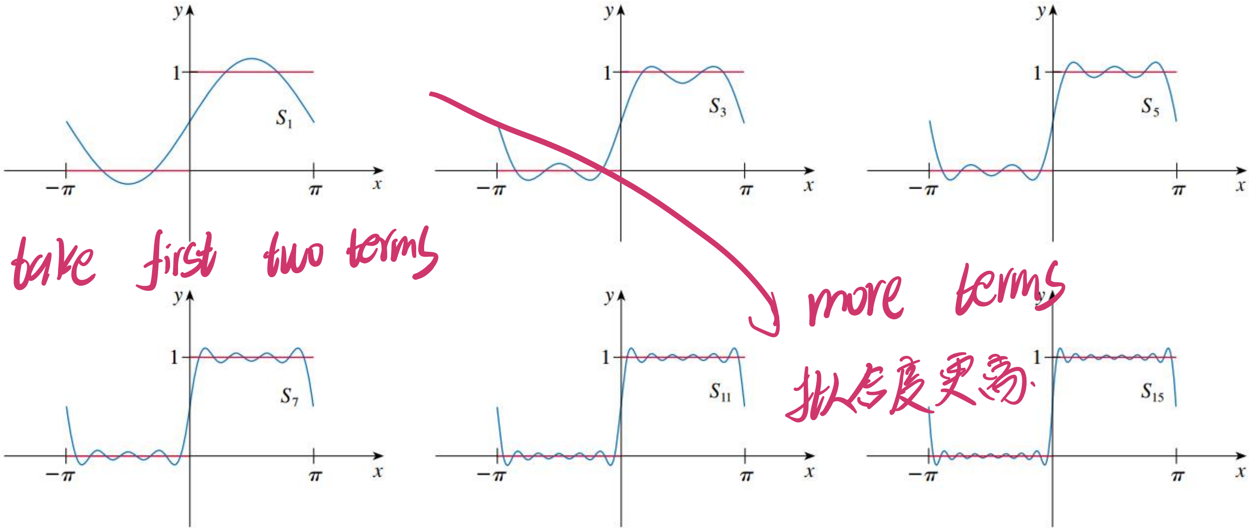
$\vec{u}_3=\begin{pmatrix}0\\6\\2\end{pmatrix}=\vec{u}_2+\vec{u}_1$ $W=\operatorname{span}\{\vec{u}_1,\vec{u}_3\rangle$

$\vec{p}'=\frac{\langle\vec{y},\vec{u}_1\rangle}{\langle\vec{u}_1,\vec{u}_1\rangle}\vec{u}_1+\frac{\langle\vec{y},\vec{u}_3\rangle}{\langle\vec{u}_3,\vec{u}_3\rangle}\vec{u}_3\neq\vec{p}$

$\searrow$ 只有 basis orthogonal 才能那么做.

Fourier series of f is

$$\frac{1}{2}+\frac{2}{\pi}\sin x+\frac{2}{3\pi}\sin 3x+\frac{2}{5\pi}\sin 5x+\cdots$$



take first two terms

more terms

拟合度更高.

Definition 5.4.12 column vector are orthonormal vector set.

An $n\times n$ matrix Q is said to be an **orthogonal matrix** if the column vectors of Q form an orthonormal set in R^n.

$<q_i, q_j> = \delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i\neq j \end{cases}$

Assume: $Q = (\vec{q_1}, \vec{q_2} \cdots \vec{q_n})$

Theorem 5.4.13

An $n\times n$ matrix Q is orthogonal if and only if $Q^TQ=I$.

$$Q^TQ = \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix} [\vec{q_1}, \vec{q_2} \cdots \vec{q_n}] = [<\vec{q_i}, \vec{q_j}>] = I.$$

$\Rightarrow Q^{-1} = Q^T$

It follows that from Theorem 5.4.13 that if Q is orthogonal, then Q is invertible and $Q^{-1}=Q^T$.

Q 相当于旋转 ↗ 不改变 x y 长度
↘ 不改变 x y 夹角

Theorem 5.4.14

If Q is an $n\times n$ orthogonal matrix, then

$$\langle Q\boldsymbol{x}, Q\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle \quad \text{and} \quad \|Q\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$$

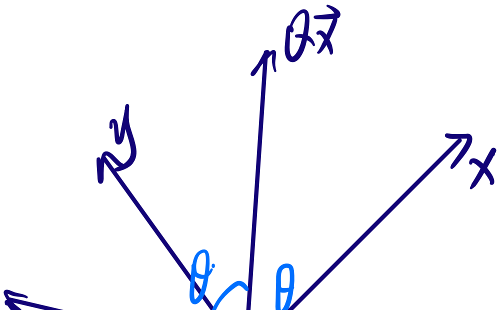
$\nearrow (Qx)^T \cdot Qy = x^Ty = <\vec{x}, \vec{y}>$

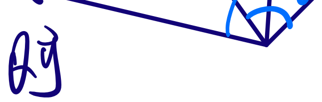
prove: $<Qx. Qy> = <\vec{x}, \vec{y}>$

$\downarrow$ $\|Qx\|_2 = \|x\|_2$ $\downarrow$

$\cancel{\|Q\vec{x}\|_2} \cancel{\|Q\vec{y}\|_2} \cdot \cos\hat{\theta} = \cancel{\|\vec{x}\|_2} \cancel{\|\vec{y}\|_2} \cos\theta$

conclusion : $\cos\hat{\theta} = \cos\theta$.





Conclusion

$Q\vec{x}, Q\vec{y}$'s angle $\hat{\theta}$ = $\vec{x}\,\vec{y}$'s angle θ.

$\langle Q\vec{x}, Q\vec{x}\rangle = (Q\vec{x})^T Q\vec{x} = x^T \cdot Q^T Q \cdot \vec{x} = x^T x$
$= \langle \vec{x}, \vec{x}\rangle$

orthogonal set

Example 5.4.15

Rotation matrix

Q ⇔ 一种 rotation

$$Q=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}$$

is orthogonal and

$$Q^{-1}=Q^{T}=\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}.$$

Rotation preserves length of vector and angle between 2 vectors (Example 4.1.16).

重新排序

△. Example 5.4.16

Permutation matrix is a matrix formed from the identity by reordering its column.

$$\begin{pmatrix}1&&\\&1&\\&&1\end{pmatrix},\begin{pmatrix}1&&\\&&1\\&1&\end{pmatrix},\begin{pmatrix}&1&\\1&&\\&&1\end{pmatrix},\begin{pmatrix}&1&\\&&1\\1&&\end{pmatrix},\begin{pmatrix}&&1\\1&&\\&1&\end{pmatrix},\begin{pmatrix}&&1\\&1&\\1&&\end{pmatrix}$$

3×3 permutation matrices

Permutation matrices are orthogonal.

Projection onto V_1

Theorem 5.4.17

Let W be a subspace of an inner product space V and let $\boldsymbol{x}\in V$. Let $\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_r\}$ be an orthogonal basis for W. If

Proj V_1 … Proj V_r.

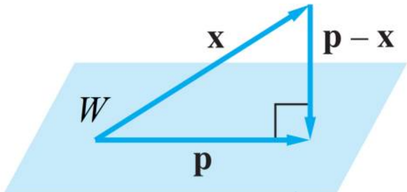
证明在下面

$$\boldsymbol{p}=\frac{\langle \boldsymbol{x},\boldsymbol{v}_1\rangle}{\langle \boldsymbol{v}_1,\boldsymbol{v}_1\rangle}\boldsymbol{v}_1+\frac{\langle \boldsymbol{x},\boldsymbol{v}_2\rangle}{\langle \boldsymbol{v}_2,\boldsymbol{v}_2\rangle}\boldsymbol{v}_2+\cdots+\frac{\langle \boldsymbol{x},\boldsymbol{v}_r\rangle}{\langle \boldsymbol{v}_r,\boldsymbol{v}_r\rangle}\boldsymbol{v}_r$$

then $\boldsymbol{x}-\boldsymbol{p}\in W^{\perp}$. $\boldsymbol{p}$ is called projection of $\boldsymbol{x}$ onto W, denoted by $\text{proj}_W \boldsymbol{x}$.

scalar projection:
$\alpha = \frac{\langle x, y\rangle}{\|y\|}$
vector projection:
$\vec{v} = \frac{\langle x, y\rangle}{\langle y, y\rangle}\cdot\vec{y}$

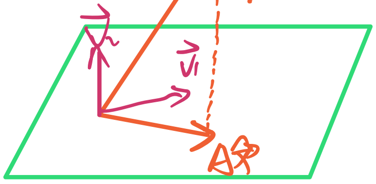
$A\vec{x} = \vec{b}$ (inconsistent)
↓ least-square
$\min\|A\hat{x} - \vec{b}\|$ $\vec{x} \in \mathbb{R}^n$



Give a example in $\mathbb{R}^3$ Projection $\vec{b}$ onto col(A)

$\text{col}(A) = \{A\vec{x} \mid \vec{x} \in \mathbb{R}^n\}$ ⇒ $A^TA\hat{x} = A^Tb$ ① Algebra. → 直接作直线

$\vec{b}$ $\vec{b} - A\hat{x}$ ⇒ $\text{Proj}\,\vec{v}_1 + \text{Proj}\,\vec{v}_2 + \text{Proj}\,\vec{v}_3 = A\hat{x}$



$\{V_1, V_2, V_3\}$ orthogonal set => project 到 basis 上
再相加得到

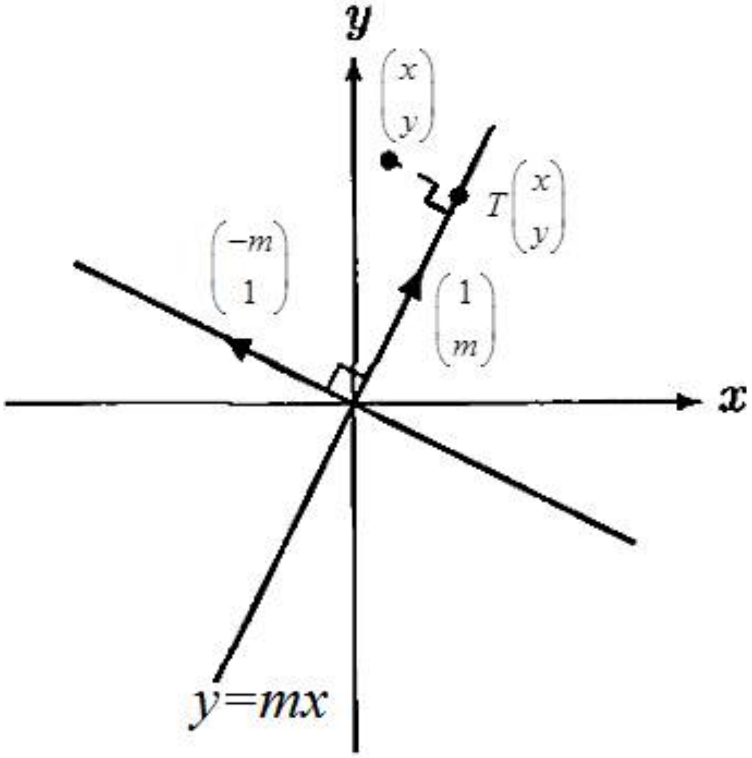
Example 5.4.18 (Projection)

Let $T:R^2\to R^2$ be linear transformation defined by $T\begin{pmatrix}1\\m\end{pmatrix}=\begin{pmatrix}1\\m\end{pmatrix}$ and $T\begin{pmatrix}-m\\1\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$. Then $T\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}\frac{1}{1+m^2} & \frac{m}{1+m^2}\\ \frac{m}{1+m^2} & \frac{m^2}{1+m^2}\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$. Let $W=span\left\{\begin{pmatrix}1\\m\end{pmatrix}\right\}$. Then $W^\perp=span\left\{\begin{pmatrix}-m\\1\end{pmatrix}\right\}$. Let $\boldsymbol{x}=\begin{pmatrix}x\\y\end{pmatrix}$.

In Example 5.1.6,

$$\boldsymbol{p}=\frac{(x,y)\begin{pmatrix}1\\m\end{pmatrix}}{(1,m)\begin{pmatrix}1\\m\end{pmatrix}}\begin{pmatrix}1\\m\end{pmatrix}=\frac{x+my}{1+m^2}\begin{pmatrix}1\\m\end{pmatrix}=\begin{pmatrix}\frac{1}{1+m^2} & \frac{m}{1+m^2}\\ \frac{m}{1+m^2} & \frac{m^2}{1+m^2}\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\mathrm{proj}_W\begin{pmatrix}x\\y\end{pmatrix}.$$

$$\begin{aligned}\boldsymbol{x}-\boldsymbol{p}&=\begin{pmatrix}x\\y\end{pmatrix}-\frac{x+my}{1+m^2}\begin{pmatrix}1\\m\end{pmatrix}\\
&=\begin{pmatrix}x-\frac{x+my}{1+m^2}\\ y-m\frac{x+my}{1+m^2}\end{pmatrix}\\
&=\begin{pmatrix}\frac{x(1+m^2)-(x+my)}{1+m^2}\\ \frac{y(1+m^2)-(mx+m^2y)}{1+m^2}\end{pmatrix}\\
&=\begin{pmatrix}\frac{xm^2-my}{1+m^2}\\ \frac{y-mx}{1+m^2}\end{pmatrix}\\
&=\frac{y-mx}{1+m^2}\begin{pmatrix}-m\\1\end{pmatrix}\\
&=\frac{(x,y)\begin{pmatrix}-m\\1\end{pmatrix}}{(-m,1)\begin{pmatrix}-m\\1\end{pmatrix}}\begin{pmatrix}-m\\1\end{pmatrix}\in W^\perp\end{aligned}$$



Indeed, $\boldsymbol{x}-\boldsymbol{p}=\mathrm{proj}_{W^\perp}\begin{pmatrix}x\\y\end{pmatrix}$ and $\boldsymbol{x}=\mathrm{proj}_W\begin{pmatrix}x\\y\end{pmatrix}+\mathrm{proj}_{W^\perp}\begin{pmatrix}x\\y\end{pmatrix}$.

$W = span\{\vec{v_1}, \vec{v_2} \cdots \vec{v_n}\}$

$\vec{x}-\vec{p} \in W^\perp$, $\vec{x}-\vec{p} \perp W \Rightarrow \forall \vec{w} \in W$, $\vec{x}-\vec{p} \perp \vec{w}$

$\vec{x}-\vec{p} \perp V_i \ (i=1,2,3\cdots n)$

$\langle \vec{x}-\vec{p}, \vec{v_i}\rangle = \langle \vec{x}, \vec{v_i}\rangle - \langle \vec{p}, \vec{v_i}\rangle = 0$

（证明 $\vec{x}-\vec{p}$ orthogonal to W
证明垂直误差最小）

$\vec{p} = \sum \frac{\langle x, \vec{v_i}\rangle}{} \vec{v_i}$ (orthogonal set)

[illegible] of R^n [illegible]

$\hat{p} = \sum_{j=1} \frac{}{\langle v_j, v_j \rangle} v_j$

$= \langle \vec{x}, \vec{v_i} \rangle - \langle \sum_{j=1} \frac{\langle x, \vec{v_j} \rangle}{\langle v_j, v_j \rangle} \vec{v_j}, v_i \rangle$

$= \langle \vec{x}, \vec{v_i} \rangle - \frac{\langle \vec{x}, \vec{v_i} \rangle}{\cancel{\langle v_i, v_i \rangle}} \cancel{\langle v_i, v_i \rangle} = 0$

W is a subspace of R, $y \in R^n$ $\hat{y} = proj_W y$

$\|\vec{y} - \hat{y}\| < \|\vec{y} - \vec{v}\|$ $\forall \vec{v} \in W$, $\vec{v} \neq \hat{y}$

$\vec{y} - \hat{y} \in W^\perp \Leftrightarrow \vec{y} - \hat{y} \perp W$ $\hat{y} - \vec{v} \in W \Rightarrow \vec{y} - \hat{y} \perp \hat{y} - \vec{v}$

Prove: $\|\vec{y} - \hat{y}\|^2 + \|\hat{y} - \vec{v}\|^2 = \|\vec{y} - \vec{v}\|^2$.

$\Rightarrow \|\vec{y} - \hat{y}\| < \|\vec{y} - \vec{v}\|$

⇒ 勾股定理.

projection of $\vec{v}$ onto W.
① Find shortest distance between $\vec{v}$ onto W
② Find orthogonal vector ($v - proj_W \vec{v}$) $\perp W$

Section 5.5 Gram-Schmidt Process

Given basis vector $\{x_1, x_2 \cdots x_n\}$ we want to estabish orthonormal vector $\{\vec{v_1}, \vec{v_2} \cdots \vec{v_n}\}$

① orthogonal set
② normlize it.

Theorem 5.5.1 (Gram-Schmidt Process)

Given a basis $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$ for a nonzero subspace W of V, define

$x_1 = v_1$

$x_2 = v_2 + \frac{\langle x_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1$

$x_3 = v_3 + \frac{\langle x_3, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2 + \frac{\langle x_3, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1$

$\vdots$

x_n could be represent by $\{v_1, v_2 \cdots v_n\}$

$$\boldsymbol{v}_1 = \boldsymbol{x}_1 \quad W_1 = span\{v_1\}$$

$$\boldsymbol{v}_2 = \boldsymbol{x}_2 - \frac{\langle \boldsymbol{x}_2, \boldsymbol{v}_1 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle} \boldsymbol{v}_1 \quad W_2 = span\{v_1, v_2\}$$

$$\boldsymbol{v}_3 = \boldsymbol{x}_3 - \frac{\langle \boldsymbol{x}_3, \boldsymbol{v}_1 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle} \boldsymbol{v}_1 - \frac{\langle \boldsymbol{x}_3, \boldsymbol{v}_2 \rangle}{\langle \boldsymbol{v}_2, \boldsymbol{v}_2 \rangle} \boldsymbol{v}_2$$

$$\vdots$$

$$\boldsymbol{v}_n = \boldsymbol{x}_n - \frac{\langle \boldsymbol{x}_n, \boldsymbol{v}_1 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle} \boldsymbol{v}_1 - \frac{\langle \boldsymbol{x}_n, \boldsymbol{v}_2 \rangle}{\langle \boldsymbol{v}_2, \boldsymbol{v}_2 \rangle} \boldsymbol{v}_2 - \cdots - \frac{\langle \boldsymbol{x}_n, \boldsymbol{v}_{n-1} \rangle}{\langle \boldsymbol{v}_{n-1}, \boldsymbol{v}_{n-1} \rangle} \boldsymbol{v}_{n-1}$$

$span\{x_1, x_2, x_3 \cdots x_n\}$

$\sqcup \ \sqcap \Rightarrow \parallel$

$span\{v_1, v_2, v_3 \cdots v_n\}$

normlize.

Then $\left\{ \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}, \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|}, \cdots, \frac{\boldsymbol{v}_n}{\|\boldsymbol{v}_n\|} \right\}$ is an orthonormal basis for W. In addition

$$span\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_k\} = span\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k\} \quad \text{for } 1 \le k \le n.$$

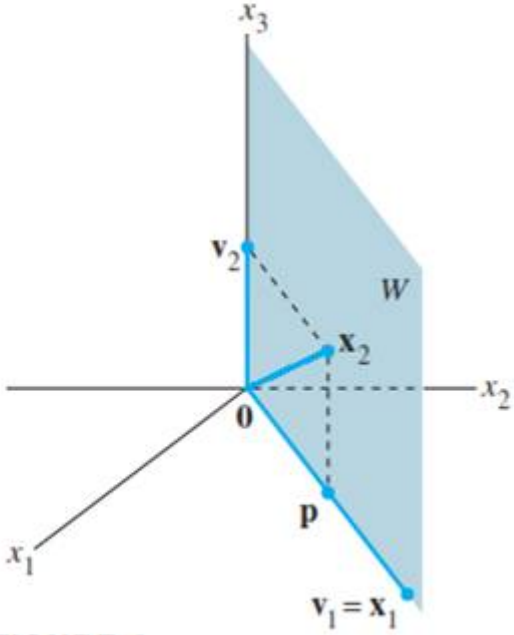


FIGURE 1

Construction of an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$.

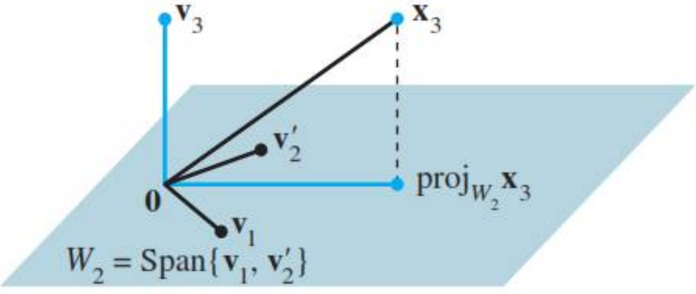


FIGURE 2 The construction of $\mathbf{v}_3$ from $\mathbf{x}_3$ and W_2.

remove noise. (课上一个问题)

$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos kx + b_k \sin kx$

$a_k \cos kx + b_k \cdot \sin kx = \left[\frac{a_k}{\sqrt{a_k^2 + b_k^2}} \cos kx + \frac{b_k}{\sqrt{a_k^2 + b_k^2}} \sin kx \right] \sqrt{a_k^2 + b_k^2}$

$= \cos(kx - \theta_k) \sqrt{a_k^2 + b_k^2}$

Example 5.5.2

Let $\boldsymbol{x}_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \boldsymbol{x}_2 = \begin{pmatrix}0\\1\\1\\1\end{pmatrix}$ and $\boldsymbol{x}_3 = \begin{pmatrix}0\\0\\1\\1\end{pmatrix}$. Construct an orthonormal basis for a subspace $span\left\{\begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \begin{pmatrix}0\\1\\1\\1\end{pmatrix}, \begin{pmatrix}0\\0\\1\\1\end{pmatrix}\right\}$ of R^4.

R^4: $\langle \vec{v}, \vec{w} \rangle = \sum_{i=1}^{n} v_i w_i$

$\langle \vec{v}, \vec{w} \rangle = \sum_{i=1}^{n} h_i v_i w_i$

$\Rightarrow (h_1, h_2 \cdots h_n) = (\frac{1}{4}, \frac{2}{4}, \cdots\cdots)$

多一个 weight function 内积不同

是否会得到不同的 orthogonal set.

Solution

Let $\boldsymbol{v}_1 = \boldsymbol{x}_1$ and $W_1 = span\{\boldsymbol{v}_1\} = span\{\boldsymbol{x}_1\}$.

Let $\boldsymbol{v}_2 = \boldsymbol{x}_2 - \text{proj}_{W_1}\boldsymbol{x}_2 = \boldsymbol{x}_2 - \dfrac{\langle \boldsymbol{x}_2, \boldsymbol{v}_1 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle}\boldsymbol{v}_1 = \begin{pmatrix}0\\1\\1\\1\end{pmatrix} - \dfrac{3}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} = \begin{pmatrix}-3/4\\1/4\\1/4\\1/4\end{pmatrix}$.

$\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is an orthogonal basis for the subspace $W_2 = \{\boldsymbol{x}_1, \boldsymbol{x}_2\}$.

$$\text{proj}_{W_2}\boldsymbol{x}_3 = \underbrace{\frac{\langle \boldsymbol{x}_3, \boldsymbol{v}_1 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle}\boldsymbol{v}_1}_{\substack{\text{projection of}\\ \boldsymbol{x}_3 \text{ onto } \boldsymbol{v}_1}} + \underbrace{\frac{\langle \boldsymbol{x}_3, \boldsymbol{v}_2 \rangle}{\langle \boldsymbol{v}_2, \boldsymbol{v}_2 \rangle}\boldsymbol{v}_2}_{\substack{\text{projection of}\\ \boldsymbol{x}_3 \text{ onto } \boldsymbol{v}_2}} = \frac{2}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} + \frac{\frac{2}{4}}{\frac{12}{16}}\begin{pmatrix}-3/4\\1/4\\1/4\\1/4\end{pmatrix} = \begin{pmatrix}0\\2/3\\2/3\\2/3\end{pmatrix}.$$

Let $\boldsymbol{v}_3 = \boldsymbol{x}_3 - \text{proj}_{W_2}\boldsymbol{x}_3 = \begin{pmatrix}0\\0\\1\\1\end{pmatrix} - \begin{pmatrix}0\\2/3\\2/3\\2/3\end{pmatrix} = \begin{pmatrix}0\\-2/3\\1/3\\1/3\end{pmatrix}$.

$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is an orthogonal basis for W.

Normalizing $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$, $\left\{\dfrac{1}{2}\begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \dfrac{1}{\sqrt{12}}\begin{pmatrix}-3\\1\\1\\1\end{pmatrix}, \dfrac{1}{\sqrt{6}}\begin{pmatrix}0\\-2\\1\\1\end{pmatrix}\right\}$ is an orthonormal basis for W.

Example 5.5.3

Let $\beta = \{1, x, x^2\}$ be a basis for P_3. Find an orthonormal basis if the inner product on P_3 is defined by

$$\langle p, q\rangle = \sum_{i=1}^{3} p(t_i)q(t_i)$$

where $t_1 = -1, t_2 = 0,$ and $t_3 = 1$.

Solution

Let $\boldsymbol{v}_1 = 1$ and $W_1 = span\{\boldsymbol{v}_1\} = span\{\boldsymbol{x}_1\}$.

Let $\boldsymbol{v}_2 = \boldsymbol{x}_2 - \mathrm{proj}_{W_1}\boldsymbol{x}_2 = \boldsymbol{x}_2 - \dfrac{\langle \boldsymbol{x}_2, \boldsymbol{v}_1\rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1\rangle}\boldsymbol{v}_1 = x - \dfrac{(-1)+0+1}{1+1+1} = x$

$\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is an orthogonal basis for the subspace $W_2 = \{\boldsymbol{x}_1, \boldsymbol{x}_2\}$.

$$\mathrm{proj}_{W_2}\boldsymbol{x}_3 = \underbrace{\frac{\langle \boldsymbol{x}_3, \boldsymbol{v}_1\rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1\rangle}\boldsymbol{v}_1}_{\substack{\text{projection of} \\ \boldsymbol{x}_3 \text{ onto } \boldsymbol{v}_1}} + \underbrace{\frac{\langle \boldsymbol{x}_3, \boldsymbol{v}_2\rangle}{\langle \boldsymbol{v}_2, \boldsymbol{v}_2\rangle}\boldsymbol{v}_2}_{\substack{\text{projection of} \\ \boldsymbol{x}_3 \text{ onto } \boldsymbol{v}_2}} = \frac{(-1)^2+0^2+1^2}{1+1+1} + \frac{(-1)^3+0^3+1^3}{(-1)^2+0^2+1^2}x = \frac{2}{3}$$

Let $\boldsymbol{v}_3 = \boldsymbol{x}_3 - \mathrm{proj}_{W_2}\boldsymbol{x}_3 = x^2 - \dfrac{2}{3}$.

$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is an orthogonal basis for W.

$$\boldsymbol{u}_1 = \frac{1}{\|\boldsymbol{v}_1\|}\boldsymbol{v}_1 = \frac{1}{\sqrt{\langle \boldsymbol{v}_1, \boldsymbol{v}_1\rangle}}\boldsymbol{v}_1 = \frac{1}{\sqrt{1+1+1}} = \frac{1}{\sqrt{3}}.$$

$$\boldsymbol{u}_2 = \frac{1}{\|\boldsymbol{v}_2\|}\boldsymbol{v}_2 = \frac{1}{\sqrt{\langle \boldsymbol{v}_2, \boldsymbol{v}_2\rangle}}\boldsymbol{v}_2 = \frac{1}{\sqrt{(-1)^2+0+1^2}}x = \frac{1}{\sqrt{2}}x$$

$$\langle \boldsymbol{v}_3, \boldsymbol{v}_3\rangle = \left((-1)^2 - \frac{2}{3}\right)^2 + \left(0^2 - \frac{2}{3}\right)^2 + \left(1^2 - \frac{2}{3}\right)^2 = \frac{2}{3}$$

$$\boldsymbol{u}_3 = \frac{1}{\|\boldsymbol{v}_3\|}\boldsymbol{v}_3 = \frac{1}{\sqrt{\langle \boldsymbol{v}_3, \boldsymbol{v}_3\rangle}}\boldsymbol{v}_3 = \frac{1}{\sqrt{\frac{2}{3}}}\left(x^2 - \frac{2}{3}\right) = \sqrt{\frac{3}{2}}\left(x^2 - \frac{2}{3}\right)$$

$\left\{\dfrac{1}{\sqrt{3}}, \dfrac{1}{\sqrt{2}}x, \sqrt{\dfrac{3}{2}}\left(x^2 - \dfrac{2}{3}\right)\right\}$ is an orthonormal basis for W.

Different inner product > different set

Example 5.5.4

Let $\beta = \{1, x, x^2\}$ be a basis for P_3. Find an orthonormal basis if the inner product on P_3 is defined by

$$\langle f(x), g(x)\rangle = \int_{-1}^{1} f(x)g(x)dx.$$

Solution

Let $\boldsymbol{v}_1 = 1$ and $W_1 = span\{\boldsymbol{v}_1\} = span\{\boldsymbol{x}_1\}$.

Let $\boldsymbol{v}_2 = \boldsymbol{x}_2 - \text{proj}_{W_1}\boldsymbol{x}_2 = \boldsymbol{x}_2 - \dfrac{\langle \boldsymbol{x}_2, \boldsymbol{v}_1\rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1\rangle}\boldsymbol{v}_1 = x - \dfrac{\int_{-1}^{1} xdx}{\int_{-1}^{1} 1dx}\boldsymbol{v}_1 = x - \dfrac{0}{2} = x$

$\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is an orthogonal basis for the subspace $W_2 = \{\boldsymbol{x}_1, \boldsymbol{x}_2\}$.

$$\text{proj}_{W_2}\boldsymbol{x}_3 = \underbrace{\frac{\langle \boldsymbol{x}_3, \boldsymbol{v}_1\rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1\rangle}\boldsymbol{v}_1}_{\substack{\text{projection of}\\ \boldsymbol{x}_3 \text{ onto } \boldsymbol{v}_1}} + \underbrace{\frac{\langle \boldsymbol{x}_3, \boldsymbol{v}_2\rangle}{\langle \boldsymbol{v}_2, \boldsymbol{v}_2\rangle}\boldsymbol{v}_2}_{\substack{\text{projection of}\\ \boldsymbol{x}_3 \text{ onto } \boldsymbol{v}_2}} = \frac{\int_{-1}^{1} x^2dx}{\int_{-1}^{1} 1dx} + \frac{\int_{-1}^{1} x^3dx}{\int_{-1}^{1} x^2dx}x = \frac{1}{3} - \frac{0}{\frac{2}{3}}x = \frac{1}{3}$$

Let $\boldsymbol{v}_3 = \boldsymbol{x}_3 - \text{proj}_{W_2}\boldsymbol{x}_3 = x^2 - \dfrac{1}{3}$.

$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is an orthogonal basis for W.

$$\boldsymbol{u}_1 = \frac{1}{\|\boldsymbol{v}_1\|}\boldsymbol{v}_1 = \frac{1}{\sqrt{\langle \boldsymbol{v}_1, \boldsymbol{v}_1\rangle}}\boldsymbol{v}_1 = \frac{1}{\sqrt{\int_{-1}^{1} 1dx}} = \frac{1}{\sqrt{2}}.$$

$$\boldsymbol{u}_2 = \frac{1}{\|\boldsymbol{v}_2\|}\boldsymbol{v}_2 = \frac{1}{\sqrt{\langle \boldsymbol{v}_2, \boldsymbol{v}_2\rangle}}\boldsymbol{v}_2 = \frac{1}{\sqrt{\int_{-1}^{1} x^2dx}}x = \sqrt{\frac{3}{2}}x$$

$$\int_{-1}^{1}\left(x^2 - \frac{1}{3}\right)^2 dx = \int_{-1}^{1} x^4 - \frac{2x^2}{3} + \frac{1}{9}dx = \left[\frac{x^5}{5} - \frac{2x^3}{9} + \frac{x}{9}\right]_{-1}^{1} = \frac{8}{45}$$

$$\boldsymbol{u}_3 = \frac{1}{\|\boldsymbol{v}_3\|}\boldsymbol{v}_3 = \frac{1}{\sqrt{\langle \boldsymbol{v}_3, \boldsymbol{v}_3\rangle}}\boldsymbol{v}_3 = \frac{1}{\sqrt{\int_{-1}^{1}\left(x^2 - \frac{1}{3}\right)^2 dx}}\left(x^2 - \frac{1}{3}\right) = \sqrt{\frac{5}{8}}\left(3x^2 - 1\right)$$

$\left\{\dfrac{1}{\sqrt{2}}, \sqrt{\dfrac{3}{2}}x, \sqrt{\dfrac{5}{8}}\left(3x^2 - 1\right)\right\}$ is an orthonormal basis for W.

normalize.

$$A = (x_1 \vdots x_2 \vdots \cdots x_n) = (v_1 \vdots v_2 \vdots \cdots v_n)\begin{pmatrix} 1 & \frac{\langle x_2, v_1\rangle}{\langle v_1, v_1\rangle} & \cdots & \frac{\langle x_n, v_1\rangle}{\langle v_1, v_1\rangle} \\ 0 & 1 & \ddots & \vdots \\ 0 & 0 & & \frac{\langle x_n, v_{n-1}\rangle}{\langle v_{n-1}, v_{n-1}\rangle} \\ \vdots & \vdots & & \\ 0 & 0 & & 1 \end{pmatrix}$$

$$\left(\frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}\,\middle|\,\frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|}\,\middle|\,\cdots\,\middle|\,\frac{\boldsymbol{v}_n}{\|\boldsymbol{v}_n\|}\right)\begin{pmatrix} \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}\cdot\boldsymbol{v}_1 & \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}\cdot\boldsymbol{x}_2 & \cdots & \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}\cdot\boldsymbol{x}_n \\ 0 & \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|}\cdot\boldsymbol{v}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{\boldsymbol{v}_{n-1}}{\|\boldsymbol{v}_{n-1}\|}\cdot\boldsymbol{x}_n \\ 0 & \cdots & 0 & \frac{\boldsymbol{v}_n}{\|\boldsymbol{v}_n\|}\cdot\boldsymbol{v}_n \end{pmatrix}$$

$Q_{m\times n} = \left(\frac{\vec{v}_1}{\|\vec{v}_1\|} \vdots \cdots \frac{\vec{v}_n}{\|\vec{v}_n\|}\right)$

$\boxed{\text{orthogonal set}} \Rightarrow m \geq n.$

$Q^T_{n\times m}Q = \begin{pmatrix} \frac{\vec{v}_1}{\|\vec{v}_1\|}^T \\ \frac{\vec{v}_2}{\|\vec{v}_2\|}^T \\ \vdots \\ \frac{\vec{v}_n}{\|\vec{v}_n\|}^T \end{pmatrix}\left(\frac{\vec{v}_1}{\|\vec{v}_1\|} \vdots \frac{\vec{v}_2}{\|\vec{v}_2\|} \vdots \cdots \frac{\vec{v}_n}{\|\vec{v}_n\|}\right)$

$(Q^TQ)_{ij} = \frac{v_i^T}{\|v_i\|}\cdot\frac{v_j}{\|v_j\|} = \begin{cases} 0 & i\neq j \\ 1 & i=j \end{cases}$

$Q^TQ = I_{n\times n}.$

$A = (\vec{x}_1 \vdots \vec{x}_2 \vdots \cdots \vdots \vec{x}_n) = (\vec{v}_1 \vdots \vec{v}_2 \vdots \cdots \vec{v}_n) = \begin{pmatrix} 1 & \frac{\langle \vec{x}_2, \vec{v}_1\rangle}{\langle \vec{v}_1, \vec{v}_1\rangle} & \frac{\langle \vec{x}_3, \vec{v}_1\rangle}{\langle \vec{v}_1, \vec{v}_1\rangle} & \cdots & \frac{\langle \vec{x}_n, \vec{v}_1\rangle}{\langle \vec{v}_1, \vec{v}_1\rangle} \\ 0 & 1 & \frac{\langle \vec{x}_3, \vec{v}_2\rangle}{\langle \vec{v}_2, \vec{v}_2\rangle} & & \frac{\langle \vec{x}_n, \vec{v}_2\rangle}{\langle \vec{v}_2, \vec{v}_2\rangle} \\ 0 & 0 & 1 & & \vdots \\ \vdots & \vdots & \vdots & & \frac{\langle \vec{x}_n, \vec{v}_{n-1}\rangle}{\langle \vec{v}_{n-1}, \vec{v}_{n-1}\rangle} \\ 0 & 0 & 0 & & 1 \end{pmatrix}$

$(\vec{v}_1 \vdots \vec{v}_2 \vdots \cdots \vec{v}_n)\begin{pmatrix} \frac{1}{\|\vec{v}_1\|} & & \\ & \frac{1}{\|\vec{v}_2\|} & \\ & & \ddots & \\ & & & \frac{1}{\|\vec{v}_n\|} \end{pmatrix}\begin{bmatrix} \|v_1\| & & & \\ & \|v_2\| & & \\ & & \ddots & \\ & & & \|v_n\| \end{bmatrix}$

Theorem 5.5.5 (QR factorization)

$m \geq n$

Use the same notation as in Theorem 5.6.1 where $V = R^m$. If $A = (\boldsymbol{x}_1 \mid \boldsymbol{x}_2 \mid \cdots \mid \boldsymbol{x}_n)$ is an $m \times n$ matrix of rank n, then A can be factored into a product QR, where

熟记 QR 的结构表达

$$Q = \left(\frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|} \;\middle|\; \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|} \;\middle|\; \cdots \;\middle|\; \frac{\boldsymbol{v}_n}{\|\boldsymbol{v}_n\|} \right) \quad \text{and} \quad R = \begin{pmatrix} \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}\cdot\boldsymbol{v}_1 & \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}\cdot\boldsymbol{x}_2 & \cdots & \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}\cdot\boldsymbol{x}_n \\ 0 & \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|}\cdot\boldsymbol{v}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{\boldsymbol{v}_{n-1}}{\|\boldsymbol{v}_{n-1}\|}\cdot\boldsymbol{x}_n \\ 0 & \cdots & 0 & \frac{\boldsymbol{v}_n}{\|\boldsymbol{v}_n\|}\cdot\boldsymbol{v}_n \end{pmatrix}$$

$A_{m\times n} = Q_{m\times n} \cdot R_{n\times n}$

Example 5.5.6

Let $A = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array}$ and $Q = \begin{array}{c} \begin{matrix} \frac{v_1}{\|v_1\|} & \frac{v_2}{\|v_2\|} & \frac{v_3}{\|v_3\|} \end{matrix} \\ \begin{pmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{pmatrix} \end{array}$ as in Example 5.5.2. We have

$$\left\{ \underset{v_1}{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}, \underset{v_2}{\begin{bmatrix} -\frac{3}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}}, \underset{v_3}{\begin{bmatrix} 0 \\ -\frac{2}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}} \right\}$$

$$\frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}\cdot\boldsymbol{v}_1 = 2 \qquad \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}\cdot\boldsymbol{x}_2 = 3/2 \qquad \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}\cdot\boldsymbol{x}_3 = 1$$

$$\frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|}\cdot\boldsymbol{v}_2 = \sqrt{3}/2 \qquad \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|}\cdot\boldsymbol{x}_3 = 1/\sqrt{3}$$

$$\frac{\boldsymbol{v}_3}{\|\boldsymbol{v}_3\|}\cdot\boldsymbol{v}_3 = \sqrt{2}/\sqrt{3}$$

Then $\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} 2 & 3/2 & 1 \\ 0 & \sqrt{3}/2 & 1/\sqrt{3} \\ 0 & 0 & \sqrt{2}/\sqrt{3} \end{pmatrix}$.

QR factorization $\Rightarrow$ $Q^TQ = I_n$.

Theorem 5.5.7

If A is an $m \times n$ matrix of rank n, then the least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$ is given by $\hat{\boldsymbol{x}} = R^{-1}Q^T\boldsymbol{b}$, where Q and R are the matrices obtained from QR factorization.

$Ax = b \quad A^TA\hat{x} = A^Tb \Rightarrow \hat{x} = (A^TA)^{-1}\cdot A^Tb \qquad Q_{m\times n}$

$= (R^TQ^TQR)^{-1}QRb = R^{-1}\cdot R^{-T}R^TQ^Tb = R^{-1}Q^Tb \Rightarrow m \geq n.$

Special case $m = n \Rightarrow \hat{x} = R^{-1}Q^{-1}\cdot b$

special case: $m=n \Rightarrow$

Example 5.5.8

Using QR factorization, find the least square solution of

$$\begin{pmatrix}1&0&0\\1&1&0\\1&1&1\\1&1&1\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}1\\2\\3\\4\end{pmatrix}.$$

Solution

$$A=\begin{array}{c}\begin{array}{ccc}x_1&x_2&x_3\end{array}\\\begin{pmatrix}1&0&0\\1&1&0\\1&1&1\\1&1&1\end{pmatrix}\end{array},\ Q=\begin{pmatrix}1/2&-3/\sqrt{12}&0\\1/2&1/\sqrt{12}&-2/\sqrt{6}\\1/2&1/\sqrt{12}&1/\sqrt{6}\\1/2&1/\sqrt{12}&1/\sqrt{6}\end{pmatrix}\text{ and } R=\begin{pmatrix}2&3/2&1\\0&\sqrt{3}/2&1/\sqrt{3}\\0&0&\sqrt{2}/\sqrt{3}\end{pmatrix}.$$

Then $R^{-1}=\begin{pmatrix}1/2&-\sqrt{3}/2&0\\0&2/\sqrt{3}&-\sqrt{2}/\sqrt{3}\\0&0&\sqrt{3}/\sqrt{2}\end{pmatrix}$. 好好求.

$$\begin{aligned}\hat{\boldsymbol{x}}&=R^{-1}Q^T\boldsymbol{b}\\&=\begin{pmatrix}1/2&-\sqrt{3}/2&0\\0&2/\sqrt{3}&-\sqrt{2}/\sqrt{3}\\0&0&\sqrt{3}/\sqrt{2}\end{pmatrix}\begin{pmatrix}1/2&1/2&1/2&1/2\\-3/\sqrt{12}&1/\sqrt{12}&1/\sqrt{12}&1/\sqrt{12}\\0&-2/\sqrt{6}&1/\sqrt{6}&1/\sqrt{6}\end{pmatrix}\begin{pmatrix}1\\2\\3\\4\end{pmatrix}\\&=\begin{pmatrix}1&0&0&0\\-1&1&0&0\\0&-1&\frac{1}{2}&\frac{1}{2}\end{pmatrix}\begin{pmatrix}1\\2\\3\\4\end{pmatrix}\\&=\begin{pmatrix}1\\1\\\frac{3}{2}\end{pmatrix}\end{aligned}$$

不见得有多简单

The answer is compatible with

$$\hat{\boldsymbol{x}}=\left(A^TA\right)^{-1}A^T\boldsymbol{b}=\begin{pmatrix}4&3&2\\3&3&2\\2&2&2\end{pmatrix}^{-1}\begin{pmatrix}1&1&1&1\\0&1&1&1\\0&0&1&1\end{pmatrix}\begin{pmatrix}1\\2\\3\\4\end{pmatrix}=\begin{pmatrix}1&-1&0\\-1&2&-1\\0&-1&\frac{3}{2}\end{pmatrix}\begin{pmatrix}10\\9\\7\end{pmatrix}=\begin{pmatrix}1\\1\\\frac{3}{2}\end{pmatrix}$$

$$A=QR=\begin{cases}\text{Incompatible} & A\vec{x}=\vec{b},\ \hat{x}=R^{-1}Q^Tb\\[2ex]\text{compatible} & A\vec{x}=\vec{b}\Rightarrow QR\vec{x}=\vec{b}\ \begin{cases}R\vec{x}=y\\Qy=\vec{b}\end{cases}\end{cases}$$

Section 6.1 Eigenvectors and Eigenvalues

$A\vec{v} = \lambda\vec{v}$

$\Rightarrow (A-\lambda I_n)\vec{v} = 0 \quad \vec{v} \neq \vec{0}$

$\det(A-\lambda I_n) = 0$

$A-\lambda I_n$ is singular.

Definition 6.1.1

Let A be an $n \times n$ matrix. A scalar λ is said to be an **eigenvalue** of A if there exists a nonzero vector $\boldsymbol{v} \in R^n$ such that $A\boldsymbol{v} = \lambda\boldsymbol{v}$. The vector $\boldsymbol{v}$ is said to be an **eigenvector** corresponding to λ.

Theorem 6.1.2

Let A be an $n \times n$ matrix and λ be a scalar. The following statements are equivalent:

(a) λ is an eigenvalue of A.

(b) $(A - \lambda I_n)\boldsymbol{v} = \boldsymbol{0}$ has a nontrivial solution.

(c) $\det(A - \lambda I_n) = 0$.

Definition 6.1.3

Let A be an $n \times n$ matrix. The polynomial $p(\lambda) = \det(\lambda I_n - A)$ is called the **characteristic polynomial** of A.

Remark 6.1.4

λ is an eigenvalue of A if and only if λ is a root of characteristic polynomial of A.

Any nonzero vector $\boldsymbol{v} \in N(A - \lambda I_n)$ is an eigenvector of A corresponding to eigenvalue λ. In particular, a nonzero vector $\boldsymbol{v} \in N(A)$ is an eigenvector of A corresponding to 0.

Remark 6.1.5

Let A be an $n \times n$ matrix, λ be an eigenvalue of A corresponding to eigenvector $\boldsymbol{v}$. Then λ^k is an eigenvalue of A^k corresponding to eigenvector $\boldsymbol{v}$ for positive integer k. Suppose k is a negative integer and A is invertible. Then λ^k is an eigenvalue of A^k corresponding to eigenvector $\boldsymbol{v}$. Let a be a constant. Then $a\lambda$ is an eigenvalue of aA corresponding to eigenvector $\boldsymbol{v}$.

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \quad \det(A-\lambda I_n) = \begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ \vdots & \ddots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

$A\vec{P_k} = \lambda_k \vec{P_k}$ $\quad \vec{A}(\vec{P_1}, \vec{P_2} \cdots \vec{P_n}) = (\vec{P_1}, \vec{P_2} \cdots \vec{P_n}) \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$

$\Rightarrow (A\vec{P_1}, A\vec{P_2} \cdots A\vec{P_n}) = (\lambda_1\vec{P_1}, \lambda_2\vec{P_2} \cdots \lambda_n\vec{P_n})$

$\begin{bmatrix} -\lambda & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = P^{-1}AP = \det A.$

$\vec{A} = \begin{bmatrix} v_1 \vdots \cdots \end{bmatrix} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \lambda_{\ddots} & \\ & & \cdot \end{bmatrix}} \begin{bmatrix} v_1 \vdots v_n \end{bmatrix}^{-1}$

Theorem 6.1.6

Suppose $p(\lambda) = \det(\lambda I_n - A) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0$. Then $a_0 = (-1)^n \det(A)$ and $a_{n-1} = -Tr(A)$.

Definition 6.1.7

An $n \times n$ matrix A is said to be **diagonalizable** if there exists a nonsingular matrix P and a diagonal matrix D such that

$$A = PDP^{-1}. \quad = \text{Diagonalize.}$$

The k-th diagonal entry of D is an eigenvalue of A corresponding to the eigenvector, the k-th column vector of P.

basis dimension

independ

$V = \text{span}\{\vec{v}_1 \cdots \vec{v}_k\}$.

$\text{span}\{\vec{v}_1 \cdots \vec{v}_r\}$ $r<k$.

Theorem 6.1.8

An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

$A =$ $\alpha_1 \cdots \alpha_k$ distinct

$\vec{v}_1 \cdots \vec{v}_k$.

Theorem 6.1.9

If $\lambda_1, \cdots, \lambda_k$ are distinct eigenvalues of an $n \times n$ matrix A corresponding to eigenvectors $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k$ then $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ is linearly independent. If A has n distinct eigenvalues then A is diagonalizable.

Example 6.1.10

Find all eigenvectors and eigenvalues of $A = \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix}$. Hence diagonalize A.

Solution

$A_{2\times2}$ $\det(A - \lambda I_2) = \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = \lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21})$.

The characteristic polynomial is

$$\det(\lambda I_2 - A) = \det\begin{pmatrix} \lambda - 1 & -3 \\ -4 & \lambda - 2 \end{pmatrix} = (\lambda - 2)(\lambda - 1) - 12 = \lambda^2 - 3\lambda - 10 = (\lambda - 5)(\lambda + 2)$$

The eigenvalues of A are $\lambda_1 = 5$ and $\lambda_2 = -2$.

λ_1 λ_2 are the root.

$\det(A - \lambda I_n) = 0$

$\lambda_1 + \lambda_2 = a_{11} + a_{22} = tr(A) = -\frac{b}{a} =$

$\lambda_1 \cdot \lambda_2 = a_{11}a_{22} - a_{12}a_{21} = \det(A)$

Consider $\lambda_1 = 5$.

$$N(A-5I_2) = N\begin{pmatrix} 1-5 & 3 \\ 4 & 2-5 \end{pmatrix} = N\begin{pmatrix} -4 & 3 \\ 4 & -3 \end{pmatrix} = span\left\{\begin{pmatrix} 3 \\ 4 \end{pmatrix}\right\}.$$

Eigenvector $\boldsymbol{v}_1 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$.

Consider $\lambda_2 = -2$.

$$N(A+2I_2) = N\begin{pmatrix} 1+2 & 3 \\ 4 & 2+2 \end{pmatrix} = N\begin{pmatrix} 3 & 3 \\ 4 & 4 \end{pmatrix} = span\left\{\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right\}.$$

Eigenvector $\boldsymbol{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

$$\begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 4 & -1 \end{pmatrix}\begin{pmatrix} 5 & 0 \\ 0 & -2 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 4 & -1 \end{pmatrix}^{-1}$$

Definition 6.1.11

Let λ be an eigenvalue of A with characteristic polynomial $f(x)$. The **algebraic multiplicity** of λ is the largest positive integer k for which $(x-\lambda)^k$ is a factor of $f(x)$, i.e., there are exactly k copies of eigenvalue λ.

Theorem 6.1.12 ① $A=PBP^{-1}$

If two $n \times n$ matrices A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues (with the same multiplicities).

Theorem 6.1.13

Let A be an $n \times n$ matrix and $\lambda_1, \cdots, \lambda_n$ be all eigenvalues of A. Then

$$Tr(A) = \lambda_1 + \cdots + \lambda_n \text{ and } \det(A) = \lambda_1 \cdots \lambda_n.$$

Example 6.1.14

Find all eigenvectors and eigenvalues of $A=\begin{pmatrix} a & a & \cdots & a \\ a & a & \ddots & \vdots \\ \vdots & \ddots & \ddots & a \\ a & \cdots & a & a \end{pmatrix}_{n\times n}$ and $B=\begin{pmatrix} b & a & \cdots & a \\ a & b & \ddots & \vdots \\ \vdots & \ddots & \ddots & a \\ a & \cdots & a & b \end{pmatrix}_{n\times n}$.

Solution

Since $rank(A)=1, \dim N(A)=n-1$ by rank-nullity Theorem. Clearly, $\begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix}$ span $N(A)$. $\begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix}$ are $n-1$ eigenvectors of A corresponding to 0.

Let λ be the last eigenvalue of A. By Theorem 6.1.14, we have $0+\cdots+0+\lambda=Tr(A)=na.$

Consider $N(A-\lambda I_n)=N\begin{pmatrix} -(n-1)a & a & \cdots & a \\ a & -(n-1)a & \ddots & \vdots \\ \vdots & \ddots & \ddots & a \\ a & \cdots & a & -(n-1)a \end{pmatrix}_{n\times n}$. Clearly, $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ spans $N(A-\lambda I_n)$.

$$\begin{array}{lccccc} \text{eigenvectors}: & \begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, & \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, & \cdots, & \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix}, & \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix} \\ \text{eigenvalues}: & 0, & 0 & ,\cdots, & 0 & ,na \end{array}$$

$$B=\begin{pmatrix} b & a & \cdots & a \\ a & b & \ddots & \vdots \\ \vdots & \ddots & \ddots & a \\ a & \cdots & a & b \end{pmatrix}_{n\times n}=\begin{pmatrix} a & a & \cdots & a \\ a & a & \ddots & \vdots \\ \vdots & \ddots & \ddots & a \\ a & \cdots & a & a \end{pmatrix}_{n\times n}+\begin{pmatrix} b-a & 0 & \cdots & 0 \\ 0 & b-a & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & b-a \end{pmatrix}_{n\times n}=A+(b-a)I_n.$$

Let λ be an eigenvalue of A corresponding to eigenvector $\boldsymbol{v}$, i.e., $A\boldsymbol{v} = \lambda\boldsymbol{v}$. We have

$$B\boldsymbol{v} = \left(A + (b-a)I_n\right)\boldsymbol{v} = A\boldsymbol{v} + (b-a)\boldsymbol{v} = \lambda\boldsymbol{v} + (b-a)\boldsymbol{v} = (\lambda + b - a)\boldsymbol{v}.$$

$$\text{eigenvectors}: \begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}$$

$$\text{eigenvalues}: b-a,\ b-a, \cdots, b-a\ , b+(n-1)a$$

Definition 6.1.15

Let $f(x) = a_n x^n + \cdots + a_1 x + a_0$ be a polynomial with real coefficient. If A is an $n \times n$ matrix, we define

$A^0 = I_n$

$$f(A) = a_n A^n + \cdots + a_1 A + a_0 I_n.$$

$f(\lambda) = a_n\lambda^n + a_{n-1}\lambda^{n-1} \cdots a_1\lambda + a_0$

$B = f(A) = a_n A^n + a_{n-1}A^{n-1} + \cdots a_1 A + a_0 I_n$

replace $x \Rightarrow A$

Theorem 6.1.16 (Cayley-Hamilton)

Let A be an $n \times n$ matrix and let $f(\lambda) = \det(\lambda I_n - A)$ be the characteristic polynomial of A. Then $f(A) = 0_n$, the $n \times n$ zero matrix.

Corollary 6.1.17

For every positive integer p, we have

$$A^p \in span\{I_n, A, \cdots, A^{n-1}\}$$

i.e., every power of A can be expressed as linear combination of $I_n, A, \cdots, A^{n-1}$.

Remark 6.1.18

$f(A) = \det(AI_n - A) = \det(0) = 0$. Is it correct? No.

λ is a scalar but A is a matrix. We cannot substitute A into λ.

algebraic multiplicity
geometric multiplicity > diagonalization.

zero function of $A \Rightarrow A^{-1}$

Example 6.1.19

Let $A=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Its characteristic polynomial is

$$\det(\lambda I_2-A)=\det\begin{pmatrix} \lambda-a & -b \\ -c & \lambda-d \end{pmatrix}=(\lambda-a)(\lambda-d)-bc=\lambda^2-(a+d)\lambda+ad-bc.$$

$$\begin{aligned}
A^2-(a+d)A+(ad-bc)I_2 &= A\big(A-(a+d)I_2\big)+(ad-bc)I_2 \\
&= \begin{pmatrix} a & b \\ c & d \end{pmatrix}\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}-\begin{pmatrix} a+d & 0 \\ 0 & a+d \end{pmatrix}\right)+\begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} \\
&= \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} -d & b \\ c & -a \end{pmatrix}+\begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} \\
&= \begin{pmatrix} -ad+bc & 0 \\ 0 & -ad+bc \end{pmatrix}+\begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}
\end{aligned}$$

Suppose $\det(A)=ad-bc\neq 0$. We have $A\big((a+d)I_2-A)\big)=\big((a+d)I_2-A)\big)A=(ad-bc)I_2$. Then

$$A^{-1}=\frac{1}{ad-bc}\big((a+d)I_2-A)\big)=\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

零化多项式有啥用

Example 6.1.20

$f(A)=A^2-A-20I_3$ ✗.

Let $A=\begin{pmatrix} 3 & 1 & -3 \\ 2 & 4 & 3 \\ -4 & 2 & -1 \end{pmatrix}$. Show that $A^2-A=20I_3$. Hence find A^{-1}.

…… 可提供了.

Solution

Clearly, $A^2-A=A(A-I_3)=\begin{pmatrix} 3 & 1 & -3 \\ 2 & 4 & 3 \\ -4 & 2 & -1 \end{pmatrix}\begin{pmatrix} 2 & 1 & -3 \\ 2 & 3 & 3 \\ -4 & 2 & -2 \end{pmatrix}=20I_3$.

We have $A\left(\frac{1}{20}(A-I_3)\right)=\left(\frac{1}{20}(A-I_3)\right)A=I_3$. Then $A^{-1}=\frac{1}{20}(A-I_3)=\frac{1}{20}\begin{pmatrix} 2 & 1 & -3 \\ 2 & 3 & 3 \\ -4 & 2 & -2 \end{pmatrix}$.

$\phi(\lambda)=\det(\lambda I_n-A)$

$=a_n\lambda^n+a_{n-1}\lambda^{n-1}+\cdots a_1\lambda+a_0$

Prove : $a_nA^n+a_{n-1}A^{n-1}+\cdots a_0I_n=0_{n\times n}$.

Let $B=$ adjoint (λI_n-A)

$(\lambda I_n-A)\cdot B=\det(\lambda I_n-B)\cdot I_n$

$A\cdot A^{*}=(\det A)I_n$

$B=\lambda^{n-1}B_{n-1}+\lambda^{n-2}B_{n-2}+\cdots\lambda B_1+B_0$

$(\lambda I_n-A)B=\lambda B-AB$

$=\lambda^nB_{n-1}+\lambda^{n-1}B_{n-2}+\cdots\lambda^2B_1+\lambda B_0-$

$\lambda^{n-1}AB_{n-1}-\cdots-\lambda^2AB_1-\lambda AB_1-AB_0$

$A\vec{v}=\lambda\vec{v}$

$\Rightarrow(A-\lambda I_n)\vec{v}=0$

$E_{\lambda_i} = N(A - \lambda_i I_n)$

Definition 6.1.21

Let A be an $n \times n$ matrix and λ an eigenvalue of A. $E_\lambda = N(A - \lambda I_n)$ is called the **eigenspace** of A corresponding to eigenvalue λ.

$\dim(E_\lambda)$ is called the geometric multiplity of λ

Example 6.1.22

In Example 6.1.14, the algebraic multiplicities of $\lambda = b - a$ is $n-1$ and $\lambda = b + (n-1)a$ is 1.

$$E_{b-a} = span\left\{\begin{pmatrix}1\\-1\\0\\\vdots\\0\end{pmatrix}, \begin{pmatrix}1\\0\\-1\\0\\\vdots\\0\end{pmatrix}, \cdots, \begin{pmatrix}1\\0\\\vdots\\0\\-1\end{pmatrix}\right\} \quad \text{and} \quad E_{b+(n-1)a} = span\left\{\begin{pmatrix}1\\1\\\vdots\\1\\1\end{pmatrix}\right\}$$

Remark 6.1.23

Are all matrices diagonalizable? No.

Diagonalization

Algebraic multipility

$P(\lambda) = \det(\lambda I_n - A) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_k)^{m_k}$.

Geometric multiplity $\quad (A - \lambda_i I_n)\vec{v} = 0 \Rightarrow E_{\lambda_i} = N(A - \lambda_i I_n)$

$\dim(E_{\lambda_i})$

Example 6.1.24 (复数域的 eignvalue)

Consider rotation matrix $A = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}$. Geometrically, for any nonzero vector $\boldsymbol{v}$, $\boldsymbol{v}$ and $A\boldsymbol{v}$ are not collinear (if $\theta \neq n\pi$); hence $A\boldsymbol{v}$ is not a multiple of $\boldsymbol{v}$. Therefore A has no eigenvectors and, consequently, no eigenvalues.

A is not diagonalizable in $\boldsymbol{R}$ but diagonalizable in $\boldsymbol{C}$.

Characteristic polynomial of $A = \det\begin{pmatrix}\lambda - \cos\theta & \sin\theta\\ -\sin\theta & \lambda - \cos\theta\end{pmatrix} = (\lambda - \cos\theta)^2 + \sin^2\theta$.

There are 2 eigenvalues $\cos\theta + i\sin\theta \,(= e^{i\theta})$ and $\cos\theta - i\sin\theta \,(= e^{-i\theta})$ corresponding to eigenvectors $\begin{pmatrix}1\\-i\end{pmatrix}$ and $\begin{pmatrix}1\\i\end{pmatrix}$ respectively. Then

$$\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix} = \begin{pmatrix}1 & 1\\ -i & i\end{pmatrix}\begin{pmatrix}e^{i\theta} & 0\\ 0 & e^{-i\theta}\end{pmatrix}\begin{pmatrix}1 & 1\\ -i & i\end{pmatrix}^{-1}.$$

$P^{-1}AP = D$

A is diagonalization $\Leftrightarrow$ A has n linearly independent eigenvector.

$\Leftrightarrow$ Algebraic multiplity $\Leftrightarrow$ Geometric multiplity

For each λ_i

In the rest of this course, we consider vector spaces and matrices over C.

Example 6.1.25

Consider $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Its characteristic polynomial is $p(\lambda) = \lambda^2$. 0 is the only eigenvalue with algebraic multiplicity 2. If A is diagonalizable, then there exists an invertible matrix P such that

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = A = P\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}P^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Contradiction.

Theorem 6.1.26

Let $A \in M_{n\times n}(C)$ be an $n \times n$ matrix in C and $\lambda_1, \lambda_2, \cdots, \lambda_k$ the distinct eigenvalues of A. Then

(a) $\dim E_{\lambda_i} \le$ algebraic multiplicity of λ_i for all i.

(b) All "=" in (a) hold if and only if A is diagonalizable.

Theorem 6.1.27 $A = QDQ^{-1}$

If A is a real symmetric matrix, then there is an orthogonal matrix Q such that $Q^T AQ = D$, where D is diagonal.

In Theorem 6.1.26 (b), A is not diagonalizable because eigenspace E_λ does not provide enough eigenvector. We may extend the eigenspace E_λ to generalized eigenspace K_λ.

Definition 6.1.28

Let $A \in M_{n\times n}(C)$ and λ be an eigenvalue of A. A nonzero vector $\boldsymbol{v} \in C^n$ is called a **generalized eigenvector** of A corresponding to λ if $(A - \lambda I_n)^p(\boldsymbol{v}) = 0$ for some positive integer p. The **generalized eigenspace** of A corresponding to λ, denoted K_λ, is defined by

$$K_\lambda = \left\{\boldsymbol{v} \in C^n : (A - \lambda I_n)^p(\boldsymbol{v}) = 0 \text{ for some positive } p\right\}.$$

Theorem 6.1.29

Let $A \in M_{n\times n}(C)$ be an $n\times n$ matrix in C and $\lambda_1,\lambda_2,\cdots,\lambda_k$ be distinct eigenvalues of A with corresponding multiplicities $m_1,m_2,\cdots m_k$. For $1\le i\le k$, let β_i be an ordered basis for K_{λ_i}. Then

(a) $\beta_i \cap \beta_j = \varnothing$ for $i \ne j$.

(b) $\beta = \beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$ is an ordered basis for C^n.

(c) $K_{\lambda_i} = N\left((A-\lambda I_n)^{m_i}\right)$ and $\dim\left(K_{\lambda_i}\right) = m_i$ for all i.

Theorem 6.1.30

Let $A \in M_{n\times n}(C)$ and $\boldsymbol{v} \in C^n$ be a generalized eigenvector of A corresponding to the eigenvalue λ. Suppose that p is the smallest positive integer such that $(A-\lambda I_n)^p(\boldsymbol{v}) = \boldsymbol{0}$. Let $\gamma = \left\{(A-\lambda I_n)^{p-1}(\boldsymbol{v}),\cdots,(A-\lambda I_n)(\boldsymbol{v}),\boldsymbol{v}\right\}$. Then

$$[A]_\gamma = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \lambda & 1 \\ 0 & \cdots & 0 & 0 & \lambda \end{pmatrix}.$$

The above matrix is called is called Jordan block corresponding to eigenvalue λ

Hence
$$A = P\begin{pmatrix} \boxed{\begin{matrix} \lambda_1 & 1 & & \\ & \lambda_1 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_1 \end{matrix}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boxed{\begin{matrix} \lambda_2 & 1 & & \\ & \lambda_2 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_2 \end{matrix}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boxed{\begin{matrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{matrix}} \end{pmatrix} P^{-1}$$
where

$$P = \left((A-\lambda_1 I_n)^{p_1-1}(\boldsymbol{v}_1),\cdots,(A-\lambda_1 I_n)(\boldsymbol{v}_1),\boldsymbol{v}_1,\cdots,(A-\lambda_k I_n)^{p_k-1}(\boldsymbol{v}_k),\cdots,(A-\lambda_k I_n)(\boldsymbol{v}_k),\boldsymbol{v}_k\right).$$

Δ. Example 6.1.31

Consider $A=\begin{pmatrix}9&4&5\\-4&0&-3\\-6&-4&-2\end{pmatrix}$. Its characteristic polynomial is $\det(\lambda I_3-A)=\lambda^3-7\lambda^2+16\lambda-12=(\lambda-2)^2(\lambda-3)$.

There are 2 eigenvalues $\lambda=2$ (multiplicity $=2$), $\lambda=3$ (multiplicity $=1$).

Consider $E_2=N(A-2I_3)=N\begin{pmatrix}7&4&5\\-4&-2&-3\\-6&-4&-4\end{pmatrix}=\operatorname{span}\left\{\begin{pmatrix}-2\\1\\2\end{pmatrix}\right\}$.

Since $\dim E_2=1\neq 2$ (multiplicity of $\lambda=2$), A is not diagonalizable.

Extend E_2 to $K_2=N\left((A-2I_3)^2\right)=N\begin{pmatrix}3&0&3\\-2&0&-2\\-2&0&-2\end{pmatrix}$. Clearly, $\begin{pmatrix}1\\0\\-1\end{pmatrix}\in K_2$. Then

$$\beta_1=\left\{\underbrace{(A-2I_3)\begin{pmatrix}1\\0\\-1\end{pmatrix}=\overset{v_1}{\begin{pmatrix}2\\-1\\-2\end{pmatrix}}}_{E_2},\underbrace{\overset{v_2}{\begin{pmatrix}1\\0\\-1\end{pmatrix}}}_{K_2\setminus E_2}\right\}$$ will form a basis of K_2.

Consider $E_3=N(A-3I_3)=N\begin{pmatrix}6&4&5\\-4&-3&-3\\-6&-4&-5\end{pmatrix}=\operatorname{span}\left\{\begin{pmatrix}3\\-2\\-2\end{pmatrix}\right\}$.

Hence $\beta=\left\{\begin{pmatrix}2\\-1\\-2\end{pmatrix},\begin{pmatrix}1\\0\\-1\end{pmatrix},\begin{pmatrix}3\\-2\\-2\end{pmatrix}\right\}$ will form a basis of C^3. We have

注意顺序

$$\begin{pmatrix}9&4&5\\-4&0&-3\\-6&-4&-2\end{pmatrix}=\begin{pmatrix}2&1&3\\-1&0&-2\\-2&-1&-2\end{pmatrix}\left(\begin{array}{cc|c}2&\textcircled{1}&0\\0&2&0\\\hline 0&0&3\end{array}\right)\begin{pmatrix}2&1&3\\-1&0&-2\\-2&-1&-2\end{pmatrix}^{-1}$$

$v_1 \quad v_2 \quad K_1$

$$=\begin{pmatrix}2&1&3\\-1&0&-2\\-2&-1&-2\end{pmatrix}\left(\begin{array}{cc|c}2&1&0\\0&2&0\\\hline 0&0&3\end{array}\right)\begin{pmatrix}-2&-1&-2\\2&2&1\\1&0&1\end{pmatrix}.$$

$\dim E_{\lambda_i} < m_i$

$\dim K_{\lambda_i} = m_i \Rightarrow$ stop

Example 6.1.32

Consider $A=\begin{pmatrix}2&4&-8\\0&0&4\\0&-1&4\end{pmatrix}$. Its characteristic polynomial is $\det(\lambda I_3-A)=(\lambda-2)^3$.

two jordan chain.

⇗

There is 1 eigenvalue $\lambda=2$ (multiplicity $=3$).

Consider $E_2=N(A-2I_3)=N\begin{pmatrix}0&4&-8\\0&-2&4\\0&-1&2\end{pmatrix}=\operatorname{span}\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\2\\1\end{pmatrix}\right\}$.

Since $\dim E_2=2\neq 3$ (multiplicity of $\lambda=2$), A is not diagonalizable.

Extend E_2 to $K_2=N\left((A-2I_3)^2\right)=N\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}=C^3$. $\dim K_2=3$. Choose a vector, say

key-step

$\begin{pmatrix}0\\0\\1\end{pmatrix}\in N\left((A-2I_3)^2\right)\setminus N(A-2I_3)$.

Then $\beta=\left\{\underbrace{\begin{pmatrix}1\\0\\0\end{pmatrix},(A-2I_3)\begin{pmatrix}0\\0\\1\end{pmatrix}=\begin{pmatrix}-8\\4\\2\end{pmatrix}}_{E_2},\underbrace{\begin{pmatrix}0\\0\\1\end{pmatrix}}_{K_2\setminus E_2}\right\}$ and $\gamma=\left\{\underbrace{(A-2I_3)\begin{pmatrix}0\\0\\1\end{pmatrix}=\begin{pmatrix}-8\\4\\2\end{pmatrix}}_{E_2},\underbrace{\begin{pmatrix}0\\0\\1\end{pmatrix}}_{K_2\setminus E_2},\begin{pmatrix}1\\0\\0\end{pmatrix}\right\}$ will form a basis of C^3. We have

$$\begin{pmatrix}2&4&-8\\0&0&4\\0&-1&4\end{pmatrix}=\begin{pmatrix}1&-8&0\\0&4&0\\0&2&1\end{pmatrix}\left(\begin{array}{c|cc}2&0&0\\\hline 0&2&1\\0&0&2\end{array}\right)\begin{pmatrix}1&-8&0\\0&4&0\\0&2&1\end{pmatrix}^{-1}=\begin{pmatrix}1&-8&0\\0&4&0\\0&2&1\end{pmatrix}\left(\begin{array}{c|cc}2&0&0\\\hline 0&2&1\\0&0&2\end{array}\right)\begin{pmatrix}1&2&0\\0&\frac14&0\\0&-\frac12&1\end{pmatrix}.$$

or

⇓ 调换顺序

$$\begin{pmatrix}2&4&-8\\0&0&4\\0&-1&4\end{pmatrix}=\begin{pmatrix}-8&0&1\\4&0&0\\2&1&0\end{pmatrix}\left(\begin{array}{cc|c}2&1&0\\0&2&0\\\hline 0&0&2\end{array}\right)\begin{pmatrix}-8&0&1\\4&0&0\\2&1&0\end{pmatrix}^{-1}=\begin{pmatrix}-8&0&1\\4&0&0\\2&1&0\end{pmatrix}\left(\begin{array}{cc|c}2&1&0\\0&2&0\\\hline 0&0&2\end{array}\right)\begin{pmatrix}0&\frac14&0\\0&-\frac12&1\\1&2&0\end{pmatrix}.$$

Example 6.1.33

Consider $A=\begin{pmatrix}3&1&0\\-1&1&0\\3&2&2\end{pmatrix}$. Its characteristic polynomial is $\det(\lambda I_3-A)=(\lambda-2)^3$.

There is 1 eigenvalue $\lambda=2$ (multiplicity $=3$). $\Rightarrow$ one jordan chain.

Consider $E_2=N(A-2I_3)=N\begin{pmatrix}1&1&0\\-1&-1&0\\3&2&0\end{pmatrix}=\text{span}\left\{\begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$.

Since $\dim E_2=1\neq 3$ (multiplicity of $\lambda=2$), A is not diagonalizable.

Extend E_2 to $K_2=N\left((A-2I_3)^2\right)=N\begin{pmatrix}0&0&0\\0&0&0\\1&1&0\end{pmatrix}=\text{span}\left\{\begin{pmatrix}0\\0\\1\end{pmatrix},\begin{pmatrix}1\\-1\\0\end{pmatrix}\right\}$. $\dim K_2=2\neq 3$. The generalized eigenspace K_2 is not large enough to provide 3 eigenvectors. Let us consider $N\left((A-2I_3)^3\right)=N\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}=C^3$, which has dimension 3. Choose a vector, say $\begin{pmatrix}1\\0\\0\end{pmatrix}\in N\left((A-2I_3)^3\right)\setminus N\left((A-2I_3)^2\right)$.

Then $\beta=\left\{\underbrace{(A-2I_3)^2\begin{pmatrix}1\\0\\0\end{pmatrix}=\overset{\vec{v}_1}{\begin{pmatrix}0\\0\\1\end{pmatrix}}}_{E_2},\underbrace{(A-2I_3)\begin{pmatrix}1\\0\\0\end{pmatrix}=\overset{\vec{v}_2}{\begin{pmatrix}1\\-1\\3\end{pmatrix}}}_{K_2\setminus E_2},\underbrace{\overset{\vec{v}_3}{\begin{pmatrix}1\\0\\0\end{pmatrix}}}_{C^3\setminus K_2}\right\}$ will form a basis of C^3. We have

$$\begin{aligned}\begin{pmatrix}3&1&0\\-1&1&0\\3&2&2\end{pmatrix}&=\begin{pmatrix}0&1&1\\0&-1&0\\1&3&0\end{pmatrix}\begin{pmatrix}2&1&0\\0&2&1\\0&0&2\end{pmatrix}\begin{pmatrix}0&1&1\\0&-1&0\\1&3&0\end{pmatrix}^{-1}\\&=\begin{pmatrix}0&1&1\\0&-1&0\\1&3&0\end{pmatrix}\begin{pmatrix}2&1&0\\0&2&1\\0&0&2\end{pmatrix}\begin{pmatrix}0&3&1\\0&-1&0\\1&1&0\end{pmatrix}.\end{aligned}$$

real - symmetric matrix

$\Downarrow$

diagonalize

Section 6.2 Definite Matrices

quadratic form

$$\begin{cases} \frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \\ \frac{x^2}{a^2}-\frac{y^2}{b^2}=1 \\ y^2=cx \quad y=cx^2. \end{cases}$$

Definition 6.2.1

对称!!!

前提非常关键.

An $n\times n$ real symmetric matrix A is said to be

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

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

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

可以取0

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

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

不定.

$$f(x,y)=ax^2+bxy+cy^2 =(x,y)\begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c\end{pmatrix}\begin{pmatrix} x \\ y\end{pmatrix}$$

$$f(x,y)=\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \quad =(x,y)\begin{pmatrix} \frac{1}{a^2} & 0 \\ 0 & \frac{1}{b^2}\end{pmatrix}\begin{pmatrix} x \\ y\end{pmatrix}$$

$$f(x,y)=\frac{x^2}{a^2}-\frac{y^2}{b^2}=1 \quad =(x,y)\begin{pmatrix} \frac{1}{a^2} & 0 \\ 0 & -\frac{1}{b^2}\end{pmatrix}\begin{pmatrix} x \\ y\end{pmatrix}$$

Theorem 6.2.2

Let A be an $n\times n$ real symmetric matrix.

eignvalue 判定.

(i) A is positive definite if and only if all eigenvalues of A are positive.

(ii) A is negative definite if and only if all eigenvalues of A are negative.

(iii) A is indefinite if and only if some eigenvalues of A are positive and some eigenvalues of A are negative.

Suppose an everywhere infinitely differentiable function $z=f(x,y)$ has a critical point (x_0,y_0), i.e., $\frac{\partial f}{\partial x}(x_0,y_0)=\frac{\partial f}{\partial y}(x_0,y_0)=0$. (horizontal tangent plane at (x_0,y_0)). Taylor expansion of f at (x_0,y_0) is

condition

$$\frac{\partial f}{\partial x}(x_0,y_0)(x-x_0) \rightarrow 0$$

$$\frac{\partial f}{\partial y}(x_0,y_0)(y-y_0)$$

$$f(x_0+x,y_0+y)=\underbrace{f(x_0,y_0)+x\frac{\partial f}{\partial x}(x_0,y_0)+y\frac{\partial f}{\partial y}(x_0,y_0)}_{\text{equation of tangent plane at }(x_0,y_0)}$$

$$+\frac{1}{2}\left[x^2\frac{\partial^2 f}{\partial x^2}(x_0,y_0)+2xy\frac{\partial^2 f}{\partial x\partial y}(x_0,y_0)+y^2\frac{\partial^2 f}{\partial y^2}(x_0,y_0)\right]+\text{Error}$$

0

Δx 0

$$\approx f(x_0,y_0)+\frac{1}{2}\left[x^2\underbrace{\frac{\partial^2 f}{\partial x^2}}(x_0,y_0)+2xy\underbrace{\frac{\partial^2 f}{\partial x\partial y}}(x_0,y_0)+y^2\underbrace{\frac{\partial^2 f}{\partial y^2}}(x_0,y_0)\right]$$

Δy

两次项)

$$f(x_1,x_2,x_3)=ax_1^2+bx_1x_2+cx_2^2+dx_1x_3+ex_2x_3+fx_3^2$$

$$\Rightarrow (x_1,x_2,x_3)\begin{pmatrix} a & \frac{b}{2} & \frac{d}{2} \\ \frac{b}{2} & c & \frac{e}{2} \\ \frac{d}{2} & \frac{e}{2} & f\end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3\end{pmatrix}$$

$$\Downarrow$$

$$(x,y)\begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(x_0,y_0) & \frac{\partial^2 f}{\partial x\partial y}(x_0,y_0) \\ \frac{\partial^2 f}{\partial x\partial y}(x_0,y_0) & \frac{\partial^2 f}{\partial y^2}(x_0,y_0)\end{pmatrix}$$

A is real symmetric matrix

$\Updownarrow$

$$A=P\begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}P^{-1} \quad \underbrace{P^TP=I}$$

$$x^TAx = x^TP\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}P^{-1}x$$

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad = \vec{y}^T\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}\vec{y}$$

$$= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots \lambda_n y_n^2$$

$\lambda_1, \lambda_2 \cdots \lambda_n > 0$

$y^TDy > 0$ For $y \neq 0$

$\Downarrow$

$x^TAx > 0 \Rightarrow A$ is positive define

where $\dfrac{\text{Error}}{h^2+k^2} \to 0$ as $(h,k) \to (0,0)$.

Define the Hessian of f at (x_0, y_0) to be

$$\frac{(x-4)^2}{a^2} + \frac{y^2}{b^2} = 1$$

$\downarrow$ standard

$$\frac{\bar{x}^2}{a^2} + \frac{\bar{y}^2}{b^2} = 1 \Rightarrow \begin{cases} \bar{x} = x-4 \\ \bar{y} = y \end{cases}$$

$$H(x_0,y_0) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2}(x_0,y_0) & \dfrac{\partial^2 f}{\partial x \partial y}(x_0,y_0) \\ \dfrac{\partial^2 f}{\partial x \partial y}(x_0,y_0) & \dfrac{\partial^2 f}{\partial y^2}(x_0,y_0) \end{pmatrix}.$$

Then

$$x^2\frac{\partial^2 f}{\partial x^2}(x_0,y_0) + 2xy\frac{\partial^2 f}{\partial x \partial y}(x_0,y_0) + y^2\frac{\partial^2 f}{\partial y^2}(x_0,y_0) = (x,y)H(x_0,y_0)\begin{pmatrix} x \\ y \end{pmatrix}.$$

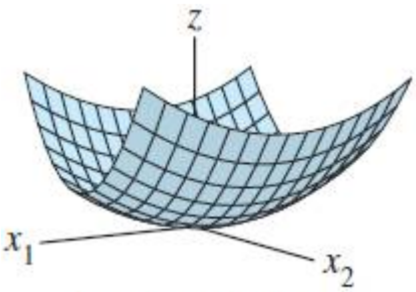
Theorem 6.2.3

Suppose an everywhere infinitely differentiable function $z = f(x,y)$ has a critical point (x_0, y_0). Let λ_1, λ_2 be eigenvalues of $H(x_0, y_0)$. Then

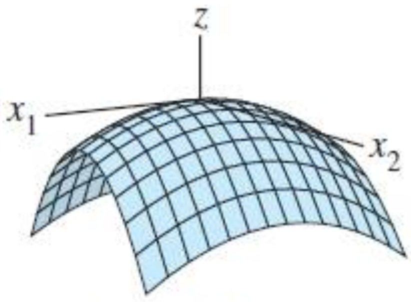
(i) f has a local minimum at (x_0,y_0) if $\lambda_1 > 0, \lambda_2 > 0$. ($H(x_0,y_0)$ is positive definite)

(ii) f has a local maximum at (x_0,y_0) if $\lambda_1 < 0, \lambda_2 < 0$. ($H(x_0,y_0)$ is negative definite)

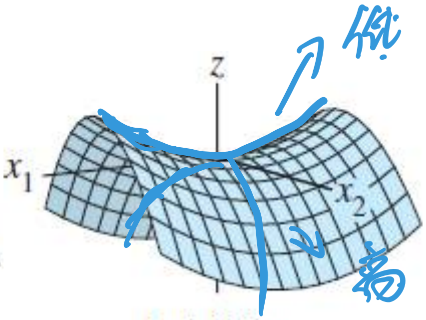
(iii) f has a saddle at (x_0,y_0) if $\lambda_1\lambda_2 < 0$. ($H(x_0,y_0)$ is indefinite)



Positive definite



Negative definite



Indefinite

Remark 6.2.4

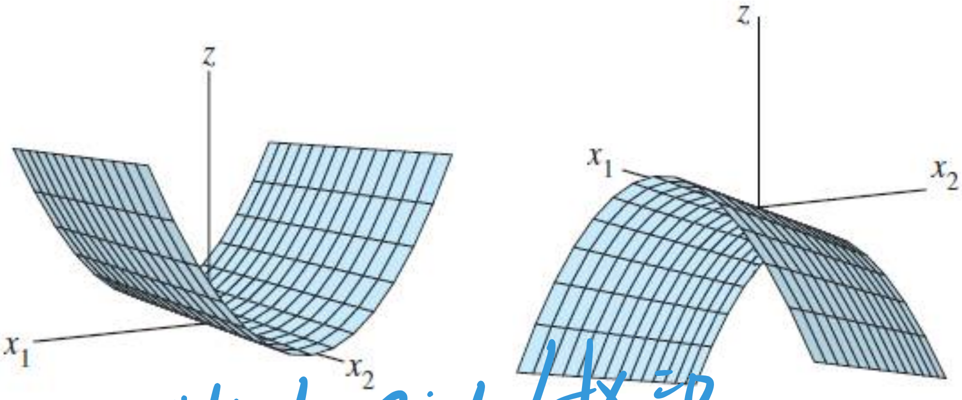
If one of eigenvalues of $H(x_0,y_0)$ equal 0, then $z = f(x,y)$ looks like:

negative/posive

quadratic form locate the local extrem value.

1. eignvalue

2. principal minor



Find the critical point $\begin{cases} f_x = 0 \\ f_y = 0 \end{cases}$

Theorem 6.2.5

Suppose we have the same condition in Theorem 6.2.3.

Principal minor ⇒ $(-1)^k D_k > 0$ negative define ↘

(i) If $\det H(x_0, y_0) > 0$ and $\frac{\partial^2 f}{\partial x^2}(x_0, y_0) < 0$, then f has a local maximum value at (x_0, y_0).

Positive define ↘

(ii) If $\det H(x_0, y_0) > 0$ and $\frac{\partial^2 f}{\partial x^2}(x_0, y_0) > 0$, then f has a local minimum value at (x_0, y_0).

$\lambda_1 \cdot \lambda_2 < 0$ 一正一负.

(iii) If $\det H(x_0, y_0) < 0$, then f has a saddle at (x_0, y_0).

(iv) If $\det H(x_0, y_0) = 0$, then no conclusion.

Example 6.2.6

Let $f(x, y) = 4x^3 + 2x^2 y + xy^2 - 4x$. Then

仅对 critical point 有作用

$$\frac{\partial f}{\partial x} = 12x^2 + 4xy + y^2 - 4, \quad \frac{\partial f}{\partial y} = 2x^2 + 2xy,$$

$$\frac{\partial^2 f}{\partial x^2} = 24x + 4y, \quad \frac{\partial^2 f}{\partial x \partial y} = 4x + 2y, \quad \frac{\partial^2 f}{\partial y^2} = 2x.$$

$$\frac{\partial f}{\partial y} = 2x^2 + 2xy = 0 \quad \Rightarrow \quad x = 0 \text{ or } y = -x.$$

$$x = 0, \frac{\partial f}{\partial x} = 12x^2 + 4xy + y^2 - 4 = 0 \quad \Rightarrow \quad y = \pm 2.$$

$$y = -x, \frac{\partial f}{\partial x} = 12x^2 + 4xy + y^2 - 4 = 0 \quad \Rightarrow \quad 9x^2 = 4 \quad \Rightarrow \quad x = \pm\frac{2}{3} \text{ and } y = \mp\frac{2}{3}.$$

Critical point:

(0,2) (0,-2)

$(\frac{2}{3}, -\frac{2}{3})$ $(-\frac{2}{3}, \frac{2}{3})$

Case I: critical point = $(0,2)$

$H(0,2)=\begin{pmatrix}8 & 4\\ 4 & 0\end{pmatrix}$ is indefinite because $\det\begin{pmatrix}8 & 4\\ 4 & 0\end{pmatrix}=-16<0$. f has a saddle point at $(0,2)$.

Case II: critical point = $(0,-2)$

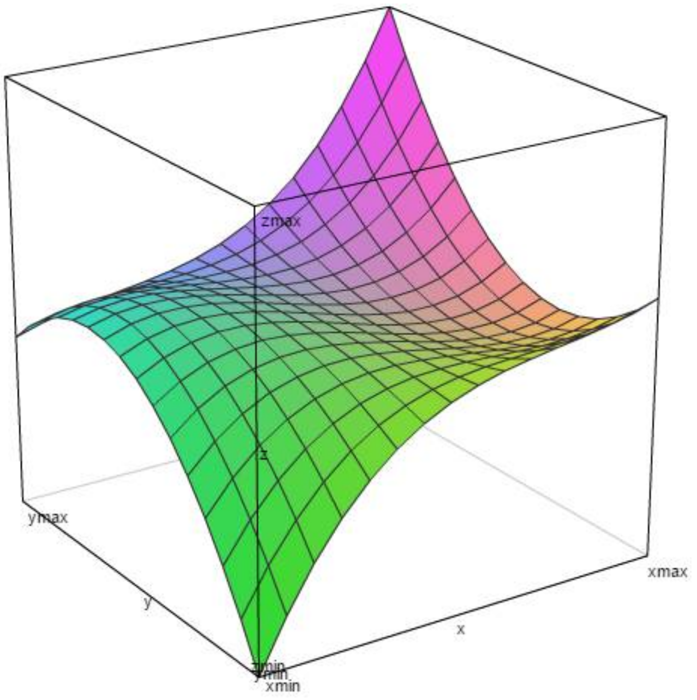
$H(0,-2)=\begin{pmatrix}-8 & -4\\ -4 & 0\end{pmatrix}$ is indefinite because $\det\begin{pmatrix}-8 & -4\\ -4 & 0\end{pmatrix}=-16<0$. f has a saddle point at $(0,-2)$.

Case III: critical point = $\left(\frac{2}{3},-\frac{2}{3}\right)$

$H\left(\frac{2}{3},-\frac{2}{3}\right)=\begin{pmatrix}\frac{40}{3} & \frac{4}{3}\\ \frac{4}{3} & \frac{4}{3}\end{pmatrix}$ is positive definite because $\frac{\partial^2 f}{\partial x^2}\left(\frac{2}{3},-\frac{2}{3}\right)=\frac{40}{3}>0$ and $\det\begin{pmatrix}\frac{40}{3} & \frac{4}{3}\\ \frac{4}{3} & \frac{4}{3}\end{pmatrix}>0$. f has a minimum point at $\left(\frac{2}{3},-\frac{2}{3}\right)$.

Case IV: critical point = $\left(-\frac{2}{3},\frac{2}{3}\right)$

$H\left(-\frac{2}{3},\frac{2}{3}\right)=\begin{pmatrix}-\frac{40}{3} & -\frac{4}{3}\\ -\frac{4}{3} & -\frac{4}{3}\end{pmatrix}$ is negative definite because $\frac{\partial^2 f}{\partial x^2}\left(-\frac{2}{3},\frac{2}{3}\right)=-\frac{40}{3}<0$ and $\det\begin{pmatrix}-\frac{40}{3} & -\frac{4}{3}\\ -\frac{4}{3} & -\frac{4}{3}\end{pmatrix}>0$. f has a maximum point at $\left(-\frac{2}{3},\frac{2}{3}\right)$.



Example 6.2.7

Let $f(x, y) = 2x^4 + y^4$. Then

$$\frac{\partial f}{\partial x} = 8x^3, \quad \frac{\partial f}{\partial y} = 4y^3, \quad \frac{\partial^2 f}{\partial x^2} = 24x^2, \quad \frac{\partial^2 f}{\partial x \partial y} = 0, \quad \frac{\partial^2 f}{\partial y^2} = 12y^2.$$

f has a critical point at $(0, 0)$.

$H(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. f has a minimum point at $(0, 0)$.

⇒ 用其它方法进行判定

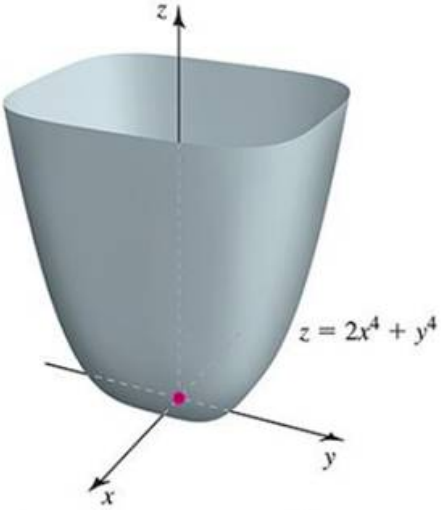
Example 6.2.8

Let $f(x, y) = 2 - xy^2$. Then

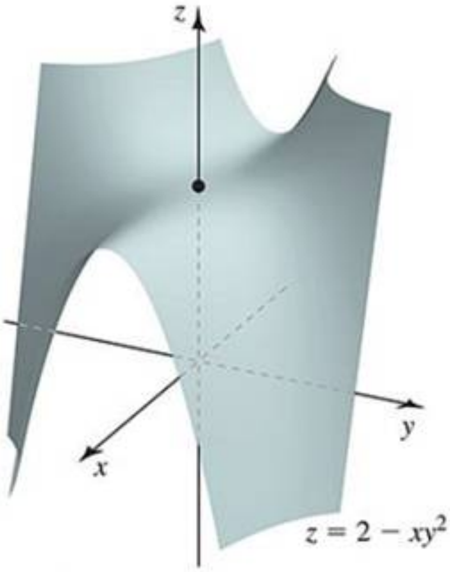
$$\frac{\partial f}{\partial x} = -y^2, \quad \frac{\partial f}{\partial y} = -2xy, \quad \frac{\partial^2 f}{\partial x^2} = 0, \quad \frac{\partial^2 f}{\partial x \partial y} = -2y, \quad \frac{\partial^2 f}{\partial y^2} = -2x.$$

f has a critical point at $(0, 0)$.

$H(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. f has a saddle point at $(0, 0)$.



Local minimum at (0, 0), but the Second Derivative Test is inconclusive.



Second derivative test fails to detect saddle point at (0, 0).

Step1: 找 A

两种 / 1. 矩阵法.

Step 2: 对角化·

方
法 $\Big\{$ 2. 代数法

Section 6.3 Recurrence relation

Theorem 6.3.1

Let A be an $n\times n$ matrix

$$\begin{pmatrix} -c_{n-1} & \cdots & -c_2 & -c_1 & -c_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \rightarrow \text{全为0}$$

where $c_0, c_1, \cdots, c_{n-1} \in R$. Then the characteristic polynomial of A is

$$\lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0.$$

If α is an eigenvalue of A, then its corresponding eigenvector is $\begin{pmatrix} \alpha^{n-1} \\ \vdots \\ \alpha \\ 1 \end{pmatrix}$.

Conclusion: ↗

application

Example 6.3.2

Consider Fibonacci sequence: $\{1, 1, 2, 3, 5, 8, 13, \cdots\}$. The recurrence relation is $a_{n+2} = a_{n+1} + a_n$, where $a_2 = a_1 = 1$. Find a_n.

Solution

Clearly, $\begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} a_{n-1} + a_{n-2} \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_{n-1} \\ a_{n-2} \end{pmatrix}$.

The characteristic polynomial of $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ is $\lambda^2 - \lambda - 1 = \left(\lambda - \frac{1+\sqrt{5}}{2}\right)\left(\lambda - \frac{1-\sqrt{5}}{2}\right)$.

$\lambda = \frac{1+\sqrt{5}}{2}, \frac{1-\sqrt{5}}{2}$ are eigenvalues corresponding to eigenvectors $\begin{pmatrix} \frac{1+\sqrt{5}}{2} \\ 1 \end{pmatrix}, \begin{pmatrix} \frac{1-\sqrt{5}}{2} \\ 1 \end{pmatrix}$ respectively.

example:

$a_n = a\cdot a_{n-1} + b\cdot a_{n-2}$

$\begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_{n-1} \\ a_{n-2} \end{pmatrix} = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix}^2\begin{pmatrix} a_{n-2} \\ a_{n-3} \end{pmatrix} = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix}^{n-1}\begin{pmatrix} a_1 \\ a_0 \end{pmatrix}$

Then

diagonalize

$$\begin{pmatrix}1&1\\1&0\end{pmatrix}=\begin{pmatrix}\frac{1+\sqrt{5}}{2}&\frac{1-\sqrt{5}}{2}\\1&1\end{pmatrix}\begin{pmatrix}\frac{1+\sqrt{5}}{2}&0\\0&\frac{1-\sqrt{5}}{2}\end{pmatrix}\begin{pmatrix}\frac{1+\sqrt{5}}{2}&\frac{1-\sqrt{5}}{2}\\1&1\end{pmatrix}^{-1}$$

Hence

to power n.

$$\begin{aligned}\begin{pmatrix}1&1\\1&0\end{pmatrix}^{n}&=\begin{pmatrix}\frac{1+\sqrt{5}}{2}&\frac{1-\sqrt{5}}{2}\\1&1\end{pmatrix}\begin{pmatrix}\frac{1+\sqrt{5}}{2}&0\\0&\frac{1-\sqrt{5}}{2}\end{pmatrix}^{n}\begin{pmatrix}\frac{1+\sqrt{5}}{2}&\frac{1-\sqrt{5}}{2}\\1&1\end{pmatrix}^{-1}\\
&=\frac{1}{\frac{1+\sqrt{5}}{2}-\frac{1-\sqrt{5}}{2}}\begin{pmatrix}\frac{1+\sqrt{5}}{2}&\frac{1-\sqrt{5}}{2}\\1&1\end{pmatrix}\begin{pmatrix}\left(\frac{1+\sqrt{5}}{2}\right)^{n}&0\\0&\left(\frac{1-\sqrt{5}}{2}\right)^{n}\end{pmatrix}\begin{pmatrix}1&-\frac{1-\sqrt{5}}{2}\\-1&\frac{1+\sqrt{5}}{2}\end{pmatrix}\\
&=\frac{1}{\sqrt{5}}\begin{pmatrix}\frac{1+\sqrt{5}}{2}&\frac{1-\sqrt{5}}{2}\\1&1\end{pmatrix}\begin{pmatrix}\left(\frac{1+\sqrt{5}}{2}\right)^{n}&0\\0&\left(\frac{1-\sqrt{5}}{2}\right)^{n}\end{pmatrix}\begin{pmatrix}1&-\frac{1-\sqrt{5}}{2}\\-1&\frac{1+\sqrt{5}}{2}\end{pmatrix}\end{aligned}$$

$$\begin{aligned}\begin{pmatrix}a_n\\a_{n-1}\end{pmatrix}&=\begin{pmatrix}1&1\\1&0\end{pmatrix}\begin{pmatrix}a_{n-1}\\a_{n-2}\end{pmatrix}\\
&=\begin{pmatrix}1&1\\1&0\end{pmatrix}^{n-2}\begin{pmatrix}a_2\\a_1\end{pmatrix}\\
&=\frac{1}{\sqrt{5}}\begin{pmatrix}\frac{1+\sqrt{5}}{2}&\frac{1-\sqrt{5}}{2}\\1&1\end{pmatrix}\begin{pmatrix}\left(\frac{1+\sqrt{5}}{2}\right)^{n-2}&0\\0&\left(\frac{1-\sqrt{5}}{2}\right)^{n-2}\end{pmatrix}\begin{pmatrix}1&-\frac{1-\sqrt{5}}{2}\\-1&\frac{1+\sqrt{5}}{2}\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix}\\
&=\frac{1}{\sqrt{5}}\begin{pmatrix}\frac{1+\sqrt{5}}{2}&\frac{1-\sqrt{5}}{2}\\1&1\end{pmatrix}\begin{pmatrix}\left(\frac{1+\sqrt{5}}{2}\right)^{n-2}&0\\0&\left(\frac{1-\sqrt{5}}{2}\right)^{n-2}\end{pmatrix}\begin{pmatrix}\frac{1+\sqrt{5}}{2}\\-\frac{1-\sqrt{5}}{2}\end{pmatrix}\\
&=\frac{1}{\sqrt{5}}\begin{pmatrix}\frac{1+\sqrt{5}}{2}&\frac{1-\sqrt{5}}{2}\\1&1\end{pmatrix}\begin{pmatrix}\left(\frac{1+\sqrt{5}}{2}\right)^{n-1}\\-\left(\frac{1-\sqrt{5}}{2}\right)^{n-1}\end{pmatrix}\end{aligned}$$

$$=\frac{1}{\sqrt{5}}\begin{pmatrix}\left(\frac{1+\sqrt{5}}{2}\right)^{n}-\left(\frac{1-\sqrt{5}}{2}\right)^{n}\\ \left(\frac{1+\sqrt{5}}{2}\right)^{n-1}-\left(\frac{1-\sqrt{5}}{2}\right)^{n-1}\end{pmatrix}$$

Hence

$$a_n=\frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^{n}-\left(\frac{1-\sqrt{5}}{2}\right)^{n}\right).$$

Example 6.3.3

Calculate $D_n=\begin{vmatrix}a+b & b & 0 & \cdots & 0\\ a & a+b & b & \ddots & \vdots\\ 0 & a & \ddots & \ddots & 0\\ \vdots & \ddots & \ddots & a+b & b\\ 0 & \cdots & 0 & a & a+b\end{vmatrix}_{n\times n}$. ⇒ 利用 determine 性质.

Solution

Expand along the first row, we have

$$D_n=\begin{vmatrix}a+b & b & 0 & \cdots & 0\\ a & a+b & b & \ddots & \vdots\\ 0 & a & \ddots & \ddots & 0\\ \vdots & \ddots & \ddots & a+b & b\\ 0 & \cdots & 0 & a & a+b\end{vmatrix}_{n\times n}$$

$$=(a+b)\begin{vmatrix}a+b & b & 0 & \cdots & 0\\ a & a+b & b & \ddots & \vdots\\ 0 & a & \ddots & \ddots & 0\\ \vdots & \ddots & \ddots & a+b & b\\ 0 & \cdots & 0 & a & a+b\end{vmatrix}_{(n-1)\times(n-1)}-b\begin{vmatrix}a & b & 0 & \cdots & 0\\ 0 & a+b & b & \ddots & \vdots\\ 0 & a & \ddots & \ddots & 0\\ \vdots & \ddots & \ddots & a+b & b\\ 0 & \cdots & 0 & a & a+b\end{vmatrix}_{(n-1)\times(n-1)}$$

$$=(a+b)\begin{vmatrix}a+b & b & 0 & \cdots & 0\\ a & a+b & b & \ddots & \vdots\\ 0 & a & \ddots & \ddots & 0\\ \vdots & \ddots & \ddots & a+b & b\\ 0 & \cdots & 0 & a & a+b\end{vmatrix}_{(n-1)\times(n-1)}-ab\begin{vmatrix}a+b & b & 0 & \cdots & 0\\ a & a+b & b & \ddots & \vdots\\ 0 & a & \ddots & \ddots & 0\\ \vdots & \ddots & \ddots & a+b & b\\ 0 & \cdots & 0 & a & a+b\end{vmatrix}_{(n-2)\times(n-2)}$$

$$=(a+b)D_{n-1}-abD_{n-2}$$

⇓ 矩阵化.

Then $\begin{pmatrix} D_n \\ D_{n-1} \end{pmatrix} = \begin{pmatrix} (a+b)D_{n-1} - abD_{n-2} \\ D_{n-1} \end{pmatrix} = \begin{pmatrix} a+b & -ab \\ 1 & 0 \end{pmatrix} \begin{pmatrix} D_{n-1} \\ D_{n-2} \end{pmatrix}$.

The characteristic polynomial of $\begin{pmatrix} a+b & -ab \\ 1 & 0 \end{pmatrix}$ is $\lambda^2 - (a+b)\lambda + ab = (\lambda - a)(\lambda - b)$.

Case (i): $a \neq b$ ⇒ Diagonalize

$\lambda = a, b$ are eigenvalues corresponding to eigenvectors $\begin{pmatrix} a \\ 1 \end{pmatrix}, \begin{pmatrix} b \\ 1 \end{pmatrix}$ respectively. Then

$$\begin{pmatrix} a+b & -ab \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} a & b \\ 1 & 1 \end{pmatrix}^{-1}$$

Hence

$$\begin{pmatrix} a+b & -ab \\ 1 & 0 \end{pmatrix}^n = \begin{pmatrix} a & b \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}^n \begin{pmatrix} a & b \\ 1 & 1 \end{pmatrix}^{-1} = \frac{1}{a-b} \begin{pmatrix} a & b \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a^n & 0 \\ 0 & b^n \end{pmatrix} \begin{pmatrix} 1 & -b \\ -1 & a \end{pmatrix}.$$

$$\begin{aligned}
\begin{pmatrix} D_n \\ D_{n-1} \end{pmatrix} &= \begin{pmatrix} a+b & -ab \\ 1 & 0 \end{pmatrix} \begin{pmatrix} D_{n-1} \\ D_{n-2} \end{pmatrix} \\
&= \begin{pmatrix} a+b & -ab \\ 1 & 0 \end{pmatrix}^{n-2} \begin{pmatrix} D_2 \\ D_1 \end{pmatrix} \\
&= \frac{1}{a-b} \begin{pmatrix} a & b \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a^{n-2} & 0 \\ 0 & b^{n-2} \end{pmatrix} \begin{pmatrix} 1 & -b \\ -1 & a \end{pmatrix} \begin{pmatrix} D_2 \\ D_1 \end{pmatrix} \\
&= \frac{1}{a-b} \begin{pmatrix} a & b \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a^{n-2} & -ba^{n-2} \\ -b^{n-2} & ab^{n-2} \end{pmatrix} \begin{pmatrix} D_2 \\ D_1 \end{pmatrix} \\
&= \frac{1}{a-b} \begin{pmatrix} a^{n-1} - b^{n-1} & ab^{n-1} - ba^{n-1} \\ a^{n-2} - b^{n-2} & ab^{n-2} - ba^{n-2} \end{pmatrix} \begin{pmatrix} D_2 \\ D_1 \end{pmatrix}
\end{aligned}$$

Since $D_1 = a+b$ and $D_2 = \begin{vmatrix} a+b & b \\ a & a+b \end{vmatrix} = (a+b)^2 - ab = a^2 + ab + b^2$,

直接求所要的幂就行

$$\begin{aligned}
D_n &= \frac{1}{a-b}\left(\left(a^2+ab+b^2\right)\left(a^{n-1}-b^{n-1}\right) + (a+b)\left(ab^{n-1} - ba^{n-1}\right)\right) \\
&= \frac{1}{a-b}\left(\left(a^{n+1} + a^n b + a^{n-1}b^2\right) - \left(a^2 b^{n-1} + ab^n + b^{n+1}\right) + \left(a^2 b^{n-1} - ba^n\right) + \left(ab^n - b^2 a^{n-1}\right)\right) \\
&= \frac{1}{a-b}\left(a^{n+1} - b^{n+1}\right)
\end{aligned}$$

两种对角化 $\begin{cases} A = PDP^{-1} \\ A = PJP^{-1} \end{cases} \Rightarrow \begin{cases} A^n = PD^nP^{-1} \\ A^n = PJ^nP^{-1} \end{cases}$

Case (ii): $a=b$ ⇒ jordan form.

$\lambda = a$ is an eigenvalue of $\begin{pmatrix} 2a & -a^2 \\ 1 & 0 \end{pmatrix}$ (with multiplicity = 2) corresponding to eigenvector $\begin{pmatrix} a \\ 1 \end{pmatrix}$.

$A=\begin{pmatrix} 2a & -a^2 \\ 1 & 0 \end{pmatrix}$ is not diagonalizable. Consider $K_a = N\left((A-aI_2)^2\right) = N\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

⇒但不在 eigenspace里.

Clearly, $\begin{pmatrix} 1 \\ 0 \end{pmatrix} \in K_a$ and $\beta = \left\{ (A-aI_2)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$ forms a basis of $K_a = R^2$. Then

important conclusion

$$\begin{pmatrix} 2a & -a^2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}\begin{pmatrix} a & 1 \\ 1 & 0 \end{pmatrix}^{-1}.$$

=二项式展开

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}^n = \left(\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right)^n$$ binomial expansion

两次以上项全为0

$$= \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}^n + n\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}^{n-1}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a^n & na^{n-1} \\ 0 & a^n \end{pmatrix}$$

$$\begin{pmatrix} 2a & -a^2 \\ 1 & 0 \end{pmatrix}^n = \begin{pmatrix} a & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}^n\begin{pmatrix} a & 1 \\ 1 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} a & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a^n & na^{n-1} \\ 0 & a^n \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & -a \end{pmatrix}.$$

$$\begin{aligned}
\begin{pmatrix} D_n \\ D_{n-1} \end{pmatrix} &= \begin{pmatrix} 2a & -a^2 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} D_{n-1} \\ D_{n-2} \end{pmatrix} \\
&= \begin{pmatrix} 2a & -a^2 \\ 1 & 0 \end{pmatrix}^{n-2}\begin{pmatrix} D_2 \\ D_1 \end{pmatrix} \\
&= \begin{pmatrix} a & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a^{n-2} & (n-2)a^{n-3} \\ 0 & a^{n-2} \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & -a \end{pmatrix}\begin{pmatrix} D_2 \\ D_1 \end{pmatrix} \\
&= \begin{pmatrix} a & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} (n-2)a^{n-3} & -(n-3)a^{n-2} \\ a^{n-2} & -a^{n-1} \end{pmatrix}\begin{pmatrix} D_2 \\ D_1 \end{pmatrix} \\
&= \begin{pmatrix} (n-1)a^{n-2} & -(n-2)a^{n-1} \\ (n-2)a^{n-3} & -(n-3)a^{n-2} \end{pmatrix}\begin{pmatrix} D_2 \\ D_1 \end{pmatrix}
\end{aligned}$$

Since $D_1 = 2a$ and $D_2 = \begin{vmatrix} 2a & a \\ a & 2a \end{vmatrix} = 3a^2$, ✓

$$D_n = 3(n-1)a^n - 2(n-2)a^n = (n+1)a^n.$$

代数法

$\downarrow\downarrow\downarrow$

Any recurrence relation of the form $x_n = ax_{n-1}$ is called a first order homogeneous linear recurrence relation. Its solution is

$$x_n = Ca^n.$$

$\Uparrow$ b=0's special case.

Any recurrence relation of the form

$$x_n = ax_{n-1} + bx_{n-2} \tag{1}$$

is called a second order homogeneous linear recurrence relation. $r^2 - ar - b = 0$ is called the characteristic equation of (1). In particular, if $b = 0$, (1) reduces to first order homogeneous linear recurrence relation.

Theorem 6.3.4

If the characteristic equation has two distinct roots r_1 and r_2, then the homogeneous solution for (1) is given by

$$x_n = C_1 r_1^n + C_2 r_2^n.$$

If the characteristic equation has only one root r, then the homogeneous solution for (1) is given by

$$x_n = C_1 r_1^n + C_2 n r_1^n.$$

A recurrence relation of the form

General form $\Rightarrow$

$$x_n = ax_{n-1} + bx_{n-2} + f(n) \tag{2}$$

is called a non-homogeneous recurrence relation.

Let $x_n^{(p)}$ be a solution of (2), called a particular solution. Then the general solution for (2) is

$$x_n = x_n^{(h)} + x_n^{(p)}$$

homogeneous solution + particular solution

where $x_n^{(h)}$ is the homogeneous solution of recurrence relation of $x_n = ax_{n-1} + bx_{n-2}$.

The solution that we guess is suggested by the form of $f(n) = \left(c_t n^t + c_{t-1} n^{t-1} + \cdots + c_0\right) r^n$:

deal with $f(n)$:

step 1: 将 $f(n)$ 的形式变成 $f(n) = (A_m n^m + A_{m-1} n^{m-1} + \cdots A_0) r^n$

step 2: 考虑 $f(n)$ 中的 r^n. r 在不在 r_1 r_2 里.

$\nearrow$ 不在: $x_n^{(p)} = (d_t n^t + \cdots d_0) r^n$.

Step3: ↘ 在 : $x_n^{(p)} = n^m(d_t n + d_{t-1}n^{t-1} + \cdots d_0)r^n$

↓ multiplicity

- If r is not a root of the characteristic equation of the associated homogeneous recurrence relation, then $x_n^{(p)} = \left(d_t n^t + d_{t-1}n^{t-1} + \cdots + d_0\right) r^n$.
- If r is a root with multiplicity m, then $x_n^{(p)} = n^m\left(d_t n^t + d_{t-1}n^{t-1} + \cdots + d_0\right) r^n$.

Example 6.3.5

Find $D_n = \begin{vmatrix} a+b & b & \cdots & b \\ a & a+b & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ a & \cdots & a & a+b \end{vmatrix}_{n\times n}$.

Solution

$$D_n = \begin{vmatrix} a+b & b & b & \cdots & b \\ a & a+b & b & \ddots & \vdots \\ a & a & \ddots & \ddots & b \\ \vdots & \ddots & \ddots & a+b & b \\ a & \cdots & a & a & a+b \end{vmatrix}_{n\times n}$$

为什么是这样的?

$$= \left|\begin{array}{ccccc} a & 0 & 0 & \cdots & 0 \\ \hline a & a+b & b & \cdots & b \\ a & a & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & a+b & b \\ a & \cdots & a & a & a+b \end{array}\right|_{n\times n} + \left|\begin{array}{ccccc} b & b & b & \cdots & b \\ \hline a & a+b & b & \cdots & b \\ a & a & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & a+b & b \\ a & \cdots & a & a & a+b \end{array}\right|_{n\times n}$$

$$= a\begin{vmatrix} a+b & b & \cdots & b \\ a & a+b & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ a & \cdots & a & a+b \end{vmatrix}_{(n-1)\times(n-1)} + b\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ a & a+b & b & \cdots & b \\ a & a & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & a+b & b \\ a & \cdots & a & a & a+b \end{vmatrix}_{n\times n}$$

$a_n = a_{n-1} + a_{n-2}$.

$r^2 - r - 1 = 0 \Rightarrow r = \frac{1\pm\sqrt{5}}{2}$

$a_n = c_1\left(\frac{1+\sqrt{5}}{2}\right)^n + c_2\left(\frac{1-\sqrt{5}}{2}\right)^n \Rightarrow \begin{cases} a_1 = 1 \\ a_2 = 1 \end{cases}$ 结合求 c_1/c_2.

$$= aD_{n-1} + b\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & b & b-a & \cdots & b-a \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b & b-a \\ 0 & \cdots & 0 & 0 & b \end{vmatrix}_{n\times n}$$

[In the 2nd determinant, $R_i - aR_1 \to R_i, i \geq 2$]

$$= aD_{n-1} + b^n$$

The homogeneous solution is $D_n^{(h)} = Ca^n$.

分类讨论

If $b \neq a$, the particular solution is $D_n^{(p)} = Ab^n$. $D_n^{(p)} = aD_{n-1}^{(p)} + b^n \Rightarrow Ab^n = aAb^{n-1} + b^n$.

↓

根在不在特征方程里

$$A = \frac{b}{b-a} \quad \text{and} \quad D_n^{(p)} = \frac{b^{n+1}}{b-a}.$$

We have $D_n = D_n^{(h)} + D_n^{(p)} = Ca^n + \frac{b^{n+1}}{b-a}$. $D_1 = a + b \Rightarrow a + b = Ca + \frac{b^2}{b-a} \Rightarrow C = \frac{-a}{b-a}$

$$D_n = \frac{b^{n+1} - a^{n+1}}{b-a}.$$

If $b = a$, the particular solution is $D_n^{(p)} = Ana^n$. $D_n^{(p)} = aD_{n-1}^{(p)} + a^n \Rightarrow Ana^n = aA(n-1)a^{n-1} + a^n$.

$$\checkmark A = 1 \quad \text{and} \quad D_n^{(p)} = na^n.$$

We have $D_n = D_n^{(h)} + D_n^{(p)} = Ca^n + na^n$. $D_1 = 2a \Rightarrow C = 1$

$$\underbrace{D_n = (n+1)a^n}.$$

△ Example 6.3.6 能再做一次

Solve the recurrence relation

$$a_{n+2} - 4a_{n+1} + 3a_n = \overset{f(n)}{\boxed{-200 + 4n,}} \quad n \geq 0, \quad a_0 = 3000, \quad a_1 = 3299.$$

Solution

Here $a_n^{(h)} = C_1(3^n) + C_2(1^n) = C_1(3^n) + C_2$. Since $f(n) = -200 = -200(1^n)$ is a solution of the associated homogeneous relation, here $a_n^{(p)} = An + Bn^2$. for some constant A and B. This leads us to

$n \times (An+B) \cdot 1^n$

$$\begin{aligned} &A(n+2) - 4A(n+1) + 3An + B(n+2)^2 - 4B(n+1)^2 + 3Bn^2 = -200 + 4n, \\ &-2A - 4Bn = -200 + 4n \\ &A = 100, \quad B = -1. \end{aligned}$$

Hence $a_n = C_1(3^n) + C_2 + 100n - n^2$. With $a_0 = 3000$ and $a_1 = 3299$, we have

$$a_n = 100(3^n) + 2900 + 100n - n^2, \quad n \geq 0.$$

$a_n = a \cdot a_{n-1} + b a_{n-2} + f(n)$

①. $a_n = a \cdot a_{n-1} + b \cdot a_{n-2}$ (Homogenous)

$\Rightarrow$ Solution $\begin{cases} a_n = C_1 \cdot r_1^n + C_2 \cdot r_2^n \\ a_n = C_1 \cdot r_1^n + C_2 \cdot n \cdot r_1^n \end{cases}$

② Calculate $a_n^{(p)}$ use definition.

③ General solution: $a_n = a_n^{(h)} + a_n^{(P)}$

Section 6.4 System of Linear Differential Equations

Example 6.4.1

Solve the system of linear differential equation

确定 C 系数. $\Uparrow$

$$y''-3y'+2y=0$$

Decoupled.

with initial condition $y(0)=5,\ y'(0)=2.$

$$\begin{cases} y_1' = 5y_1 \\ y_2' = 3y_2 \end{cases} \Rightarrow \begin{cases} y_1 = C_1 e^{5x} \\ y_2 = C_2 e^{3x} \end{cases}$$

Solution

coupled

Clearly, $\begin{pmatrix} y'' \\ y' \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} y' \\ y \end{pmatrix}.$

目标: Decouple → coupled.

The characteristic polynomial of $\begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix}$ is $\det\begin{pmatrix} \lambda-3 & 2 \\ -1 & \lambda-0 \end{pmatrix} = \lambda^2-3\lambda+2=(\lambda-1)(\lambda-2).$

$\lambda = 2, 1$ are eigenvalues corresponding to eigenvectors $\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ respectively. Then

对角化.

$$\begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

We have

$$\begin{pmatrix} y'' \\ y' \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} y' \\ y \end{pmatrix}$$

$$\text{看成整体}\ \underbrace{\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} y'' \\ y' \end{pmatrix}} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\underbrace{\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} y' \\ y \end{pmatrix}}\ \text{看成整体.}$$

Let $\begin{pmatrix} u(x) \\ v(x) \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} y' \\ y \end{pmatrix} = \begin{pmatrix} y'-y \\ -y'+2y \end{pmatrix}$. Then $\begin{pmatrix} u'(x) \\ v'(x) \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} u(x) \\ v(x) \end{pmatrix}$, i.e.,

$$\begin{cases} u'(x) = 2u(x) \\ v'(x) = v(x) \end{cases}$$

Hence $u(x) = C_1 e^{2x}$ and $v(x) = C_2 e^{x}$.

At $x=0$, $\begin{pmatrix} u(0) \\ v(0) \end{pmatrix} = \begin{pmatrix} y'(0)-y(0) \\ -y'(0)+2y(0) \end{pmatrix} \Rightarrow \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} -3 \\ 8 \end{pmatrix}.$

$y' = xy$

$y' \cdot e^{-\frac{x^2}{2}} = x \cdot e^{-\frac{x^2}{2}} y$

$y' \cdot e^{-\frac{x^2}{2}} - x \cdot e^{-\frac{x^2}{2}} y = 0$

$\frac{d(y e^{-\frac{x^2}{2}})}{dx} = 0$

$y \cdot e^{-\frac{x^2}{2}} = C$

$y = C \cdot e^{\frac{x^2}{2}}$

积分因子.
令 微分 方便微分.

$\downarrow$

Generlize: $y' = P(x) y$

$\mu(x) = e^{\int -P(x) dx}$

积分因子

$$y' \cdot e^{\int -P(x)dx} = P(x)e^{\int -P(x)dx}\, y$$

$$\Rightarrow \frac{d(ye^{\int -P(x)dx})}{dx} = 0$$

$$y = C \cdot e^{\int P(x)dx}$$

$$\begin{pmatrix} y' \\ y \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} u(x) \\ v(x) \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -3e^{2x} \\ 8e^{x} \end{pmatrix} = \begin{pmatrix} -6e^{2x} + 8e^{x} \\ -3e^{2x} + 8e^{x} \end{pmatrix}.$$

$y(x) = -3e^{2x} + 8e^{x}$.

Example 6.4.2

Solve the system of linear differential equation

$$\text{coupled} \Rightarrow \quad \begin{aligned} y_1' &= y_1 + 3y_2 \\ y_2' &= 4y_1 + 2y_2 \end{aligned}$$

with initial condition $y_1(0) = 5,\ y_2(0) = 2$.

条件表示.

Solution

By Example 6.1.10, $\begin{pmatrix} y_1' \\ y_2' \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 4 & -1 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 4 & -1 \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$

Let $\begin{pmatrix} u(t) \\ v(t) \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 4 & -1 \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \dfrac{1}{-7} \begin{pmatrix} -1 & -1 \\ -4 & 3 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$. Then $\begin{pmatrix} u'(t) \\ v'(t) \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} u(t) \\ v(t) \end{pmatrix}$, i.e.,

$$\begin{cases} u'(t) = 5u(t) \\ v'(t) = -2v(t) \end{cases} \Rightarrow \text{Decoupled.}$$

Hence $u(t) = C_1 e^{5t}$ and $v(t) = C_2 e^{-2t}$.

待定系数法.

At $t = 0$, $\begin{pmatrix} u(0) \\ v(0) \end{pmatrix} = \dfrac{1}{-7} \begin{pmatrix} -y_1(0) - y_2(0) \\ -4y_1(0) + 3y_2(0) \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

$$\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 4 & -1 \end{pmatrix} \begin{pmatrix} u(t) \\ v(t) \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 4 & -1 \end{pmatrix} \begin{pmatrix} e^{5t} \\ 2e^{-2t} \end{pmatrix} = \begin{pmatrix} 3e^{5t} + 2e^{-2t} \\ 4e^{5t} - 2e^{-2t} \end{pmatrix}.$$

Definition 6.4.3

For any $n \times n$ matrix A, the **matrix exponential** e^{A} is defined in terms of the convergent power series

$$e^{A} = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots$$

connect to taylor series:

$$e^{x} = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots$$

$$\Downarrow$$

$$e^{A} = I_n + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots$$

$$D=\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}_{n\times n} \Rightarrow e^D=\begin{pmatrix}1 & & \\ & \ddots & \\ & & 1\end{pmatrix}_{n\times n}+\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}+\frac{1}{2!}\begin{pmatrix}\lambda_1^2 & & & \\ & \lambda_2^2 & & \\ & & \ddots & \\ & & & \lambda_n^2\end{pmatrix}+\cdots\frac{1}{n!}\begin{pmatrix}\lambda_1^n & & & \\ & \lambda_2^n & & \\ & & \ddots & \\ & & & \lambda_n^n\end{pmatrix}$$

各项相加 $\Downarrow$

$$=\begin{pmatrix}1+\lambda_1+\frac{1}{2!}\lambda_1^2+\cdots\frac{1}{n!}\lambda_1^n & & & \\ & 1+\lambda_2+\frac{1}{2!}\lambda_2^2+\cdots\frac{1}{n!}\lambda_2^n & & \\ & & \ddots & \\ & & & 1+\lambda_n+\frac{1}{2!}\lambda_n+\cdots\frac{1}{n!}\lambda_n^n\end{pmatrix}=\begin{pmatrix}e^{\lambda_1} & & & \\ & e^{\lambda_2} & & \\ & & \ddots & \\ & & & e^{\lambda_n}\end{pmatrix}$$

In the case of a diagonal matrix

$$D=\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}$$

Proof

the matrix exponential can be defined as follows:

$$e^D=\lim_{m\to\infty}\left(I+D+\frac{1}{2!}D^2+\cdots+\frac{1}{m!}D^m\right)$$
$$=\lim_{m\to\infty}\begin{pmatrix}\sum_{k=1}^{m}\frac{\lambda_1^k}{k!} & & & \\ & \sum_{k=1}^{m}\frac{\lambda_2^k}{k!} & & \\ & & \ddots & \\ & & & \sum_{k=1}^{m}\frac{\lambda_n^k}{k!}\end{pmatrix}$$
$$=\begin{pmatrix}e^{\lambda_1} & & & \\ & e^{\lambda_2} & & \\ & & \ddots & \\ & & & e^{\lambda_n}\end{pmatrix}$$

Properities:

1. $A=PDP^{-1} \Rightarrow D=P^{-1}AP$

$$e^A=I+A+\frac{1}{2!}A^2+\cdots\frac{1}{n!}A^n$$
$$=I+PDP^{-1}+\frac{1}{2!}(PDP^{-1})^2+\cdots\frac{1}{n!}(PDP^{-1})^n$$
$$=P(I+D+\frac{1}{2!}D^2+\cdots\frac{1}{n!}D^n)P^{-1}$$

$$e^A=Pe^DP^{-1}$$

Proof.

For a general $n\times n$ matrix A, it is difficult to compute the matrix exponent. However if A is diagonalizable, then $A^k=PD^kP^{-1}$ for $k=1,2,\cdots$ where P is formed by eigenvector of A.

$$e^{tA}=I+tA+\frac{1}{2!}(tA)^2+\frac{1}{3!}(tA)^3+\cdots$$
$$=PIP^{-1}+P(tD)P^{-1}+\frac{1}{2!}P(tD)^2P^{-1}+\frac{1}{3!}P(tD)^3P^{-1}+\cdots$$
$$=P\left(I+(tD)+\frac{1}{2!}(tD)^2+\frac{1}{3!}(tD)^3+\cdots\right)P^{-1}$$
$$=Pe^{tD}P^{-1}$$

Remark 6.4.4

$e^B\cdot e^A\neq e^A\cdot e^B\neq e^{A+B}$ in general.

Proof:

$$e^{A+B}=I+(A+B)+\frac{1}{2!}(A+B)(A+B)+\cdots$$

$$e^A=I+A+\frac{1}{2!}A^2+\cdots$$

也不同

$$e^B=I+B+\frac{1}{2!}B^2+\cdots$$

不定相同 ✓

$$e^{A} \cdot e^{B} = I^{2} + B + A + \boxed{AB} + \cdots$$

$$e^{B} \cdot e^{A} = I^{2} + A + B + \boxed{BA} + \cdots$$

Let $A=\begin{pmatrix}1&0\\0&0\end{pmatrix}$ and $B=\begin{pmatrix}0&0\\1&0\end{pmatrix}$. Clearly, $A^2=\begin{pmatrix}1&0\\0&0\end{pmatrix}^2=\begin{pmatrix}1&0\\0&0\end{pmatrix}=A$, $B^2=\begin{pmatrix}0&0\\1&0\end{pmatrix}^2=\begin{pmatrix}0&0\\0&0\end{pmatrix}$.

Then $A+B=\begin{pmatrix}1&0\\1&0\end{pmatrix}$ and $(A+B)^2=\begin{pmatrix}1&0\\1&0\end{pmatrix}^2=\begin{pmatrix}1&0\\1&0\end{pmatrix}=A+B$.

$$\begin{aligned}e^{A}&=I_2+A+\frac{1}{2!}A^2+\frac{1}{3!}A^3+\cdots\\&=I_2+\left(1+\frac{1}{2!}+\frac{1}{3!}+\cdots\right)A\\&=\begin{pmatrix}1&0\\0&1\end{pmatrix}+(e-1)A\\&=\begin{pmatrix}e&0\\0&1\end{pmatrix}.\end{aligned}$$

$$e^{B}=I_2+B+\frac{1}{2!}B^2+\cdots=I_2+B=\begin{pmatrix}1&0\\0&1\end{pmatrix}+\begin{pmatrix}0&0\\1&0\end{pmatrix}=\begin{pmatrix}1&0\\1&1\end{pmatrix}.$$

$$e^{A}\cdot e^{B}=\begin{pmatrix}e&0\\0&1\end{pmatrix}\begin{pmatrix}1&0\\1&1\end{pmatrix}=\begin{pmatrix}e&0\\1&1\end{pmatrix}.$$

$$e^{B}\cdot e^{A}=\begin{pmatrix}1&0\\1&1\end{pmatrix}\begin{pmatrix}e&0\\0&1\end{pmatrix}=\begin{pmatrix}e&0\\e&1\end{pmatrix}.$$

$$\begin{aligned}e^{A+B}&=I_2+(A+B)+\frac{1}{2!}(A+B)^2+\frac{1}{3!}(A+B)^3+\cdots\\&=I_2+\left(1+\frac{1}{2!}+\frac{1}{3!}+\cdots\right)(A+B)\\&=\begin{pmatrix}1&0\\0&1\end{pmatrix}+(e-1)(A+B)\\&=\begin{pmatrix}e&0\\e-1&1\end{pmatrix}.\end{aligned}$$

They are not equal to each other because $AB=\begin{pmatrix}0&0\\0&0\end{pmatrix}\neq\begin{pmatrix}0&0\\1&0\end{pmatrix}=BA$.

If $AB=BA$, then $e^{B}\cdot e^{A}=e^{A}\cdot e^{B}=e^{A+B}$.

Example 6.4.5

Compute e^{tA} for $A=\begin{pmatrix}1&3\\4&2\end{pmatrix}$.

$$e^{tA}=Pe^{tD}P^{-1}$$

$= P e^{tJ} P^{-1}$

$\vec{m}_i' = A \cdot \vec{m}_i$

$M = (\vec{m}_1 \ \vec{m}_2 \ \cdots \ \vec{m}_n)$

$\downarrow$

$(\vec{m}_1', \vec{m}_2' + \cdots \vec{m}_n')$

$(\vec{Y}(t))' = A(Y(t)) \quad = A(\vec{m}_1, \vec{m}_2 \cdots \vec{m}_n)$

$M = e^{tA} = I + tA + \frac{1}{2!}(tA)^2 + \cdots \frac{1}{n!}(tA)^n$

$M' = A + tA^2 + \cdots + \frac{t^{n-1}}{(n-1)!} A^n$

$= A[I + tA \cdots \frac{t^{n-1}}{(n-1)!} A^{n-1} + \cdots] = A \cdot e^{tA} = AM$

$\vec{Y}(t) = c_1 \vec{m}_1 + c_2 \vec{m}_2 + \cdots c_n \vec{m}_n$

$= (\vec{m}_1, \vec{m}_2 \cdots \vec{m}_n)\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$

$= M \cdot \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$

$\downarrow$

$= e^{tA} \cdot \vec{C}$

$(\vec{Y}(t))' = A(\vec{Y}(t))$

$\vec{Y} = e^{tA}$

$Y' = A\vec{Y} \quad \vec{Y} = e^{tA}$

Similarly. $y' = ay \quad y = Ce^{ax}$

Solution

By Example 6.1.10, $P = \begin{pmatrix} 3 & 1 \\ 4 & -1 \end{pmatrix}$ and $D = \begin{pmatrix} 5 & 0 \\ 0 & -2 \end{pmatrix}$.

compute step by step.

$$\begin{aligned}
e^{tA} &= \begin{pmatrix} 3 & 1 \\ 4 & -1 \end{pmatrix}\begin{pmatrix} e^{5t} & 0 \\ 0 & e^{-2t} \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 4 & -1 \end{pmatrix}^{-1} \\
&= \frac{1}{-7}\begin{pmatrix} 3e^{5t} & e^{-2t} \\ 4e^{5t} & -e^{-2t} \end{pmatrix}\begin{pmatrix} -1 & -1 \\ -4 & 3 \end{pmatrix} \\
&= \frac{1}{7}\begin{pmatrix} 3e^{5t} + 4e^{-2t} & 3e^{5t} - 3e^{-2t} \\ 4e^{5t} - 4e^{-2t} & 4e^{5t} + 3e^{-2t} \end{pmatrix}
\end{aligned}$$

Theorem 6.4.6

Consider the initial value problem $\boldsymbol{Y}' = A\boldsymbol{Y}$, $\boldsymbol{Y}(0) = \boldsymbol{Y}_0$. Then the solution can be express as

$$\boldsymbol{Y}(t) = e^{tA}\boldsymbol{Y}_0.$$

Example 6.4.7

Use the matrix exponential to solve the initial value problem

$$\begin{aligned}
y_1' &= y_1 + 3y_2 \\
y_2' &= 4y_1 + 2y_2
\end{aligned}$$

with initial condition $y_1(0) = 5,\ y_2(0) = 2$.

$D = \begin{pmatrix} 5 & 0 \\ 0 & -2 \end{pmatrix} \quad tD = \begin{pmatrix} 5t & 0 \\ 0 & -2t \end{pmatrix}$

Solution

By Example 6.4.5,

$e^{tD} = P\begin{pmatrix} e^{5t} & \\ & e^{-2t} \end{pmatrix}P^{-1}$

$$\begin{aligned}
\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} &= \frac{1}{7}\begin{pmatrix} 3e^{5t} + 4e^{-2t} & 3e^{5t} - 3e^{-2t} \\ 4e^{5t} - 4e^{-2t} & 4e^{5t} + 3e^{-2t} \end{pmatrix}\begin{pmatrix} 5 \\ 2 \end{pmatrix} \\
&= \frac{1}{7}\begin{pmatrix} 21e^{5t} + 14e^{-2t} \\ 28e^{5t} - 14e^{-2t} \end{pmatrix} \\
&= \begin{pmatrix} 3e^{5t} + 2e^{-2t} \\ 4e^{5t} - 2e^{-2t} \end{pmatrix} \quad \text{same as the answer in Example 6.4.2}
\end{aligned}$$

Example 6.4.8 [illegible]级-级

Solve the initial value problem

$$\boldsymbol{Y}' = A\boldsymbol{Y}, \quad \boldsymbol{Y}(0) = \boldsymbol{Y}_0.$$

where $A = \begin{pmatrix} 9 & 4 & 5 \\ -4 & 0 & -3 \\ -6 & -4 & -2 \end{pmatrix}$, $\boldsymbol{Y}_0 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

Solution

Method 1

By Example 6.1.31, $\begin{pmatrix} y_1' \\ y_2' \\ y_3' \end{pmatrix} = \begin{pmatrix} 9 & 4 & 5 \\ -4 & 0 & -3 \\ -6 & -4 & -2 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 3 \\ -1 & 0 & -2 \\ -2 & -1 & -2 \end{pmatrix}\left(\begin{array}{cc|c} 2 & 1 & 0 \\ 0 & 2 & 0 \\ \hline 0 & 0 & 3 \end{array}\right)\underbrace{\begin{pmatrix} -2 & -1 & -2 \\ 2 & 2 & 1 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}}_{\text{看成整体}}$

Let $\begin{pmatrix} u(t) \\ v(t) \\ w(t) \end{pmatrix} = \begin{pmatrix} -2 & -1 & -2 \\ 2 & 2 & 1 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$. Then $\begin{pmatrix} u'(t) \\ v'(t) \\ w'(t) \end{pmatrix} = \left(\begin{array}{cc|c} 2 & 1 & 0 \\ 0 & 2 & 0 \\ \hline 0 & 0 & 3 \end{array}\right)\begin{pmatrix} u(t) \\ v(t) \\ w(t) \end{pmatrix}$, i.e.,

$$\begin{cases} u'(t) = 2u(t) + v(t) \\ v'(t) = 2v(t) \\ w'(t) = 3w(t) \end{cases}$$

Hence $v(t) = C_2 e^{2t}$ and $w(t) = C_3 e^{3t}$.

$$\begin{aligned} u'(t) = 2u(t) + v(t) \quad &\Rightarrow \quad u'(t) - 2u(t) = C_2 e^{2t} \\ &\Rightarrow \quad e^{-2t}u'(t) - 2e^{-2t}u(t) = C_2 \\ &\Rightarrow \quad \left(e^{-2t}u(t)\right)' = C_2 \\ &\Rightarrow \quad e^{-2t}u(t) = C_1 + C_2 t \\ &\Rightarrow \quad u(t) = \left(C_1 + C_2 t\right)e^{2t} \end{aligned}$$

↓ 操作目的: 右边是常数

At $t = 0$, $\begin{pmatrix} u(0) \\ v(0) \\ w(0) \end{pmatrix} = \begin{pmatrix} -2 & -1 & -2 \\ 2 & 2 & 1 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} y_1(0) \\ y_2(0) \\ y_3(0) \end{pmatrix} = \begin{pmatrix} -2 & -1 & -2 \\ 2 & 2 & 1 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -10 \\ 9 \\ 4 \end{pmatrix} \Rightarrow \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} -10 \\ 9 \\ 4 \end{pmatrix}$.

$$\begin{pmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{pmatrix} = \begin{pmatrix} 2 & 1 & 3 \\ -1 & 0 & -2 \\ -2 & -1 & -2 \end{pmatrix}\begin{pmatrix} u(t) \\ v(t) \\ w(t) \end{pmatrix}$$

解决 Jordan form 有两种方法

1) coupled → decoupled

② 分块 (Jordan block) 取高阶级

e^{2+2} $e^2 \cdot e^2$

$$= \begin{pmatrix}2&1&3\\-1&0&-2\\-2&-1&-2\end{pmatrix}\begin{pmatrix}(-10+9t)e^{2t}\\9e^{2t}\\4e^{3t}\end{pmatrix}$$

$$= \begin{pmatrix}(-11+18t)e^{2t}+12e^{3t}\\(10-9t)e^{2t}-8e^{3t}\\(11-18t)e^{2t}-8e^{3t}\end{pmatrix}.$$

Method 2

In Example 6.1.31, we have $A = PJP^{-1}$, where $P = \begin{pmatrix}2&1&3\\-1&0&-2\\-2&-1&-2\end{pmatrix}$ and $J = \left(\begin{array}{cc|c}2&1&0\\0&2&0\\\hline 0&0&3\end{array}\right)$.

$$A^n = PJ^nP^{-1} = \begin{pmatrix}2&1&3\\-1&0&-2\\-2&-1&-2\end{pmatrix}\left(\begin{array}{c|c}\boxed{\begin{pmatrix}2&1\\0&2\end{pmatrix}^n}&\begin{matrix}0\\0\end{matrix}\\\hline \begin{matrix}0&0\end{matrix}&\boxed{3^n}\end{array}\right)\begin{pmatrix}-2&-1&-2\\2&2&1\\1&0&1\end{pmatrix}$$

Let $B = \begin{pmatrix}2&1\\0&2\end{pmatrix}$ and $E = \begin{pmatrix}0&1\\0&0\end{pmatrix} = B - 2I_2$. Since $E^2 = \begin{pmatrix}0&0\\0&0\end{pmatrix}$, $e^{tE} = I_2 + tE = \begin{pmatrix}1&t\\0&1\end{pmatrix}$. We have

$\begin{bmatrix}e^t&0\\0&e^t\end{bmatrix}$

$$e^{tB} = e^{2tI_2}\cdot e^{tE} = \begin{pmatrix}e^{2t}&0\\0&e^{2t}\end{pmatrix}(I_2+tE) = \boxed{e^{2t}}\begin{pmatrix}1&t\\0&1\end{pmatrix}$$ and

$$e^{tA} = Pe^{tJ}P^{-1} = P\left(\begin{array}{c|c}e^{tB}&\begin{matrix}0\\0\end{matrix}\\\hline \begin{matrix}0&0\end{matrix}&e^{3t}\end{array}\right)P^{-1} = P\left(\begin{array}{cc|c}e^{2t}&te^{2t}&0\\0&e^{2t}&0\\\hline 0&0&e^{3t}\end{array}\right)P^{-1}.$$

$$\begin{pmatrix}y_1(t)\\y_2(t)\\y_3(t)\end{pmatrix} = \begin{pmatrix}2&1&3\\-1&0&-2\\-2&-1&-2\end{pmatrix}\left(\begin{array}{cc|c}e^{2t}&te^{2t}&0\\0&e^{2t}&0\\\hline 0&0&e^{3t}\end{array}\right)\begin{pmatrix}-2&-1&-2\\2&2&1\\1&0&1\end{pmatrix}\begin{pmatrix}1\\2\\3\end{pmatrix}$$

$$= \begin{pmatrix}2e^{2t}&2te^{2t}+e^{2t}&3e^{3t}\\-e^{2t}&-te^{2t}&-2e^{3t}\\-2e^{2t}&-2te^{2t}-e^{2t}&-2e^{3t}\end{pmatrix}\begin{pmatrix}-10\\9\\4\end{pmatrix}$$

$$= \begin{pmatrix}(-11+18t)e^{2t}+12e^{3t}\\(10-9t)e^{2t}-8e^{3t}\\(11-18t)e^{2t}-8e^{3t}\end{pmatrix}$$

Section 6.5 Markov Chain

Definition 6.5.1

随机的

A **stochastic** process is any sequence of experiments for which the outcome at any stage depends on chance. A **Markov** process is a stochastic process with the following properties:

I. The set of possible outcomes or states is finite.

II. The probability of the next outcome depends only on the previous outcome.

III. The probabilities are constant over time.

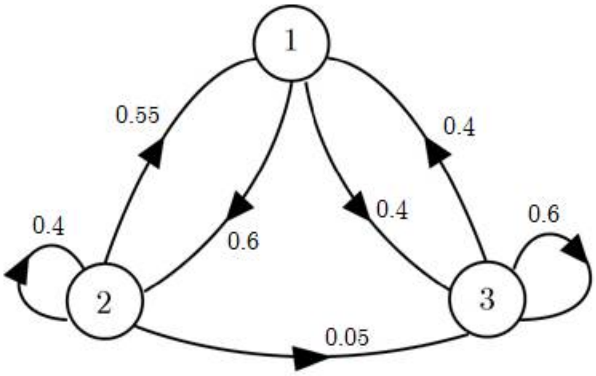
Definition 6.5.2

Any square matrix $\boldsymbol{P}$ having (i) nonnegative entries and (ii) all rows that sum to 1, is called a **transition matrix**. For an arbitrary $n \times n$ transition matrix, the rows and columns correspond to n states, and the entry $\boldsymbol{P}_{ij}$ represents the probability of moving from state i to state j in one stage.

转换矩阵

Example 6.5.3

Consider the following Markov chain



We have transition matrix $\boldsymbol{P}$.

$$\begin{array}{cc} & \overbrace{\begin{array}{ccc}1 & \quad 2 & \quad 3\end{array}}^{\text{To}} \\ \text{From}\left\{\begin{matrix}1\\2\\3\end{matrix}\right. & \begin{pmatrix}1\to1 & 1\to2 & 1\to3\\ 2\to1 & 2\to2 & 2\to3\\ 3\to1 & 3\to2 & 3\to3\end{pmatrix}\end{array} = \begin{pmatrix}0 & 0.6 & 0.4\\ 0.55 & 0.4 & 0.05\\ 0.4 & 0 & 0.6\end{pmatrix}$$

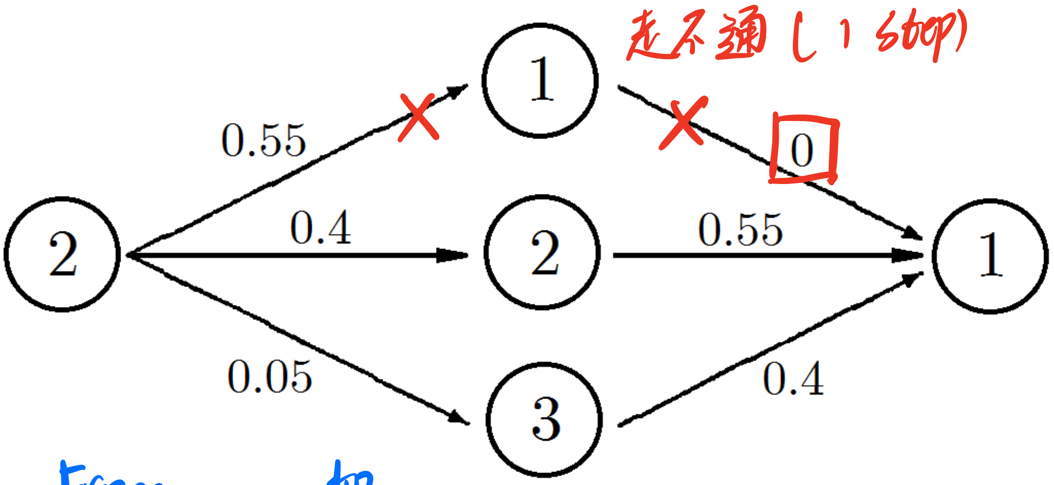
transition matrix

P_{ij}: the probability that goes from i to j with 1 step

P^2_{ij} : the probability that goes from i to j with 2 steps

② → ① (要两步)

Let us calculate the probability of moving to state 1 from state 2 in 2 transitions.



From   to

$$P^2 = \begin{pmatrix} * & * & * \\ 0.55 & 0.4 & 0.05 \\ * & * & * \end{pmatrix}\begin{pmatrix} 0 & * & * \\ 0.55 & * & * \\ 0.4 & * & * \end{pmatrix} = \begin{pmatrix} * & * & * \\ (0)(0.55)+(0.4)(0.55)+(0.05)(0.4) & * & * \\ * & * & * \end{pmatrix}$$

从②出发   到达 ①.

P^2 is the transition matrix in 2 years.

Theorem 6.5.4 (**Chapman–Kolmogorov**)

In general, for any transition matrix $\boldsymbol{P}$, the entry $\left(\boldsymbol{P}^m\right)_{ij}$ represents the probability of moving from state i to state j in m years.

↓

多少 step m 就是多少

Theorem 6.5.5

Every transition matrix has 1 as an eigenvalue with eigenvector $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$.

Example 6.5.6

Consider Example 6.5.3. Find $\boldsymbol{P}^m$. Hence verify Theorem 6.5.4 for $i = 2, j = 3$ and $m = 4$.

Solution

Characteristic polynomial of $\boldsymbol{P}$ is

x ⇒ initial probability $(P^m)_{ij}$ the probability of

$P(\vec{X}_0) =$

$$\det(\lambda I_3 - \boldsymbol{P}) = \det\begin{pmatrix} \lambda & -0.6 & -0.4 \\ -0.55 & \lambda-0.4 & -0.05 \\ -0.4 & 0 & \lambda-0.6 \end{pmatrix}$$
$$= \lambda^3 - \lambda^2 - 0.25\lambda + 0.25$$
$$= (\lambda-1)\left(\lambda^2 - 0.25\right) \quad [1 \text{ is an eigenvalue}]$$
$$= (\lambda-1)(\lambda-0.5)(\lambda+0.5)$$

Its eigenvalues are $\lambda_1 = 1, \lambda_2 = 0.5$ and $\lambda_3 = -0.5$.

1. For $\lambda_1 = 1$, the corresponding eigenvector is $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. Theorem 6.5.5

2. For $\lambda_2 = 0.5$, consider $N(0.5I_3 - \boldsymbol{P}) = N\begin{pmatrix} 0.5 & -0.6 & -0.4 \\ -0.55 & 0.1 & -0.05 \\ -0.4 & 0 & -0.1 \end{pmatrix} = span\left\{\begin{pmatrix} 2 \\ 7 \\ -8 \end{pmatrix}\right\}$. The corresponding eigenvector is $\begin{pmatrix} 2 \\ 7 \\ -8 \end{pmatrix}$.

3. For $\lambda_3 = -0.5$, consider $N(-0.5I_3 - \boldsymbol{P}) = N\begin{pmatrix} -0.5 & -0.6 & -0.4 \\ -0.55 & -0.9 & -0.05 \\ -0.4 & 0 & -1.1 \end{pmatrix} = span\left\{\begin{pmatrix} 22 \\ -13 \\ -8 \end{pmatrix}\right\}$. The corresponding eigenvector is $\begin{pmatrix} 22 \\ -13 \\ -8 \end{pmatrix}$.

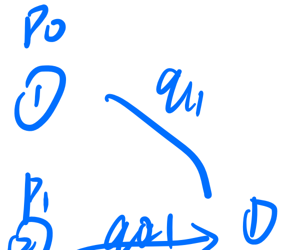
Then

$$\boldsymbol{P} = \begin{pmatrix} 0 & 0.6 & 0.4 \\ 0.55 & 0.4 & 0.05 \\ 0.4 & 0 & 0.6 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & -0.5 \end{pmatrix}\begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix}^{-1}$$
$$= \frac{1}{100}\begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & -0.5 \end{pmatrix}\begin{pmatrix} 32 & 32 & 36 \\ 1 & 6 & -7 \\ 3 & -2 & -1 \end{pmatrix}.$$

$\vec{X}_0 = (P_0, P_1, P_2)$

$$(P_0, P_1, P_2)\begin{pmatrix} q_{11} & q_{12} & q_{13} \\ a_{21} & q_{22} & q_{23} \end{pmatrix}$$

可能



$\begin{pmatrix} a_{31} & a_{32} & a_{33} \end{pmatrix}$

以状态.

从 ③

a_{31}

第一步后在 ① 的概率

$$\boldsymbol{P}^m = \cancel{\frac{1}{100}}\begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & -0.5 \end{pmatrix}^m \begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix}^{-1}$$

$$= \frac{1}{100}\begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5^m & 0 \\ 0 & 0 & (-0.5)^m \end{pmatrix}\begin{pmatrix} 32 & 32 & 36 \\ 1 & 6 & -7 \\ 3 & -2 & -1 \end{pmatrix}$$

$$\left(\boldsymbol{P}^4\right)_{2,3} = \frac{1}{100}\begin{pmatrix} 1 & 7 & -13 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5^4 & 0 \\ 0 & 0 & (-0.5)^4 \end{pmatrix}\begin{pmatrix} 36 \\ -7 \\ -1 \end{pmatrix} = \frac{1}{100}\left[36 - 49(0.5^4) + 13(-0.5)^4\right] = 0.3375$$

选P中第二行

选P中第三列

$$P(2 \xrightarrow{0.55} 1 \xrightarrow{0.6} 2 \xrightarrow{0.55} 1 \xrightarrow{0.4} 3) = 0.55^2 \times 0.6 \times 0.4 = 0.0726$$

$$P(2 \xrightarrow{0.55} 1 \xrightarrow{0.6} 2 \xrightarrow{0.4} 2 \xrightarrow{0.05} 3) = 0.55 \times 0.6 \times 0.4 \times 0.05 = 0.0066$$

$$P(2 \xrightarrow{0.55} 1 \xrightarrow{0.6} 2 \xrightarrow{0.05} 3 \xrightarrow{0.6} 3) = 0.55 \times 0.6^2 \times 0.05 = 0.0099$$

$$P(2 \xrightarrow{0.55} 1 \xrightarrow{0.4} 3 \xrightarrow{0.4} 1 \xrightarrow{0.4} 3) = 0.55 \times 0.4^3 = 0.0352$$

$$P(2 \xrightarrow{0.55} 1 \xrightarrow{0.4} 3 \xrightarrow{0.6} 3 \xrightarrow{0.6} 3) = 0.55 \times 0.4 \times 0.6^2 = 0.0792$$

$$P(2 \xrightarrow{0.4} 2 \xrightarrow{0.55} 1 \xrightarrow{0.6} 2 \xrightarrow{0.05} 3) = 0.55 \times 0.6 \times 0.4 \times 0.05 = 0.0066$$

$$P(2 \xrightarrow{0.4} 2 \xrightarrow{0.55} 1 \xrightarrow{0.4} 3 \xrightarrow{0.6} 3) = 0.55 \times 0.6 \times 0.4^2 = 0.0528$$

$$P(2 \xrightarrow{0.4} 2 \xrightarrow{0.4} 2 \xrightarrow{0.55} 1 \xrightarrow{0.4} 3) = 0.4^3 \times 0.55 = 0.0352$$

$$P(2 \xrightarrow{0.4} 2 \xrightarrow{0.4} 2 \xrightarrow{0.4} 2 \xrightarrow{0.05} 3) = 0.4^3 \times 0.05 = 0.0032$$

$$P(2 \xrightarrow{0.4} 2 \xrightarrow{0.4} 2 \xrightarrow{0.05} 3 \xrightarrow{0.6} 3) = 0.4^2 \times 0.05 \times 0.6 = 0.0048$$

$$P(2 \xrightarrow{0.4} 2 \xrightarrow{0.05} 3 \xrightarrow{0.4} 1 \xrightarrow{0.4} 3) = 0.4^3 \times 0.05 = 0.0032$$

$$P(2 \xrightarrow{0.4} 2 \xrightarrow{0.05} 3 \xrightarrow{0.6} 3 \xrightarrow{0.6} 3) = 0.4 \times 0.05 \times 0.6^2 = 0.0072$$

$$P(2 \xrightarrow{0.05} 3 \xrightarrow{0.4} 1 \xrightarrow{0.6} 2 \xrightarrow{0.05} 3) = 0.05^2 \times 0.4 \times 0.6 = 0.0006$$

$$P(2 \xrightarrow{0.05} 3 \xrightarrow{0.4} 1 \xrightarrow{0.4} 3 \xrightarrow{0.6} 3) = 0.05 \times 0.4^2 \times 0.6 = 0.0048$$

$$P(2 \xrightarrow{0.05} 3 \xrightarrow{0.6} 3 \xrightarrow{0.4} 1 \xrightarrow{0.4} 3) = 0.05 \times 0.4^2 \times 0.6 = 0.0048$$

$$P(2 \xrightarrow{0.05} 3 \xrightarrow{0.6} 3 \xrightarrow{0.6} 3 \xrightarrow{0.6} 3) = 0.05 \times 0.6^3 = 0.0108$$

Sum of all these probabilities is exactly 0.3375.

Definition 6.5.7

A row vector contains nonnegative entries that sum to 1 is called a **probability vector**.

Theorem 6.5.8

The product of two $n \times n$ transition matrices is an $n \times n$ transition matrix. The product of a transition matrix and a probability vector is a probability vector.

$x_1 = P(x_{01}, x_{02}\ \ x_{03})$

经过m步后在 state 1 的

$x_0 = P(x_{11}, x_{12}, x_{13})$ 概率

Definition 6.5.9

A **Markov chain** is a sequence of probability vectors $\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \cdots$ together with a transition matrix $\boldsymbol{P}$ such that

$$\boldsymbol{x}_1 = \boldsymbol{x}_0\boldsymbol{P}, \quad \boldsymbol{x}_2 = \boldsymbol{x}_1\boldsymbol{P} = \boldsymbol{x}_0\boldsymbol{P}^2, \quad \boldsymbol{x}_3 = \boldsymbol{x}_2\boldsymbol{P} = \boldsymbol{x}_0\boldsymbol{P}^3, \quad \cdots$$

$\boldsymbol{x}_k$ is called a k-th **state vector**. $x_n = x_0 \cdot P^n$

Example 6.5.10

Consider Example 6.5.3. Find the limiting probability $\lim_{m\to\infty} \boldsymbol{x}_0\boldsymbol{P}^m$.

Solution

不用考虑 x_0?

$$\lim_{m\to\infty} \boldsymbol{P}^m = \lim_{m\to\infty} \begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & -0.5 \end{pmatrix}^m \begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix}^{-1}$$

$$= \lim_{m\to\infty} \begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5^m & 0 \\ 0 & 0 & (-0.5)^m \end{pmatrix} \begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix}^{-1}$$

$$= \frac{1}{100} \begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 32 & 32 & 36 \\ 1 & 6 & -7 \\ 3 & -2 & -1 \end{pmatrix}$$

$$= \frac{1}{100} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 32 & 32 & 36 \\ 1 & 6 & -7 \\ 3 & -2 & -1 \end{pmatrix}$$

$$= \frac{1}{100} \begin{pmatrix} 32 & 32 & 36 \\ 32 & 32 & 36 \\ 32 & 32 & 36 \end{pmatrix}$$

①	②
0.1	0.90
0.1	0.05
0.8	0.05

$\lim_{m\to\infty} \boldsymbol{x}_0\boldsymbol{P}^m = (0.32, 0.32, 0.36)$.

走到这个结果

Definition 6.5.11

Let $\boldsymbol{P}$ be a transition matrix. A probability vector π is called a **steady-state vector** for $\boldsymbol{P}$ if $\pi\boldsymbol{P} = \pi$.

$$\begin{cases} \pi P = \pi \\ \pi P^2 = \pi \end{cases}$$

找到一个特殊的 π 永远是 π

Probability vector. $\xrightarrow{\text{converge to}}$ steady-state vector

$\lim_{m\to\infty} x^m = x_0 \lim_{m\to\infty} P^m$

$x_0 = x_1 = x_2 \cdots x_m$

$\pi P = \pi.$

之间联系、

Example 6.5.12

Let $\boldsymbol{P}=\begin{pmatrix}0 & 0.6 & 0.4\\ 0.55 & 0.4 & 0.05\\ 0.4 & 0 & 0.6\end{pmatrix}$. Find the steady-state vector $\pi=(\pi_1,\pi_2,\pi_3)$ for $\boldsymbol{P}$.

Solution

We need to solve the equation $\pi\boldsymbol{P}=\pi$.

$$\pi\boldsymbol{P}=\pi \quad\Rightarrow\quad \pi\boldsymbol{P}-\pi=\boldsymbol{0} \quad\Rightarrow\quad \pi(\boldsymbol{P}-I_3)=\boldsymbol{0} \quad\Rightarrow\quad (\pi_1,\pi_2,\pi_3)\begin{pmatrix}-1 & 0.6 & 0.4\\ 0.55 & -0.6 & 0.05\\ 0.4 & 0 & -0.4\end{pmatrix}=(0,0,0)$$

Solving

直接解出来

$$\begin{cases}-\pi_1+0.55\pi_2+0.4\pi_3=0\\ 0.6\pi_1-0.6\pi_2=0\\ 0.4\pi_1+0.05\pi_2-0.4\pi_3=0\\ \pi_1+\pi_2+\pi_3=1\end{cases}$$

we have $\pi=(\pi_1,\pi_2,\pi_3)=(0.32,0.32,0.36)$, the same as limiting probability (Example 6.5.10). Coincidence???

↑ 终极结果

Theorem 6.5.13

Let $\lambda_1=1,\lambda_2,\cdots,\lambda_n$ be all eigenvalues of a transition matrix $\boldsymbol{P}$. If $|\lambda_i|<1$ for all $i=2,\cdots,n$ then the Markov chain $\boldsymbol{x}_0,\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots$ will converge to a steady state vector.

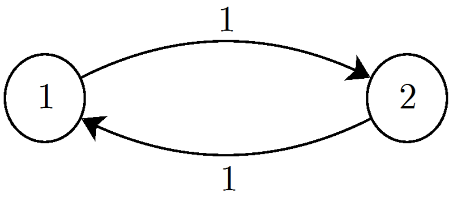
Remark 6.5.14

Limit of powers of a transition matrix need not exist.

Consider $\boldsymbol{P}=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$.

1. transition matrix P
2. calculate P^m (eigen space)
3. Probability vector (initial $\vec{x}_0$, $\vec{x}_n=\vec{x}_0P^m$
4. Steady state. vector $\pi P=\pi$

↑ [illegible] special case



1. when $\lim_{m\to\infty}\vec{x}_m$ exist?

2. Given $\lim\vec{x}_m$ exist, is it a steady-state vector?

2. Given

Set $\pi = \lim_{m\to\infty} \vec{x}_m$, whether $\pi P = \pi$

$\pi P = \vec{x}_0 \lim_{m\to\infty} P^m \cdot P = \vec{x}_0 \cdot \lim_{m\to\infty} P^{m+1}$

$= \pi$

Characteristic polynomial of $\boldsymbol{P}$ is

$$\det(\lambda I_2 - \boldsymbol{P}) = \det\begin{pmatrix} \lambda & -1 \\ -1 & \lambda \end{pmatrix} = \lambda^2 - 1 = (\lambda - 1)(\lambda + 1).$$

Its eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = -1$. Theorem 6.5.13 does not guarantee the existence of limit of powers of $\boldsymbol{P}$.

For $\lambda_1 = 1$, the corresponding eigenvector is $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

For $\lambda_2 = -1$, consider $N(-I_3 - \boldsymbol{P}) = N\begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} = span\left\{\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right\}$. The corresponding eigenvector is $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

极限不存在，
个

Then $\boldsymbol{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1}$. We have

$$\lim_{m\to\infty} \boldsymbol{P}^m = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \lim_{m\to\infty} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}^m \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1} = \lim_{m\to\infty} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & (-1)^m \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1}.$$

Since $\lim_{m\to\infty} \begin{pmatrix} 1 & 0 \\ 0 & (-1)^m \end{pmatrix}$ does not exist, $\lim_{m\to\infty} \boldsymbol{P}^m$ does not exist.

Indeed, $\lim_{m\to\infty} \boldsymbol{P}^{2m+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $\lim_{m\to\infty} \boldsymbol{P}^{2m} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Then $\lim_{m\to\infty} \boldsymbol{P}^m$ does not exist. Hence limiting probability $\lim_{m\to\infty} \boldsymbol{x}_0 \boldsymbol{P}^m$ need not exist. For instance, let $\boldsymbol{x}_0 = (1,0)$.

$$\boldsymbol{x}_0 \boldsymbol{P}^{2m+1} = (1,0)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (0,1) \text{ and } \boldsymbol{x}_0 \boldsymbol{P}^{2m} = (1,0)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = (1,0).$$

Limiting probability $\lim_{m\to\infty} \boldsymbol{x}_0 \boldsymbol{P}^m$ does not exist.

However, steady state vector π does exist.

$$\pi \boldsymbol{P} = \pi \quad \Rightarrow \quad \pi \boldsymbol{P} - \pi = \boldsymbol{0} \quad \Rightarrow \quad \pi(\boldsymbol{P} - I) = \boldsymbol{0} \quad \Rightarrow \quad (\pi_1, \pi_2)\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} = (0,0)$$

Solving

limit exist $\rightarrow$ steady-state vector.

steady-state vector $\not\Rightarrow$ limit exist.

高次元素皆大于0

$$\begin{cases} -\pi_1 + \pi_2 = 0 \\ \pi_1 + \pi_2 = 1 \end{cases}$$

we have $\pi = (\pi_1, \pi_2) = (0.5, 0.5)$.

Definition 6.5.15

A transition matrix is called **regular** if some power of the matrix contains only positive entries.

Example 6.5.16

证明 not regular

The transition matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0.2 & 0.3 & 0.5 \\ 0 & 0.4 & 0.6 \end{pmatrix}$ is not regular because the first row of A^n is $(1,0,0)$ for any positive power m.

The transition matrix $\begin{pmatrix} 0.9 & 0 & 0.1 \\ 0.5 & 0.5 & 0 \\ 0 & 0.4 & 0.6 \end{pmatrix}$ is regular because every entry of $\begin{pmatrix} 0.9 & 0 & 0.1 \\ 0.5 & 0.5 & 0 \\ 0 & 0.4 & 0.6 \end{pmatrix}^2 = \begin{pmatrix} 0.81 & 0.04 & 0.15 \\ 0.7 & 0.25 & 0.05 \\ 0.2 & 0.44 & 0.36 \end{pmatrix}$ is positive.

Theorem 6.5.17

Let $\boldsymbol{P}$ be an $n \times n$ regular transition matrix. Then

(a) The multiplicity of 1 as an eigenvalue of $\boldsymbol{P}$ is 1.

(b) $L = \lim_{m\to\infty} \boldsymbol{P}^m$ exists and is a transition matrix.

(c) $\boldsymbol{P}L = L\boldsymbol{P} = L$.

(d) Each row of L is equal to the unique probability vector π, a steady-state vector, which is also an eigenvector of $\boldsymbol{P}^T$ corresponding to the eigenvalue 1.

(e) For any probability vector $\boldsymbol{v}$, $\lim_{m\to\infty} \boldsymbol{v}\boldsymbol{P}^m = \pi$.

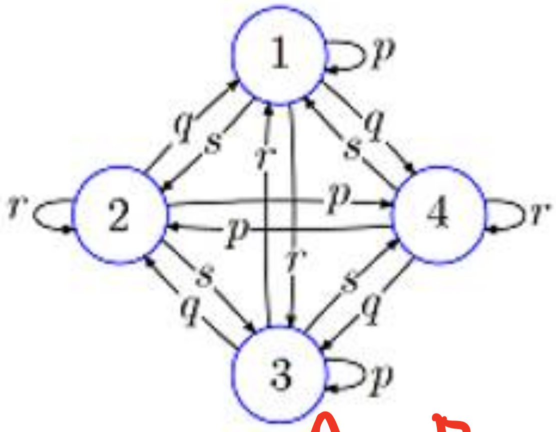
Theorem 6.5.18

Let $M_{2n\times 2n}=\left(\begin{array}{c|c} A_{n\times n} & B_{n\times n} \\ \hline C_{n\times n} & D_{n\times n}\end{array}\right)$ be a $2n\times 2n$ matrix such that $A_{n\times n}+B_{n\times n}=C_{n\times n}+D_{n\times n}$. Then the characteristic polynomial of $M_{2n\times 2n}$ is

$$\det\left(\lambda I_{2n}-M_{2n\times 2n}\right)=\det\left(\lambda I_n-(A_{n\times n}-C_{n\times n})\right)\det\left(\lambda I_n-(C_{n\times n}+D_{n\times n})\right)$$

Example 6.5.19

Consider the following Markov chain



$p=0.1, q=0.2, r=0.3, s=0.4$

Transition matrix is $\boldsymbol{P}=\overset{\begin{array}{cc}A & B\end{array}}{\underset{\begin{array}{cc}C & D\end{array}}{\left(\begin{array}{cc|cc} 0.1 & 0.4 & 0.3 & 0.2 \\ 0.2 & 0.3 & 0.4 & 0.1 \\ \hline 0.3 & 0.2 & 0.1 & 0.4 \\ 0.4 & 0.1 & 0.2 & 0.3 \end{array}\right)}}$

Since $\begin{pmatrix} 0.1 & 0.4 \\ 0.2 & 0.3 \end{pmatrix}+\begin{pmatrix} 0.3 & 0.2 \\ 0.4 & 0.1 \end{pmatrix}=\begin{pmatrix} 0.3 & 0.2 \\ 0.4 & 0.1 \end{pmatrix}+\begin{pmatrix} 0.1 & 0.4 \\ 0.2 & 0.3 \end{pmatrix}=\begin{pmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{pmatrix}$, the characteristic polynomial of $\boldsymbol{P}$ is

$$\begin{aligned}\det(\lambda I_4-\boldsymbol{P}) &= \det\left(\lambda I_2-\begin{pmatrix} -0.2 & 0.2 \\ -0.2 & 0.2 \end{pmatrix}\right)\det\left(\lambda I_2-\begin{pmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{pmatrix}\right) \\ &= \det\begin{pmatrix} \lambda+0.2 & -0.2 \\ 0.2 & \lambda-0.2 \end{pmatrix}\det\begin{pmatrix} \lambda-0.4 & -0.6 \\ -0.6 & \lambda-0.4 \end{pmatrix} \\ &= \lambda^2\left((\lambda-0.4)^2-0.6^2\right) \\ &= \lambda^2(\lambda-1)(\lambda+0.2)\end{aligned}$$

Its eigenvalues are $\lambda_1 = 1, \lambda_2 = -0.2$ and $\lambda_3 = \lambda_4 = 0$.

$$\lambda_1 = 1, N(I_4 - \boldsymbol{P}) = N\begin{pmatrix} 0.9 & -0.4 & -0.3 & -0.2 \\ -0.2 & 0.7 & -0.4 & -0.1 \\ -0.3 & -0.2 & 0.9 & -0.4 \\ -0.4 & -0.1 & -0.2 & 0.7 \end{pmatrix} = span\left\{\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}\right\}, v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\lambda_2 = -0.2, N(-0.2I_4 - \boldsymbol{P}) = N\begin{pmatrix} -0.3 & -0.4 & -0.3 & -0.2 \\ -0.2 & -0.5 & -0.4 & -0.1 \\ -0.3 & -0.2 & -0.3 & -0.4 \\ -0.4 & -0.1 & -0.2 & -0.5 \end{pmatrix} = span\left\{\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}\right\}, v_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$$

P不能对角化.

$$\lambda_3 = 0, N(0I_4 - \boldsymbol{P}) = N\begin{pmatrix} -0.1 & -0.4 & -0.3 & -0.2 \\ -0.2 & -0.3 & -0.4 & -0.1 \\ -0.3 & -0.2 & -0.1 & -0.4 \\ -0.4 & -0.1 & -0.2 & -0.3 \end{pmatrix} = span\left\{\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}\right\}$$

P is not diagonalizable ↙

Jordan form

$$N\left((0I_4 - \boldsymbol{P})^2\right) = N\begin{pmatrix} 0.26 & 0.24 & 0.26 & 0.24 \\ 0.24 & 0.26 & 0.24 & 0.26 \\ 0.26 & 0.24 & 0.26 & 0.24 \\ 0.24 & 0.26 & 0.24 & 0.26 \end{pmatrix} = span\left\{\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}\right\}$$

$$v_4 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}, v_3 = \boldsymbol{P}v_4 = \begin{pmatrix} -0.2 \\ -0.2 \\ 0.2 \\ 0.2 \end{pmatrix}, \boldsymbol{P}v_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 0.1 & 0.4 & 0.3 & 0.2 \\ 0.2 & 0.3 & 0.4 & 0.1 \\ 0.3 & 0.2 & 0.1 & 0.4 \\ 0.4 & 0.1 & 0.2 & 0.3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & -0.2 & 1 \\ 1 & -1 & -0.2 & 0 \\ 1 & 1 & 0.2 & -1 \\ 1 & -1 & 0.2 & 0 \end{pmatrix}\begin{pmatrix} 1 & & & \\ & -0.2 & & \\ & & 0 & 1 \\ & & & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 & -0.2 & 1 \\ 1 & -1 & -0.2 & 0 \\ 1 & 1 & 0.2 & -1 \\ 1 & -1 & 0.2 & 0 \end{pmatrix}^{-1}$$

$$\lim_{n\to\infty}\begin{pmatrix} 0.1 & 0.4 & 0.3 & 0.2 \\ 0.2 & 0.3 & 0.4 & 0.1 \\ 0.3 & 0.2 & 0.1 & 0.4 \\ 0.4 & 0.1 & 0.2 & 0.3 \end{pmatrix} = \lim_{n\to\infty}\begin{pmatrix} 1 & 1 & -0.2 & 1 \\ 1 & -1 & -0.2 & 0 \\ 1 & 1 & 0.2 & -1 \\ 1 & -1 & 0.2 & 0 \end{pmatrix}\begin{pmatrix} 1 & & & \\ & -0.2 & & \\ & & 0 & 1 \\ & & & 0 \end{pmatrix}^n\begin{pmatrix} 1 & 1 & -0.2 & 1 \\ 1 & -1 & -0.2 & 0 \\ 1 & 1 & 0.2 & -1 \\ 1 & -1 & 0.2 & 0 \end{pmatrix}^{-1}$$

$$= \lim_{n\to\infty}\begin{pmatrix} 1 & 1 & -0.2 & 1 \\ 1 & -1 & -0.2 & 0 \\ 1 & 1 & 0.2 & -1 \\ 1 & -1 & 0.2 & 0 \end{pmatrix}\begin{pmatrix} 1 & & \\ & (-0.2)^n & \\ & & \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^n \end{pmatrix}\begin{pmatrix} 1 & 1 & -0.2 & 1 \\ 1 & -1 & -0.2 & 0 \\ 1 & 1 & 0.2 & -1 \\ 1 & -1 & 0.2 & 0 \end{pmatrix}^{-1}$$

$$=\begin{pmatrix}1&1&-0.2&1\\1&-1&-0.2&0\\1&1&0.2&-1\\1&-1&0.2&0\end{pmatrix}\begin{pmatrix}1&&&\\&0&&\\&&0&\\&&&0\end{pmatrix}\begin{pmatrix}1&1&-0.2&1\\1&-1&-0.2&0\\1&1&0.2&-1\\1&-1&0.2&0\end{pmatrix}^{-1}$$

$$=\frac{1}{4}\begin{pmatrix}1&1&-0.2&1\\1&-1&-0.2&0\\1&1&0.2&-1\\1&-1&0.2&0\end{pmatrix}\begin{pmatrix}1&&&\\&0&&\\&&0&\\&&&0\end{pmatrix}\begin{pmatrix}1&1&1&1\\1&-1&1&-1\\0&-10&0&10\\2&-2&-2&2\end{pmatrix}$$

$$=\frac{1}{4}\begin{pmatrix}1&1&-0.2&1\\1&-1&-0.2&0\\1&1&0.2&-1\\1&-1&0.2&0\end{pmatrix}\begin{pmatrix}1&1&1&1\\&&&\\&&&\\&&&\end{pmatrix}$$

$$=\frac{1}{4}\begin{pmatrix}1&1&1&1\\1&1&1&1\\1&1&1&1\\1&1&1&1\end{pmatrix}$$

Limiting probability $\lim_{m\to\infty} \boldsymbol{x}_0\boldsymbol{P}^n=\frac{1}{4}(1,1,1,1)$ for any probability vector $\boldsymbol{x}_0$.

Consider steady state vector π. We need to solve the equation $\pi\boldsymbol{P}=\pi$.

$$\pi\boldsymbol{P}=\pi \;\Rightarrow\; \pi\boldsymbol{P}-\pi=\boldsymbol{0} \;\Rightarrow\; \pi(\boldsymbol{P}-I_4)=\boldsymbol{0} \;\Rightarrow\; (\pi_1,\pi_2,\pi_3,\pi_4)\begin{pmatrix}-0.9&0.4&0.3&0.2\\0.2&-0.7&0.4&0.1\\0.3&0.2&-0.9&0.4\\0.4&0.1&0.2&-0.7\end{pmatrix}=(0,0,0,0)$$

Solving

$$\begin{cases}-0.9\pi_1+0.4\pi_2+0.3\pi_3+0.2\pi_4=0\\0.2\;\pi_1-0.7\pi_2+0.4\pi_3+0.1\pi_4=0\\0.3\;\pi_1+0.2\pi_2-0.9\pi_3+0.4\pi_4=0\\0.4\;\pi_1+0.1\pi_2+0.2\pi_3-0.7\pi_4=0\\\pi_1+\pi_2+\pi_3+\pi_4=1\end{cases}$$

Steady state vector $\pi=(\pi_1,\pi_2,\pi_3,\pi_4)=\frac{1}{4}(1,1,1,1)$.

直接找 steady vector 简便.