MAT 2040 线性代数

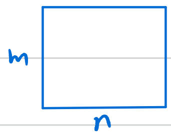
许炜源

2021 — Summer

Slide 1

1. Linear system

① m×n system
- m : # of equations
- n : # of unknowns



② Solution set

the # of elements
- 0
- 1
- ∞ (没有2,3,4…个解)

③ Equivalent system

a. Definition
- the same variables
- the same solution set

b. Theorem

$$\text{Linear System 1} \xleftrightarrow[\text{equivalent}]{\text{3 equation operations}} \text{Linear System 2}$$

(solution set remains the same)

④ Homogeneous system 齐次线性方程组

a. Definition: $\left[\ \begin{array}{c|c} & \begin{smallmatrix}0\\0\\0\end{smallmatrix}\end{array}\right]$

b. trivial solution: 零解 ; nontrivial solution: 非零解

c. eternally consistent

2. Matrix

① Row equivalence

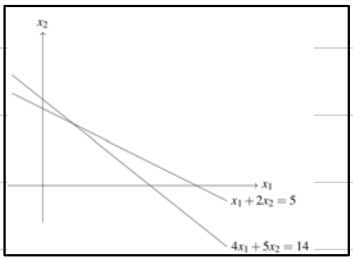
a. Elementary row (column) operations
- Interchange two rows (columns)
- $(R_i/C_i) \times k$ $(k \neq 0)$
- $(R_i/C_i) \times l + (R_j/C_j)$ $(l = R)$

b. Row equivalence

$$\text{Matrix 1} \xleftrightarrow[\text{Row equivalent}]{\text{3 elementary row operations}} \text{Matrix 2}$$

c. Equivalence of the matrix
- Row equivalence
- Column equivalence

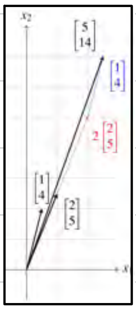
② Linear system and matrix



$$\begin{cases} x_1+2x_2=5 \\ 4x_1+5x_2=14 \end{cases}$$

$\boxed{[A|b]}$ $\left[\begin{array}{cc|c} 1 & 2 & 5 \\ 4 & 5 & 14 \end{array}\right] \longrightarrow$ Row pic

$\boxed{Ax=b}$ $\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=\begin{bmatrix} 5 \\ 14 \end{bmatrix} \longrightarrow$ Col pic $\triangleq$ $x_1\begin{bmatrix} 1 \\ 4 \end{bmatrix}+x_2\begin{bmatrix} 2 \\ 5 \end{bmatrix}=\begin{bmatrix} 5 \\ 14 \end{bmatrix} \longrightarrow$



Δ relationship: System — Augmented matrix — Ax=b (linear combination)

Equivalent Hoop

3. Solving systems

① Process

Augmented matrix $\xrightarrow{\text{Forward substitution}}$ Row echelon form $\xrightarrow{\text{Back substitution}}$ Solution set

left col → right col: make the entry below pivot = 0

by Gaussian Elimination (n×n)

Reduced row echelon form (i) → 0 (right col → left col); ii) → 1)

by Gauss—Jordan Elimination

② Row echelon form

a. Row echelon form
- nonzero row ⟶ zero row;
- 行数↑，列标(leading entry)↑ strictly

⟶ A ~ many R

b. Reduced row echelon form
- row echelon form
- nonzero leading entry is 1
- Leading 1 is the only entry of its column

⟶ A ~ unique RR

③ Variables of augmented matrix in reduced echelon form

- Lead (dependent) variable : if column i has a pivot
- free variable : if column i doesn't have a pivot

\# of lead variables = # of pivots

Leading entry in RREF

④ The cases of solution

a. Consistency
- consistent, feasible
- inconsistent, infeasible

b. # of the elements in the solution set
- Algebra exp: rank(C), rank(A) & n
- Geometric exp: # of intersection points

⟶
- Overdetermined : ① 0 ② 1 ③ ∞
- Underdetermined : ① 0 ③ ∞

Slide 2

1. Matrix notation → uppercase letter

① Matrix : $(A)_{ij} = a_{ij}$
Entry : $(a_{ij}) = A$

② Zero matrix : $0_{m \times n}$

2. Matrix Transpose

① Definition

a. In algebra

Vector i^{th} row of $A^T = a_i^T$ 注意：向量方向 / 向量由元

Entry $(A^T)_{ij} = a_{ji}$

b. In graph

Folded by main diagonal

② Properties

a. $(A+B)^T = A^T + B^T$

b. $(\alpha A)^T = \alpha A^T$

c. $(A^T)^T = A$

d. $(AB)^T = B^T A^T$

proof of d

$$\left.\begin{aligned} \left((\mathbf{AB})^{\mathrm{T}}\right)_{ij} &= (\mathbf{AB})_{ji} = \sum_{k=1}^{n} a_{jk} b_{ki} \\ \left(\mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\right)_{ij} &= \mathbf{b}_i^{\mathrm{T}} \vec{\mathbf{a}}_j^{\mathrm{T}} = \sum_{k=1}^{n} a_{jk} b_{ki} \end{aligned}\right\} (\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$$

3. Vector

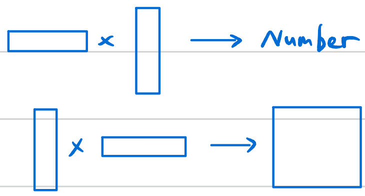
① Matrix in vector form

$$A = \begin{cases} [a_1, \cdots, a_n] \\ \begin{bmatrix} \vec{a_1} \\ \vdots \\ \vec{a_m} \end{bmatrix} \end{cases} \longrightarrow A^T = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix} = [\vec{a_1}^T \ \vec{a_2}^T \cdots \vec{a_m}^T]$$

$\vec{a_1}^T \neq a_1$

4. Matrix operation <arithmetic>

①
- Equality : $a_{ij} = b_{ij} \longrightarrow A = B$ (记为 < equal / equivalent)
- Addition : point summation
- Multiplication
  - Scalar
  - $v \times v$
    - Inner product $\xrightarrow{(1\times n)\times(n\times 1)}$ scalar
    - Outer product $\xrightarrow{(m\times 1)\times(1\times n)}$ matrix



  - $m \times v$
    - post multiplication $\begin{array}{c} a_1 a_2 a_3 \\ \begin{bmatrix} | & | & | \end{bmatrix} \end{array} \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} = \underline{b_{11}a_1 + b_{21}a_2 + b_{31}a_3}$ vector
    - pre multiplication $[b_{11}\ b_{12}\ b_{13}] \begin{bmatrix} \vec{a_1} \\ \vec{a_2} \\ \vec{a_3} \end{bmatrix} = \underline{b_{11}\vec{a_1} + b_{12}\vec{a_2} + b_{13}\vec{a_3}}$ row
  - $m \times m$
    - $(AB)_{ij} = \vec{a_i}\, b_j = \sum_{k=1}^{n} a_{ik}b_{kj}$ $\quad$ $n \rightarrow (m\times n)\times(n\times p)$ $\quad$ 乘法的表示形式：①矩阵 ②向量 ③元素
    - A is square $\longrightarrow A^k = AA\cdots A$ (k times)

We define the matrix-vector multiplication first, then by the conversion $AB = [Ab_1\ Ab_2 \cdots Ab_r]$, we define matrix multiplication

② properties

a. $A + B = B + A$

b. $\alpha(A+B) = \alpha A + \alpha B$

c. $(\alpha+\beta)A = \alpha A + \beta A$

d. $A(B+C) = AB + AC$

e. $(A+B)C = AC + BC$

} distributive

f. $\alpha(AB) = (\alpha A)B$

g. $(AB)C = A(BC)$

} associative

proof of g

Let A be an $m \times n$ matrix, B an $n \times r$ matrix, and C an $r \times s$ matrix. Let $D = AB$ and $E = BC$. We must show that $DC = AE$. By the definition of matrix multiplication,

$$d_{il} = \sum_{k=1}^{n} a_{ik}b_{kl} \quad \text{and} \quad e_{kj} = \sum_{l=1}^{r} b_{kl}c_{lj}$$

The ijth term of DC is

$$\sum_{l=1}^{r} d_{il}c_{lj} = \sum_{l=1}^{r}\left(\sum_{k=1}^{n} a_{ik}b_{kl}\right)c_{lj}$$

and the (i, j) entry of AE is

$$\sum_{k=1}^{n} a_{ik}e_{kj} = \sum_{k=1}^{n} a_{ik}\left(\sum_{l=1}^{r} b_{kl}c_{lj}\right)$$

Since

$$\sum_{l=1}^{r}\left(\sum_{k=1}^{n} a_{ik}b_{kl}\right)c_{lj} = \sum_{l=1}^{r}\left(\sum_{k=1}^{n} a_{ik}b_{kl}c_{lj}\right) = \sum_{k=1}^{n} a_{ik}\left(\sum_{l=1}^{r} b_{kl}c_{lj}\right)$$

it follows that

$$(AB)C = DC = AE = A(BC)$$

把( )内视作一个新矩阵，后讨论元素。

③ Complexity

a. Definition: Total # of operations

b. Computation:

For $A_{m\times n} \times B_{n\times k}$, Complexity $= mk(2n-1) \approx 2mkn$

$AB = [Ab_1 \ Ab_2 \ \cdots \ Ab_r]$

Ab_j has $m(2n-1)$ flops

5. Other common matrices

① Symmetric matrix & Skew-Symmetric matrix

a. Definition: If A is square $\begin{cases} \text{Symmetric matrix} & A^T = A \\ \text{Skew-Symmetric matrix} & A^T = -A \end{cases}$

b. Property:

Any square matrix A can be written as a sum of symmetric matrix and a skew-symmetric matrix:

$$\boxed{A = \underbrace{\frac{A+A^T}{2}}_{\text{Symmetric}} + \underbrace{\frac{A-A^T}{2}}_{\text{Skew-symmetric}}}$$

$\begin{cases} (A+A^T)^T = A^T + A \\ (A-A^T)^T = A^T - A \end{cases}$

② Diagonal matrix

a. Definition: $a_{ij} = 0$ when $i \neq j$ for the square matrix A

$A = \text{diag}(a_{11}, a_{22}, \cdots, a_{nn})$ (square) $= \text{diag}(a_1, a_2, \cdots, a_n)$ (此时a_i是个元素，非列向量)

b. Properties:

△ The multiplication of two diag is also diag, and $AB = \text{diag}(a_1b_1, a_2b_2, \cdots a_nb_n)$

△ The power of a diag is also diag, and $A^k = \text{diag}(a_1^k, a_2^k, \cdots, a_n^k)$

③ Identity matrix

a. Definition:

$I = I_n = \text{diag}(1, 1, \cdots, 1) \longrightarrow I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

b. Properties

△ $AI_n = I_mA = A$

△ $\alpha x = \alpha I x$

④ Triangular matrix

a. Definition:   (只需要满足白色部分为0即可)

b. Types $\begin{cases} \text{Upper triangular matrix} \\ \text{Lower triangular matrix} \end{cases}$ $\begin{cases} \text{strictly triangular matrix: all the diagonal entries are } 0 \\ \text{unit triangular matrix: all the diagonal entries are } 1 \end{cases}$

> diagonal matrix is both upper and lower triangular

c. Properties

△ $\begin{cases} L^{-1} \text{ is also lower triangular} \\ U^{-1} \text{ is also upper triangular} \end{cases}$

proof:

$[L \mid I] \to [I \mid L^{-1}]$: $L \xrightarrow[\text{Ops}]{\text{Type2 + Type3}} I$  (改变的元素在对角线下方) $\therefore I \longrightarrow L^{-1}$ Lower triangular

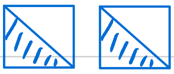
For U^{-1}, it is the same.

当 Elementary matrix 是 L | U，是为了 $\begin{cases}\text{增加} \\ \text{消除}\end{cases}$ 对应部分的元素

△ $\begin{cases} L \times L \longrightarrow L \\ U \times U \longrightarrow U \end{cases}$

proof:

Let A & B be n×n lower triangular matrix



A B

Assume $C = AB$, then

i) For $i = j$, $c_{ij} = \sum_{k=1}^{i} a_{ik}b_{kj} = a_{ii}b_{ii}$

ii) For $i < j$, $c_{ij} = 0$

$\longrightarrow$ C is lower triangular

For: $U \times U \longrightarrow U$, it is the same.

6. Matrix inversion

① Definition

A^{-1} is the inverse of A : $A^{-1}A = AA^{-1} = I$

② Properties

a. A^{-1} is unique

proof:

Assume B & C are both inverse of A, then $B = BI = B(AC) = IC = C \longrightarrow$ unique

b. $I^{-1} = I$

proof:

$[I \mid I] \to [I \mid I^{-1}] \Rightarrow I = I^{-1}$ / $\begin{cases} I \cdot I = I \\ I \cdot I = I \end{cases}$

c. $\text{diag}(a_1, \cdots, a_n)^{-1} = \text{diag}(a_1^{-1}, \cdots a_n^{-1})$ as for constant C, $C^{-1} = \frac{1}{C}$

d. $(A_1A_2 \cdots A_k)^{-1} = A_k^{-1}A_{k-1}^{-1} \cdots A_2^{-1}A_1^{-1}$

e. For square matrices A and B of the same size, if AB is invertible, then both A and B are invertible.

proof:

$|AB| \neq 0 \longrightarrow |A| \neq 0,\ |B| \neq 0$

f. If A and B are square matrices s.t. $BA = I$, then $AB = I$ $\longrightarrow$ A、B are both invertible

XX=I 是个非常好用的条件!

proof:

$A = A(BA) = (AB)A = A \Rightarrow AB = I$ (注意此处的B可能为A^T、…)

g.
$$\begin{cases} (A^T)^{-1} = (A^{-1})^T \\ (kA)^{-1} = \frac{1}{k}A^{-1} \\ (A^{-1})^{-1} = A \end{cases}$$

proof: $(A^{-1})^T \times (A)^T = (AA^{-1})^T = I^T = I$

注意: $(A+B)^{-1} \neq A^{-1} + B^{-1}$

③ Computation

a. $[A \mid I] \longrightarrow [I \mid A^{-1}]$

b. $A^{-1} = \dfrac{1}{|A|} A^*$

$|A|$: Determinant of A

A^* : $\begin{bmatrix} A_{11} & \cdots & A_{n1} \\ \vdots & \ddots & \vdots \\ A_{1n} & \cdots & A_{nn} \end{bmatrix}$, A^* is called adjoint matrix

A_{ij} in A^* : 代数余子式 $\longrightarrow$ 去除所在行、所在列元素，得新(n-1)阶矩阵，求其行列式，$A_{ij} = (-1)^{i+j}|A_{new}|$

(一阶)

拓展:
- K阶子式: 交叉部分矩阵的行列式
- K阶余子式: 除交叉部分以外矩阵的行列式
- K阶代数余子式: $(-1)^{(\Sigma行+\Sigma列)}$ · (K阶余子式)

Slide 3

1. Elementary matrix

① Definition: $I \xrightarrow{\text{once elementary row operation}} E$ (square)

② Types

a. Type 1: Row interchanges

△ properties (E_1)
- $E_1^{-1} = E_1$
- Symmetric, but not diagonal
- $E_1 A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \vec{a_1} \\ \vec{a_2} \\ \vec{a_3} \end{bmatrix} = \begin{bmatrix} \vec{a_2} \\ \vec{a_1} \\ \vec{a_3} \end{bmatrix}$

两种方式理解:
- m-v post-multiplication
- 观察E的变化 → 得出A的变化

b. Type 2: multiply k ($k \neq 0$) to a row

△ properties (E_2)
- $E_2 \xrightarrow{\text{multiply } \frac{1}{k} \text{ to the } i^{th} \text{ row}} E_2^{-1}$
- Symmetric, diagonal
- $E_2 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} \vec{a_1} \\ \vec{a_2} \\ \vec{a_3} \end{bmatrix} = \begin{bmatrix} \vec{a_1} \\ \vec{a_2} \\ 3\vec{a_3} \end{bmatrix}$

e.g. $E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \longrightarrow E_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$

c. Type 3: add $l^* R_a$ to R_b

△ properties (E_3)
- $E_3 \xrightarrow{(i,j) \text{ entry is } -l} E_3^{-1}$
- Triangular, but not symmetric
- $E_3 A = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \vec{a_1} \\ \vec{a_2} \\ \vec{a_3} \end{bmatrix} = \begin{bmatrix} \vec{a_1} + 3\vec{a_3} \\ \vec{a_2} \\ \vec{a_3} \end{bmatrix}$
- $AE_3 = [a_1 \; a_2 \; a_3] \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = [a_1 \; a_2 \; 3a_1 + a_3]$

e.g. $E_3 = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \longrightarrow E_3^{-1} = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

③ Elementary matrix & Row operation

使得 $(MA)x = Mb$ 可以推导回 $Ax = b$

a. Lemma: For any non-singular $m \times m$ matrix M, $(MA)x = Mb$ is an equivalent system to $Ax = b$.

b. Types
- Row operation —— pre-multiplying: $E_k \cdots E_2 E_1 [A \mid b]$
- Col operation —— post-multiplying: $[A \mid b] E_1 E_2 \cdots E_k$

c. Permutation matrix

△ Definition: $I \xrightarrow{\text{reorder its rows / cols}} P$, $P = E_1 E_2 \cdots E_k$ (默认列变换)

△ Properties

i) has exactly one "1" in each row/col (full rank)

ii) $P^{-1} = P^T = E_k E_{k-1} \cdots E_2 E_1$

proof:

$P^{-1} = (E_1 E_2 \cdots E_k)^{-1} = E_k^{-1} E_{k-1}^{-1} \cdots E_1^{-1} = E_k E_{k-1} \cdots E_1 = E_k^T E_{k-1}^T \cdots E_1^T = (E_1 E_2 \cdots E_k)^T = P^T$

Nature: Type 1 的行列变换可以相互抵消

2. Properties of nonsingular matrix

① Premise: square! (Only for square matrix, we consider nonsingular & singular)

② Detailed properties: (A) if and only if ⟶ 由properties证明因non-singular需要说明它是 square matrix

a. $\det \neq 0$

b. Invertible

c. $Rank = \min(n,n) = n$

d. Linear system $\begin{cases} Ax = 0 & \text{unique trivial sol} \\ Ax = b & \text{unique solution} \end{cases}$ ☆通常用该方式来求解：①先求A^{-1}；②回代求x

e. Equivalence: Row equivalent to I

f. Linear independence: Cols of A are linear independent

g. Row | Col space: $Col(A) = Row(A) = R^n$

h. $\dim(Null(A)) = 0$

Proof of (d)②

i) "Only if" Part: $A^{-1}Ax = A^{-1}b$, $x = A^{-1}b$ (unique)

ii) "If" Part (Proof by contradiction)

Assume A is singular ⟶ $\exists\, \hat{x} \neq 0$, $A\hat{x} = 0$ ⟶ $\begin{cases} Ax = b \\ A\hat{x} = 0 \end{cases}$ ⟶ $A(x+\hat{x}) = b$

⟶ $(x+\hat{x})$ is another solution ⟶ Contradiction ⟶ prove right

Proof of (e)

Since A is nonsingular, $Ax = b$ has unique sol, each col of A (reduced row echelon form) has a pivot, thus A can be converted into U, then be converted into reduced row echelon form. This reduced row echelon form matrix is I.

So, $A = E_k E_{k-1} \cdots E_1 I$ ⟶ Row equivalent to I

3. Partitioned matrix

① Definition: 十 / 一 / | (不重要)

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1t} \\ \vdots & \ddots & \vdots \\ A_{s1} & \cdots & A_{st} \end{bmatrix}$$

Partition

Block | Sub—matrix (此时用大写字母表示)

② Previous examples

属于 Partitioned matrix

- Augmented matrix $[A \mid b]$
- Row and col form $[a_1, a_2 \cdots a_n]$ / $\begin{bmatrix} \vec{a_1} \\ \vec{a_2} \\ \vdots \\ \vec{a_m} \end{bmatrix}$
- Matrix multiplication $[Ab_1 \ Ab_2 \cdots Ab_r]$

③ Block-wise multiplication

a. Premises

△ Partitions multiplication is legitimate

△ Blocks multiplication is legitimate

b. Computation

$$C_{ij} = (AB)_{ij} = \sum_k A_{ik} B_{kj} = \sum_k \sum a_{xx} b_{xx}$$

c. Partitioned matrix view of matrix multiplication

$$AB = \begin{cases} \text{Outer product expansion: } \underbrace{[a_1 \ a_2 \cdots a_n] \begin{bmatrix} \vec{b_1} \\ \vec{b_2} \\ \vdots \\ \vec{b_n} \end{bmatrix}}_{\text{"inner product"}} = \underbrace{\sum_{k=1}^{n} a_k \vec{b_k}}_{\text{Outer product}} \\ \text{"Inner product expansion": } \underbrace{\begin{bmatrix} \vec{a_1} \\ \vec{a_2} \\ \vdots \\ \vec{a_n} \end{bmatrix} [b_1 \ b_2 \cdots b_n]}_{\text{"outer product"}} = \overset{\text{inner product}}{\begin{bmatrix} \vec{a_1} b_1 & \cdots & \vec{a_1} b_n \\ \vdots & \ddots & \vdots \\ \vec{a_n} b_1 & \cdots & \vec{a_n} b_n \end{bmatrix}} = \langle \vec{a_i} b_j \rangle \end{cases}$$

看的是最终结果

⇒ ① 先把矩阵写成"Inner product"形式
② 后写出实际的 outer product expansion

d. Inverse of diagonal partition matrix

$$A = \begin{bmatrix} A_{11} & 0 & \cdots & 0 \\ 0 & A_{22} & & \\ \vdots & & \ddots & \\ 0 & & & A_{nn} \end{bmatrix} \xrightarrow{\text{Inverse}} A^{-1} = \begin{bmatrix} A_{11}^{-1} & 0 & \cdots & 0 \\ 0 & A_{22}^{-1} & & \\ \vdots & & \ddots & \\ 0 & & & A_{nn}^{-1} \end{bmatrix}$$

(Similar to inversion of diagonal matrix)

4. LU Decomposition

① Premise: square

② Process
- LU decomposition
- Substitution

a. Process I: LU decomposition $(A = LU)$

△ $E_n E_{n-1} \cdots E_1 A = U$

i> it is Gaussian elimination only with row operation III

ii> $E_1, E_2 \cdots E_n$ are all lower triangular matrix.

$\Rightarrow A \begin{cases} L = [\quad] \\ U = [\quad] \end{cases}$

△ $L = (E_n E_{n-1} \cdots E_1)^{-1} = E_1^{-1} E_2^{-1} \cdots E_n^{-1}$ 计算技巧：把 $E_1^{-1} E_2^{-1} \cdots E_{n-1}^{-1}$ 看作对 E_n^{-1} 的初等变换。

b. Process II: Forward substitution ($LUx = b$)

△ $Ly = b$

△ $Ux = y$

③ Types $\begin{cases} \text{Without row interchanges} \\ \text{With row interchanges} \end{cases}$

a. Type I: Without row interchanges

b. Type II: With row interchanges

The difference: $\begin{cases} \boxed{E_{xx} P_{xx} E_{xx} \cdots P_{xx} A = U} \\ \boxed{E_{xx} P_{xx} E_{xx} \cdots P_{xx} b = c} \end{cases}$ The only difference is in Gaussian elimination

$Ax = b$ 两边都要换，因为是方程

④ Significance

a. Lower the complexity

	div	mul	add
$[\mathbf{A} \quad \mathbf{I}] \xrightarrow{G.J.} [\mathbf{I} \quad \mathbf{A}^{-1}]$	$\approx n^2$	$\approx n^3$	$\approx n^3$
$\mathbf{A}^{-1}\mathbf{b}$		n^2	$n(n-1)$
$\mathbf{A} = \mathbf{LU}$	$\approx \frac{n^2}{2}$	$\approx \frac{n^3}{3}$	$\approx \frac{n^3}{3}$
Substitution	n	$n(n-1)$	$n(n-1)$

The normal process of finding solution

△ Find A^{-1} by "$[A \mid I] \longrightarrow [I \mid A^{-1}]$"

△ $\begin{cases} x = A^{-1}b \\ X = A^{-1}B \end{cases}$

b. Separate the decomposition and substitution, then avoid repeating (相当于预处理)

$AX = B \longrightarrow A[x_1 \ x_2 \cdots x_n] = [b_1 \ b_2 \cdots b_n]$

Same coefficient matrix $A \Rightarrow \begin{cases} \text{Once } LU \text{ decomposition} \\ \text{Multiple times substitution} \end{cases}$

Slide 4

1. Linear combination

① Definition: $a_1, a_2, \cdots, a_n$ are col. vectors of the same size, $c_1, c_2, \cdots, c_n$ are scalars,

then "$c_1a_1 + c_2a_2 + \cdots + c_na_n$" is a linear combination

if $b = c_1a_1 + c_2a_2 + \cdots + c_na_n$, then b is a linear combination of $a_1, a_2, \cdots, a_n$ (vector)

$b \in \text{Span}\{a_1, a_2, \cdots, a_n\}$

② Geometric interpretation

column graph
- Addition
- Scaling

2. Linear span

① Definition: the set of all linear combinations of $v_1, v_2, \cdots, v_n$, denoted by $\text{Span}\{v_1, v_2, \cdots, v_n\}$

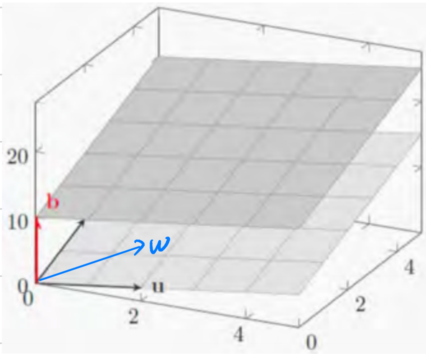
Spanning set $\{v_1, v_2, \cdots, v_n\}$ $\xrightarrow{\text{Spans}}$ $\text{Span}\{v_1, v_2, \cdots, v_n\}$

A set of vectors — A set of linear combination

"vector": vector / matrix / function

A set of vectors

* Remark: $b + W = \{b + w : w \in \text{Span}\{u, v\}\}$



线+平面⇒平面

② Linear span is the parametric vector form of the solution set of the consistent linear system

a. For Homogeneous system, the solution set is Span{vectors of free variables}

eg.

$$\begin{aligned} 2x_1 + x_2 + 7x_3 - 7x_4 &= 0 \\ -3x_1 + 4x_2 - 5x_3 - 6x_4 &= 0 \\ x_1 + x_2 + 4x_3 - 5x_4 &= 0 \end{aligned} \Rightarrow \left[\begin{array}{cccc|c} 2 & 1 & 7 & -7 & 0 \\ -3 & 4 & -5 & -6 & 0 \\ 1 & 1 & 4 & -5 & 0 \end{array}\right] \to \left[\begin{array}{cccc|c} \boxed{1} & 0 & 3 & -2 & 0 \\ 0 & \boxed{1} & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \Rightarrow \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -3x_3 + 2x_4 \\ -x_3 + 3x_4 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

Span{vectors of free variables}

b. For Non-Homogeneous system, the solution set is {Span{vectors of free variables} + constant vector}

eg.

$$\begin{aligned} 2x_1 + x_2 + 7x_3 - 7x_4 &= 8 \\ -3x_1 + 4x_2 - 5x_3 - 6x_4 &= -12 \\ x_1 + x_2 + 4x_3 - 5x_4 &= 4 \end{aligned} \Rightarrow \begin{array}{ccccc} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 & \mathbf{b} \end{array} \left[\begin{array}{cccc|c} 2 & 1 & 7 & -7 & 8 \\ -3 & 4 & -5 & -6 & -12 \\ 1 & 1 & 4 & -5 & 4 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & 0 & 3 & -2 & \boxed{4} \\ 0 & 1 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \Rightarrow \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -3x_3 + 2x_4 + \boxed{4} \\ -x_3 + 3x_4 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

{Span{vectors of free variables} + constant vector}

* Remark: Same A (coefficient matrix)

⇒ Same Span{vectors of free variables}

尽量把后面的变量转化为自由未知数

如
- x_1, x_3
- x_2, x_3 ✓: 更好

c. Theorem and Corollary derived from a & b.

△Theorem

If $\begin{cases} A\hat{x}=b \\ Az=0 \end{cases}$, $\underline{Ay=b \Leftrightarrow y=\hat{x}+z}$

即前面的向量可以不同的线性组合形成span的多条，而常数向量则应保持不变。

Nature: Span + C = Span + C + Span = Span + C / Null(A) + C = Null(A) + C + Null(A) = Null(A) + C

△ Corollary

$Ax=b$ has a unique sol $\Leftrightarrow$ $Ax=0$ has only the trivial sol

proof:

i> $Ax=b$ has a unique sol $\longrightarrow$ for $y=\hat{x}+z$, $\begin{cases} y \text{ is unique} \\ \hat{x} \text{ is unique} \end{cases}$ $\longrightarrow y=\hat{x} \longrightarrow z=0$

ii> $Ax=0$ has only the trivial sol $\longrightarrow z=0 \longrightarrow y=\hat{x} \longrightarrow Ax=b$ has a unique sol

3. Linear independence

① Definition

a. For linear combination

Vectors are $\begin{cases} \text{Linearly independent: } c_1v_1+c_2v_2+\cdots c_nv_n=0, \text{ all the scalars } c_1, c_2, \cdots c_n \text{ are } 0 \\ \text{Linearly dependent: } c_1v_1+c_2v_2+\cdots c_nv_n=0, \text{ there exists the scalars } c_1, c_2 \cdots c_n \text{ not all } 0 \end{cases}$

b. For matrix form

$\boxed{x_1a_1+x_2a_2+\cdots+x_na_n=0} \longrightarrow \boxed{[a_1\ a_2\ \cdots\ a_n]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}=0} \longrightarrow \boxed{Ax=0}$

判断线性相关：① Definition

② Theorem

$\begin{cases} \text{i> } A \xrightarrow{\text{converts}} [a_1\ a_2\ \cdots a_n] \\ \text{ii> matrix × vector (post-multiplication)} \end{cases}$

Vectors are $\begin{cases} \text{Linearly independent: the system } Ax=0 \text{ has only the trivial sol} \\ \text{Linearly dependent: the system } Ax=0 \text{ has infinitely many sols} \end{cases}$

② Theorem and Corollary

a. Theorem

Vectors $v_1, v_2, \cdots, v_n$ are linearly dependent $\longleftrightarrow$ for certain $k \in \{1, 2, \cdots, n\}$, v_k is a linear combination of other vectors proof:

i) $\longrightarrow$: Assume $c_1 \neq 0 \Rightarrow v_1 = -\frac{1}{c_1}\sum_{i=2}^{n} c_i v_i$;

ii) $\longleftarrow$: $v_1 = \sum_{i=2}^{n} c_i' v_i \Rightarrow 1 v_1 - \sum_{i=2}^{n} c_i' v_k = 0$

b. Corollary

If $b \in \text{Span}\{v_1, v_2, \cdots v_n\}$, then $b, v_1, \cdots v_n$ are linearly dependent

③ Judge the linear independence

- a. By definition
- b. By theorem

4. Vector space

① Two premises

a. Closed set

if a set V that satisfies
- Addition: For any $x, y \in V$, $(x+y)$ is unique and $(x+y) \in V$
- Scalar multiplication: For any $x \in V$, αx is unique and $\alpha x \in V$

then V is a closed set

b. Vector space axioms

A1. $\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$, $\forall \mathbf{x},\mathbf{y} \in \mathcal{V}$

A2. $(\mathbf{x}+\mathbf{y})+\mathbf{z}=\mathbf{x}+(\mathbf{y}+\mathbf{z})$ for any $\mathbf{x},\mathbf{y},\mathbf{z} \in \mathcal{V}$

A3. There exists $\mathbf{0} \in \mathcal{V}$ such that $\mathbf{x}+\mathbf{0}=\mathbf{x}$ for $\forall \mathbf{x} \in \mathcal{V}$.

A4. For each $\mathbf{x} \in \mathcal{V}$, there exists $\mathbf{x}' \in \mathcal{V}$ such that $\mathbf{x}+\mathbf{x}'=\mathbf{0}$

A5. $\alpha(\mathbf{x}+\mathbf{y})=\alpha\mathbf{x}+\alpha\mathbf{y}$ for each scalar α and any $\mathbf{x},\mathbf{y} \in \mathcal{V}$

A6. $(\alpha+\beta)\mathbf{x}=\alpha\mathbf{x}+\beta\mathbf{x}$ for any scalars α and β and $\forall \mathbf{x} \in \mathcal{V}$

A7. $(\alpha\beta)\mathbf{x}=\alpha(\beta\mathbf{x})$ for any scalars α and β and any $\mathbf{x} \in \mathcal{V}$.

A8. $1\mathbf{x}=\mathbf{x}$, $\forall \mathbf{x} \in \mathcal{V}$.

c. The relation between two premises

$$\text{Closed set} \underset{\otimes}{\overset{\checkmark}{\rightleftarrows}} \text{Eight axioms}$$

② Definition

a. Defined with set of vectors / matrices

A set that satisfies the above two premises is a vector space

* Remarks

i) vector space = linear space

ii) vector is an element in vector space

iii) $\left(\begin{array}{l} R^n \\ R^{m\times n} \end{array}\right.$ are both vector space

b. Defined with the set of functions

Let $C[a,b]$ denote the set of all defined and continuous functions on $[a,b]$.

domain

e.g. $\sin(x) \in C[-\pi,\pi]$

If $f,g \in C[a,b]$, then they satisfy $\left\{\begin{array}{ll} \text{Addition:} & (f+g)(x) = f(x)+g(x) \\ \text{Scalar multiplication:} & (af)(x) = af(x) \end{array}\right.$ Judge whether a set is $C[a,b]$

New function

e.g. Polynomial function $\left(\begin{array}{l} p(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} \\ q(x) = b_0 + b_1x + \cdots + b_{n-1}x^{n-1} \end{array}\right.$, then $\left(\begin{array}{l} (p+q)(x) = p(x)+q(x) \\ (ap)(x) = ap(x) \end{array}\right.$

③ Properties

a. The zero vector $0 \in V$ is unique

proof:

Suppose 0 and $0'$ are both zero vectors. We have $0 \overset{A_3}{=} 0+0' \overset{A_1}{=} 0'+0 \overset{A_3}{=} 0'$

b. $0x = \mathbf{0}$ for each $x \in V$

proof:

$x \overset{A_8}{=} 1x = (1+0)x \overset{A_6}{=} x+0x$

$0 \overset{A_4}{=} -x+x = -x+(x+0x) \overset{A_2}{=} (-x+x)+0x \overset{A_4}{=} 0+0x \overset{A_1,A_3}{=\!=} 0x$

c. $c0 = 0$ for any scalar c.

proof:

$c0 = c0+0 = c0+(c0+(-c0)) = (c0+c0)+(-c0) = c(0+0)+(-c0) = c0+(-c0) = 0$

d. For any $x,y \in V$, if $y+x = 0$, then $y = -x$

proof:

$y \overset{A_3}{=} y+0 \overset{A_4}{=} y+(x+(-x))$

e. $(-1)x = -x$ for each $x \in V$.

proof:

$x + (-1)x = 1x + (-1)x = (1 + (-1))x = 0x = 0$

5. Subspace

① Definition

a. Defined with the set of vectors / matrices

a nonempty, closed subset of V $\left(\text{subset} \underset{\checkmark}{\overset{\times}{\rightleftarrows}} \text{subspace}\right)$

Conditions
- a. $0 \in S$ (nonempty)
- b. $x + y \in S$
- c. $ax \in S$ for any scalar a

Remarks:

i) $\{0\}$ and V are two subspaces of V

ii) R^2 is not a subspace of R^3, but $\{(x_1, x_2, 0) \mid x_1, x_2 \in R\}$ is subspace of R^3

b. Defined with the set of functions

Let $C^n[a,b]$ be the set of all functions f that have a continuous n^{th} derivative on $[a,b]$

then $C^n[a,b]$ is a $\left(\begin{array}{l}\text{subset} \\ \text{subspace}\end{array}\right.$ of $C^{n-1}[a,b]$

$C^0[a,b] \triangleq C[a,b]$

Proved by examples

i) Subset: $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$ $\underline{(a_0 = 0)}$ $\Rightarrow$ $\left\{\begin{array}{l} f'(x) = a_1 + 2a_2x + 3a_3x^3 + \cdots + ma_mx^{m-1} \\ f''(x) = 2a_2 + 6a_3x + \cdots + m(m-1)a_mx^{m-2} \end{array}\right.$

$\left\{\begin{array}{l} f(x) \in C^2[a,b] \\ f(x), f'(x) \in C^1[a,b] \end{array}\right.$ $\Rightarrow C^2[a,b] \subset C^1[a,b]$

ii) Subspace: Let $p, q \in C^2[a,b]$, then $\left\{\begin{array}{l} 0 \in C^2[a,b] \\ p+q \in C^2[a,b] \\ ap \in C^2[a,b] \end{array}\right.$ $\Rightarrow C^2[a,b]$ is a subspace

② Theorem: Span is a subspace of vector space.

Corollary: The solution set of a homogeneous linear system is a subspace

Slide 5

1. Basis

① Two Theorems

a. Theorem 1: Minimal spanning set ⟶ the nature of basis

Suppose $S=\{v_1,v_2,\cdots,v_n\}$ is a minimal spanning set of V, then the elements of S are linearly independent

Conditions
- i> Minimal spanning set can span V
- ii> Linearly independent

proof:

Assume the elements are linearly dependent, i.e., there exists v_k that $v_k=\sum_{i\neq k} c_i' v_i$,

then $v=\sum_1^n c_i v_i=\sum_{i\neq k} c_i v_i + c_k v_k=\sum_{i\neq k} c_i v_i + c_k \sum_{i\neq k} c_i' v_i=\underbrace{\sum_{i\neq k}(c_i+c_k c_i')v_i}_{\text{a smaller span}}$

⟶ It is not the minimal spanning set

b. Theorem 2: Linear independence and Uniqueness ⟶ the uniqueness of coordinate

Iff $v_1,\cdots,v_n$ are linearly independent, $u\in \text{Span}\{v_1,v_2,\cdots v_n\}$ can be written uniquely as a linear combination of $v_1,v_2,\cdots v_n$

the constant $c_1, c_2 \cdots c_n$ are unique

Proof: (both by contradiction) ⟶ 反证的思路：假设反面，结论 ⟶ 证其与条件相悖

i> Linearly independent ⟶ Uniqueness

Assume $u\in \text{Span}\{v_1,v_2\cdots v_n\}$ has two expressions,

$u=\alpha_1 v_1+\alpha_2 v_2+\cdots+\alpha_n v_n=\beta_1 v_1+\beta_2 v_2+\cdots+\beta_n v_n$

$\Rightarrow (\alpha_1-\beta_1)v_1+(\alpha_2-\beta_2)v_2+\cdots+(\alpha_n-\beta_n)v_n=0 \Rightarrow (\alpha_1-\beta_1)=(\alpha_2-\beta_2)=\cdots=(\alpha_n-\beta_n)=0 \Rightarrow \alpha_1=\beta_1,\cdots,\alpha_n=\beta_n \Rightarrow$ unique coefficients

ii> Uniqueness ⟶ Linearly independent

Assume $v_1,v_2,\cdots v_n$ are not linearly independent, there exist $c_1,c_2\cdots,c_n$ with $c_k\neq 0$, s.t. $c_1v_1+\cdots+c_nv_n=0$

$\Rightarrow u+0=u=(\beta_1+c_1)v_1+(\beta_2+c_2)v_2+\cdots+(\beta_n+c_n)v_n$

Since $\beta_k\neq\beta_k+c_k \Rightarrow \beta_1,\beta_2\cdots\beta_n$ are not unique $\Rightarrow$ Contradiction! $\Rightarrow$ Linearly independent

② Definition

A minimal spanning set of V

Conditions
- i> $B\subseteq V$ — 向量维数要求
- ii> $v_1, v_2, v_3\cdots v_n$ are linearly independent — 当向量个数>向量维数时，一定线性相关
- iii> For any $v\in V$, $v\in \text{Span}\{B\}$ — 向量个数要求

③ Standard basis

$\mathcal{E}=\{e_1,e_2,\cdots,e_n\}$ is the standard basis of R^n.

$$\mathbf{e}_i=\begin{array}{c}\begin{bmatrix}0\\ \vdots\\ 0\\ 1\\ 0\\ \vdots\\ 0\end{bmatrix}\end{array}\begin{array}{l}\\ \\ \\ \text{The } i\text{-th entry}\\ \\ \\ \\ \end{array}$$

base vector

2. Coordinates

① Definition:

X & $[X]_B$ are both vectors

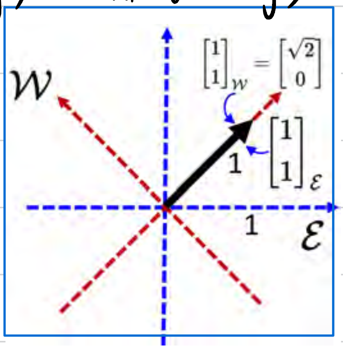
$x = B[x]_B$!!!

if $B = \{v_1, v_2, \cdots v_n\}$, $x = c_1v_1 + c_2v_2 + \cdots + c_nv_n$, we call $[x]_B = (c_1, c_2, \cdots, c_n)^T$ the coordinate vector of x w.r.t. B

(with respect to)

For $Ax = b$, $A[b]_B = b$: $[a_1\ a_2 \cdots a_n]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1a_1 + x_2a_2 + \cdots + x_na_n = b$

② Basis changes ⇒ Coordinate changes



$\Longrightarrow$

$$\boxed{\begin{bmatrix}1\\1\end{bmatrix}_{\mathcal{W}} = \begin{bmatrix}x\\y\end{bmatrix}}$$

the absence of subscript ⇒ standard basis

Basis changes　　　　Coordinate changes

3. Dimension

① Definition

For a vector space V, its dimension is the # of vectors in the basis

and
- Dimension of the subspace $\{0\}$ is 0 　e.g. 　if zero
- Finite dimensional: the # of vectors is finite 　e.g.
- Infinite dimensional: the # of vectors is infinite 　e.g. the vector space of all polynomials / $C[a,b]$

❑ A non-zero vector in $\mathbb{R}^3$ spans a 1-dimensional subspace.
❑ Two linearly independent vectors in $\mathbb{R}^3$ span a 2-dimensional subspace.
❑ Any three linearly independent vectors in $\mathbb{R}^3$ span a 3-dimensional subspace, which must be $\mathbb{R}^3$.

② Theorem

空间维数=基内向量个数≤基内向量维数(即$R^{\#}$)

$$\mathcal{B} = \left\{\begin{bmatrix}5\\-1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\2\end{bmatrix}\right\}$$

向量个数

向量维数

proof:

当基内向量个数>基内向量维数 ⟶ prove: 在基向量个数为n的情况下继续加向量, 则它们线性相关

$\{v_1, v_2, \cdots, v_n\}$ is a spanning set of V, Let $u_1, u_2, \cdots, u_m$ be m vectors in V, where $m > n$,

then $u_i = a_{i1}v_1 + a_{i2}v_2 + \cdots + a_{in}v_n$ for $i = 1, 2, \cdots, m$

then $c_1u_1 + c_2u_2 + \cdots + c_mu_m = c_1\sum_{j=1}^{n} a_{1j}v_j + c_2\sum_{j=1}^{n} a_{2j}v_j + \cdots + c_m\sum_{j=1}^{n} a_{mj}v_j = \sum_{i=1}^{m}\left[c_i\left(\sum_{j=1}^{n} a_{ij}v_j\right)\right] = \sum_{j=1}^{n}\left(\sum_{i=1}^{m} a_{ij}c_i\right)v_j$

$\sum_{i=1}^{m} a_{ij}c_i = 0$, for $j = 1, 2, \cdots, n$, this system has a non-trivial solution $(\hat{c}_1, \hat{c}_2, \cdots, \hat{c}_m)^T$ that $\hat{c}_1u_1 + \hat{c}_2u_2 + \cdots + \hat{c}_mu_m = \sum_{j=1}^{n} 0v_j = 0 \Rightarrow u_1, u_2, \cdots, u_m$ are linearly dependent

(*)

Linear system (*) has more unknowns than equations

4. Determine coordinates from different bases of the same vector space

① Lemma:

Vector space V has a basis $\mathcal{B}$ and let $x, y \in V$, then for any scalars α, β, $[\alpha x+\beta y]_{\mathcal{B}} = \alpha[x]_{\mathcal{B}} + \beta[y]_{\mathcal{B}}$

the coordinate of linear combination = the linear combination of coordinate

proof:

Let $\mathbf{x}, \mathbf{y} \in \mathcal{V}$

$$\mathbf{x} = \sum_{i=1}^{n} c_i \mathbf{b}_i \Rightarrow [\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \sum_{i=1}^{n} d_i \mathbf{b}_i \Rightarrow [\mathbf{y}]_{\mathcal{B}} = \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix}$$

$$\boxed{\alpha\mathbf{x} + \beta\mathbf{y}} = \alpha \sum_{i=1}^{n} c_i \mathbf{b}_i + \beta \sum_{i=1}^{n} d_i \mathbf{b}_i = \sum_{i=1}^{n} \alpha c_i \mathbf{b}_i + \sum_{i=1}^{n} \beta d_i \mathbf{b}_i = \sum_{i=1}^{n} (\alpha c_i + \beta d_i)\mathbf{b}_i$$

Extract the coordinates w.r.t to $\mathcal{B}$

$$\Rightarrow [\alpha\mathbf{x} + \beta\mathbf{y}]_{\mathcal{B}} = \begin{bmatrix} \alpha c_1 + \beta d_1 \\ \alpha c_2 + \beta d_2 \\ \cdots \\ \alpha c_n + \beta d_n \end{bmatrix} = \alpha\boxed{[\mathbf{x}]_{\mathcal{B}}} + \beta\boxed{[\mathbf{y}]_{\mathcal{B}}}$$

② Theorem:

$W = \{w_1, w_2, \cdots, w_n\}$ and $\mathcal{B} = \{v_1, v_2, \cdots, v_n\}$

then $[x]_{\mathcal{B}} = U \cdot [x]_W \longrightarrow [x]_W = U^{-1}[x]_{\mathcal{B}}$, where $u_j = [w_j]_{\mathcal{B}}$ (Key: compute $U \mid U^{-1}$)

Transision matrix

"内前外后"

后坐 = U · 前坐

proof:

Let $[x]_W = (d_1, d_2, \cdots, d_n)^T$, then we have $x = d_1w_1 + d_2w_2 + \cdots + d_nw_n$

Denote $u_j = [w_j]_{\mathcal{B}}$

$$[x]_{\mathcal{B}} = [d_1w_1 + d_2w_2 + \cdots + d_nw_n]_{\mathcal{B}} \overset{\text{Lemma}}{=} d_1[w_1]_{\mathcal{B}} + d_2[w_2]_{\mathcal{B}} + \cdots + d_n[w_n]_{\mathcal{B}} = d_1u_1 + d_2u_2 + \cdots + d_nu_n$$

$$= \underbrace{[u_1 \ u_2 \ \cdots \ u_n]}_{U} \underbrace{\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}}_{[x]_W} \Rightarrow [x]_{\mathcal{B}} = U \cdot [x]_W$$

If U is invertible, then $[x]_W = U^{-1}[x]_{\mathcal{B}}$, we next prove that it's invertible by proving vectors in U are linearly independent

We assume there exist non-all-zero $c_1, \cdots, c_n$ satisfying $\sum_{j=1}^{n} c_j u_j = 0$

Recall $u_j = [w_j]_{\mathcal{B}}$, we have $w_j = \sum_i u_{ij} v_i$

$\Rightarrow \sum_{j=1}^{n} c_j w_j = \sum_j c_j \sum_i u_{ij} v_i = \sum_i \left(\sum_j c_j u_{ij}\right) v_i = \sum_i 0 v_i = 0$

$\Rightarrow \{w_1, w_2, \cdots w_n\}$ are linealy dependent $\Rightarrow$ Contradiction! $\Rightarrow U$ is invertible

③ Corollary

$V=[v_1, v_2, \cdots, v_n]$, $W=[w_1, w_2, \cdots, w_n]$, then $[x]_V = \underline{U}\,[x]_W$, where $u_j=[w_j]_V$

$$= \underline{V^{-1}W}\,[x]_W$$

$\text{后}^{-1}\cdot\text{前}$

For a basis $V=\{v_1, v_2, \cdots, v_n\}$, $V=[v_1\ v_2 \cdots v_n]$

proof:

$x=[x]_E=V\cdot[x]_V=W[x]_W$

Slide 6

1. Null space

① Definition

the solution set of $Ax=0$, i.e. $Null(A) = \{x \in \mathbb{R}^n : Ax = 0\}$, $Null(A)$ is a subspace of $\mathbb{R}^n$

Null space is a kind of subspace, it is spanned by the elements in the solution set

② Example

$$\mathbf{Ax=0} \Rightarrow \begin{cases} x_1 = -2x_3 - x_5 \\ x_2 = 3x_3 - 4x_5 \\ x_3 = x_3 \\ x_4 = -2x_5 \\ x_5 = x_5 \end{cases} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = x_3 \begin{bmatrix} -2 \\ 3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -1 \\ -4 \\ 0 \\ -2 \\ 1 \end{bmatrix} \Rightarrow \boxed{\mathbf{Null(A)} = \text{Span}\left\{ \begin{bmatrix} -2 \\ 3 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ 0 \\ -2 \\ 1 \end{bmatrix} \right\}}$$

For homogenenous linear system,
the sol set = span = $Null(A)$

③ Theorems

a. The dimension of $Null(A)$ is equal to the # of free variables of $Ax=0$

From the example above, the dimension = 2

b. Square matrix A is invertible $\rightleftharpoons$ $Null(A)$ is trivial, ie.: $Null(A) = \{0\}$

$Null(A) = \{0\}$ means there is no free variables

c. If A and B are row equivalent, then $Null(A) = Null(B)$

$Ax=0$ & $Bx=0$ will have the same sol, then the same Null space.

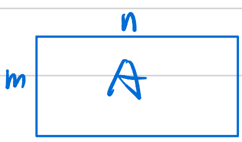
2. Row and column spaces

① Definition

$\begin{cases} \text{Row space} & Row(A) \text{ is the subspace spanned by the row vectors of } A \\ \text{Column space} & Col(A) \text{ is the subspace spanned by the col vectors of } A \end{cases}$, for $A_{m \times n}$

* Remarks

i) $\begin{cases} Row(A) \text{ is a subspace of } \mathbb{R}^n \\ Col(A) \text{ is a subspace of } \mathbb{R}^m \end{cases}$



ii) $Row(A) = Col(A^T)$

The subspace spanned by row vectors of A = col vectors of A^T

② Theorems

a. A and B are row equivalent, then $Row(A) = Row(B)$, $Col(A) \neq Col(B)$

Row equivalent $\longrightarrow$ do row operations to make their rows the same $\longrightarrow$ the spanning set is the same $\longrightarrow$ the subspace is the same

b. A has the RREF U. Let W be the set of cols of A that correspond to the pivot cols of B.

Then, W is a basis of the column space of A. Row(A) = Row(U)

Example: From P. 4, we have

Pivots on the 1st, 2nd and 4th Columns

$$\mathbf{A}\ \begin{bmatrix} \boxed{2} & \boxed{-1} & 7 & \boxed{-3} & -8 \\ \boxed{1} & \boxed{0} & 2 & \boxed{4} & 9 \\ \boxed{2} & \boxed{2} & -2 & \boxed{-1} & 8 \end{bmatrix} \xrightarrow{\text{reduced row-echelon}} \begin{bmatrix} \boxed{1} & 0 & 2 & 0 & 1 \\ 0 & \boxed{1} & -3 & 0 & 4 \\ 0 & 0 & 0 & \boxed{1} & 2 \end{bmatrix}\ \mathbf{B}$$

$$\Rightarrow \left\{ \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ -1 \end{bmatrix} \right\} \text{ is a basis of } \mathrm{Col}(\mathbf{A})$$

Brief proof:

Cols* are linearly independent in U, "A → U" will not change the independence of Cols*, so in A, Cols* form a basis

i) Find the pivotal row / pivotal col of U

ii) Match back with A

iii) The corresponding set of rows / cols is the basis of Row(A) / Col(A)

Row(A) = Span{Row vectors} = Span basis

☆ Find row(A) / col(A):

M1: Find the basis → space = Span basis

M2: (For Col(A)), Col(A) ≜ the set of b that makes the linear system consistent

c. Alternative approach to find col. space:

A linear system Ax=b is consistent iff b is in the col space of A.

$$\left[\begin{array}{ccccc|c} 1 & 2 & 7 & 1 & -1 & b_1 \\ 1 & 1 & 3 & 1 & 0 & b_2 \\ 3 & 2 & 5 & -1 & 9 & b_3 \\ 1 & -1 & -5 & 2 & 0 & b_4 \end{array}\right] \xrightarrow{\text{reduced row-echelon}} \left[\begin{array}{ccccc|c} 1 & 0 & -1 & 0 & 3 & -3b_1+5b_2-b_4 \\ 0 & 1 & 4 & 0 & -1 & b_1-b_2 \\ 0 & 0 & 0 & 1 & -2 & 2b_1-3b_2+b_4 \\ 0 & 0 & 0 & 0 & 0 & \boxed{9b_1-16b_2+b_3+4b_4} \end{array}\right]$$

The system $\mathbf{Ax} = \mathbf{b}$ is consistent if and only: $\boxed{9b_1 - 16b_2 + b_3 + 4b_4} = 0$

Therefore

$$\mathrm{Col}(\mathbf{A}) = \{\mathbf{b} : 9b_1 - 16b_2 + b_3 + 4b_4 = 0\} = \mathrm{Null}\left\{\begin{bmatrix} 9 & -16 & 1 & 4 \end{bmatrix}\right\}$$

Recall $\begin{bmatrix} b_1 & b_2 & b_3 & b_4 \end{bmatrix} \begin{bmatrix} 9 \\ -16 \\ 1 \\ 4 \end{bmatrix} = 0$

i) Find the condition to make linear system consistent

ii) Denote Col(A) by Null(B)

d. A & A^T have the same # of pivots

3. Rank

① Lemma

For an $A_{m\times n}$, dim(Row(A)) = dim(Col(A))

the same # of pivots → the same # of vectors in bases of row space & col space → the same dimension

② Definition

The rank of a matrix A, denoted by $rank(A)$, is the dimension of row space of A

$rank(A) = dim(Row(A)) = dim(Col(A)) = \#$ of pivots

③ Rank-Nullity Theorem

If A is an $m \times n$ matrix, then $\underbrace{rank(A)}_{\#\text{ of pivots}} + \underbrace{dim(Null(A))}_{\#\text{ of free variables}} = n$

④ Properties

a. $rank(A) = rank(A^T) = rank(A^TA) = rank(AA^T)$

b. $\begin{cases} rank(A) \le m \\ rank(A) \le n \end{cases} \longrightarrow rank(A) \le min(m,n)$

the # of pivots $\begin{cases} \le m \\ \le n \end{cases}$ (each row / each col at most has one pivot)

Definition: A $m \times n$ matrix is called full rank if $rank(A) = min(m,n)$

4. Determinant

① Two basic properties

a. For $n \times n$ matrices A & B, $det(AB) = det(A)det(B)$

b. For $n \times n$ triangular matrix A, $det(A) = a_{11}a_{22}\cdots a_{nn}$

From a & b, we know only the square matrix has determinant

② Characteristics

a. $det(A) = 0$, if two rows / cols of A are the same

Two rows / cols are the same $\longrightarrow$ singular $\longrightarrow$ $det(A) = 0$

b. $det(A) = det(A^T)$

c. $det(A^{-1}) = \dfrac{1}{det(A)}$

proof of b

Suppose $A = PLU$, as $A^T = U^TL^TP^T$,

$det(A^T) = det(U^T)det(L^T)det(P^T) = det(U)det(L)det(P^T)$

Since $P^T = P^{-1}$, $det(P^T) = det(P^{-1}) = \dfrac{1}{det(P)}$

Since $\begin{cases} det(P) = +1 / -1 \\ det(P^{-1}) = +1 / -1 \\ det(P) \text{ \& } det(P^{-1}) \text{ 同号} \end{cases}$, then $det(P^T) = det(P^{-1}) = det(P)$

$\Rightarrow det(A^T) = det(U)det(L)det(P^T) = det(U)det(L)det(P) = det(A)$, QED.

③ Determinants of elementary matrices

❑ Since elementary matrices are diagonal, we have

Row interchanges: $\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix} \Rightarrow \begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}$ (a) Type I, $\det(\mathbf{E}_1) = -1$

Multiply $\alpha \neq 0$ to a row: $\begin{bmatrix}1&0&0\\0&1&0\\0&0&\alpha\end{bmatrix}$ (b) Type II, $\det(\mathbf{E}_2) = \alpha$

Adding $\alpha \neq 0$ times of one row to another row: $\begin{bmatrix}1&0&\alpha\\0&1&0\\0&0&1\end{bmatrix}$ (c) Type III, $\det(\mathbf{E}_3) = 1$

a. Derivation of det of type 1 elementary matrices

Type 1 elementary matrix can be derived by Type 2 & 3 elementary matrices

❑ We have $\mathbf{E} = \mathbf{E}_4\mathbf{E}_3\mathbf{E}_2\mathbf{E}_1$ where

- $\mathbf{E}_1$ is formed by adding row i to row j.
- $\mathbf{E}_2$ is formed by adding -1 times row j to row i.
- $\mathbf{E}_3$ is formed by adding row i to row j.
- $\mathbf{E}_4$ is formed by multiplying -1 to row i.

Reference the example below

❑ $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$ are type III, while $\mathbf{E}_4$ is type II.

❑ By the two basic properties, $\det(\mathbf{E}) = -1$!

$$\underset{i=1,j=2}{\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}} \xrightarrow[\text{Type III}]{\mathbf{E}_1} \begin{bmatrix}1&0&0\\1&1&0\\0&0&1\end{bmatrix} \xrightarrow[\text{Type III}]{\mathbf{E}_2} \begin{bmatrix}0&-1&0\\1&1&0\\0&0&1\end{bmatrix} \xrightarrow[\text{Type III}]{\mathbf{E}_3} \begin{bmatrix}0&-1&0\\1&0&0\\0&0&1\end{bmatrix} \xrightarrow[\text{Type II } \alpha=-1]{\mathbf{E}_4} \begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}$$

b. Properties

① Type 1 $A \longrightarrow B$: interchanges two rows k times $\longrightarrow \det(B) = (-1)^k \det(A)$

② Type 2 $A \longrightarrow B$: $(\times k)$ for i^{th} row & $(\times l)$ for z^{th} row $\longrightarrow \det(B) = kl \det(A)$

③ Type 3 $A \longrightarrow B$: $(\times k)$ for i^{th} row + z^{th} row $\longrightarrow \det(B) = \det(A)$

④ $\det([a+b, c_2, \cdots c_n]) = \det([a, c_2, \cdots, c_n]) + \det([b, c_2, \cdots, c_n])$

Example: $\det\begin{bmatrix}1+1 & 1\\ -1+2 & 1\end{bmatrix} = \det\begin{bmatrix}1&1\\-1&1\end{bmatrix} + \det\begin{bmatrix}1&1\\2&1\end{bmatrix}$

Proof:	$\mathbf{a} \in \text{Span}\{\mathbf{c}_2, \cdots, \mathbf{c}_n\}$	$\mathbf{a} \notin \text{Span}\{\mathbf{c}_2, \cdots, \mathbf{c}_n\}$
$\mathbf{b} \in \text{Span}\{\mathbf{c}_2, \cdots, \mathbf{c}_n\}$	All matrices are singular	Case 2
$\mathbf{b} \notin \text{Span}\{\mathbf{c}_2, \cdots, \mathbf{c}_n\}$	Case 1	Case 3

Case 1 $\boxed{\mathbf{a} \in \text{Span}\{\mathbf{c}_2, \cdots, \mathbf{c}_n\}} \Rightarrow \det([\mathbf{a}, \mathbf{c}_2, \ldots \mathbf{c}_n]) = \boxed{0}$

Same for Case 2

$$\text{L.H.S} = \det([\boxed{\mathbf{a}} + \mathbf{b}, \mathbf{c}_2, \ldots \mathbf{c}_n]) = \det\left(\left[\boxed{\sum_{i=2}^{n} x_i\mathbf{c}_i} + \mathbf{b}, \mathbf{c}_2, \ldots \mathbf{c}_n\right]\right)$$

$$\overset{\text{Type 3}}{=} \det(\mathbf{E})\det([\mathbf{b}, \mathbf{c}_2, \ldots \mathbf{c}_n]) = \det([\mathbf{b}, \mathbf{c}_2, \ldots \mathbf{c}_n]) + \boxed{0} = \text{R.H.S}$$

The dimension of $\text{Span}\{\mathbf{c}_2, \ldots \mathbf{c}_n\}$ is $n-1$

$\text{Span}\{\mathbf{c}_2, \ldots, \mathbf{c}_n\} \subset \mathbb{R}^n$ whose dimension is n

$\Big\} \; \mathbf{a} \in \text{Span}\{\mathbf{b}, \mathbf{c}_2, \cdots, \mathbf{c}_n\}$ (A basis of $\mathbb{R}^n$)

Case 3

$$\text{L.H.S} = \det([\boxed{\mathbf{a}} + \mathbf{b}, \mathbf{c}_2, \ldots \mathbf{c}_n]) = \det\left(\left[\boxed{\left(\underbrace{\sum_{i=2}^{n} x_i\mathbf{c}_i}_{\text{III}} + \overset{\text{II}}{\boxed{\alpha}}\mathbf{b}\right)} + \mathbf{b}, \mathbf{c}_2, \ldots \mathbf{c}_n\right]\right)$$

$$= \det([(\alpha+1)\mathbf{b}, \mathbf{c}_2, \ldots \mathbf{c}_n]) = (\alpha+1)\det([\mathbf{b}, \mathbf{c}_2, \ldots \mathbf{c}_n])$$

$$\det([\boxed{\mathbf{a}}\, \mathbf{c}_2, \ldots \mathbf{c}_n]) = \det\left(\left[\boxed{\sum_{i=2}^{n} x_i\mathbf{c}_i + \alpha\mathbf{b}}, \mathbf{c}_2, \ldots \mathbf{c}_n\right]\right) = \boxed{\alpha\det([\mathbf{b}, \mathbf{c}_2, \ldots \mathbf{c}_n])}$$

$$\text{R.H.S} = \boxed{\det([\mathbf{a}, \mathbf{c}_2, \ldots \mathbf{c}_n])} + \det([\mathbf{b}, \mathbf{c}_2, \ldots \mathbf{c}_n]) = (\alpha+1)\det([\mathbf{b}, \mathbf{c}_2, \ldots \mathbf{c}_n])$$

④ Determinants for LU decomposition

a. For LU decomposition

$$\det(A) = \det(L)\det(U) \xrightarrow{\text{if } L \text{ is unit triangular}} \det(U)$$

b. For PLU decomposition

$$\det(A) = \det(P)\det(L)\det(U)$$

Slide 7

1. Cofactor

① Concepts

- Submatrix : M_{ij} (去除 i^{th} row, j^{th} col)
- Minor : $\det(M_{ij})$
- Cofactor : $C_{ij} = (-1)^{i+j} \det(M_{ij})$

② Cofactor expansion (Laplace Theorem)

a. Lemma: If $A = \begin{bmatrix} a_{11} & \cdots \\ 0 & M_{11} \end{bmatrix}$, then $\det(A) = a_{11}\det(M_{11})$

两者换回I的次数相同

$\begin{bmatrix} 1 & 0 \\ 0 & P \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & E_k \end{bmatrix} \cdots \begin{bmatrix} 1 & 0 \\ 0 & E_1 \end{bmatrix} = (-1)^k = \det(P)$

换回去I

proof:

$M_{11} = PLU$, since A is a diagonal partitioned matrix, then $A = \begin{bmatrix} a_{11} & \cdots \\ 0 & M_{11} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & P \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & L \end{bmatrix} \begin{bmatrix} a_{11} & \cdots \\ 0 & U \end{bmatrix}$

then $\det(A) = a_{11}\det(P)\det(L)\det(U) = a_{11}\det(M_{11})$

b. Computation of determinant: Cofactor expansion (Nature: reduce order)

$$\det(A) = \begin{cases} \text{按行展开: } a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} \\ \text{按列展开: } a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} \end{cases}$$

> Compute determinant:
> i> Two basis properties ii> Definition (Leibniz formula)
> iii> Cofactor expansion

proof:

$\mathbf{a}_1 = \boxed{a_{11}\mathbf{e}_1 + a_{21}\mathbf{e}_2 + \cdots + a_{n1}\mathbf{e}_n}$

$$\begin{aligned}
\det(\mathbf{A}) &= \det([\boxed{a_{11}\mathbf{e}_1 + \cdots + a_{n1}\mathbf{e}_n} \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n]) \\
&= \det([a_{11}\mathbf{e}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n]) + \cdots + \det([a_{n1}\mathbf{e}_n \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n]) \\
&= a_{11}\underbrace{\det([\mathbf{e}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n])}_{\overset{?}{=}\,(-1)^{1+1}\det(\mathbf{M}_{11})} + \cdots + a_{n1}\underbrace{\det([\mathbf{e}_n \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n])}_{\overset{?}{=}\,(-1)^{n+1}\det(\mathbf{M}_{n1})}
\end{aligned}$$

把单列分离，从而构造出Lemma中的形式

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} = \sum_{i=1}^{4} a_{i1}\det[\mathbf{e}_i, \mathbf{A}_{2:4}] = \sum_{i=1}^{4} a_{i1}(-1)^{i+1}\det(\mathbf{M}_{i1})$$

Using $i = 4$ as an example

$$\det\begin{bmatrix} 0 & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} = (-1)^{1}\det\begin{bmatrix} a_{41} & a_{42} & a_{43} & a_{44} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & a_{32} & a_{33} & a_{34} \\ 0 & a_{12} & a_{13} & a_{14} \end{bmatrix}$$

Need 1 row interchange to move a_{41} to the top-left corner.

$$= (-1)\cdot a_{41}\cdot\det\begin{bmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{12} & a_{13} & a_{14} \end{bmatrix} \qquad \mathbf{M}_{41} = \begin{bmatrix} a_{12} & a_{13} & a_{14} \\ a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \end{bmatrix}$$

Need 2 row interchanges to become $\mathbf{M}_{41}$

$$= (-1)^{1+2}\cdot a_{41}\cdot\det(\mathbf{M}_{41})$$

In total, we perform 3 interchanges, therefore, we have the factor of $(-1)^3$

先构造 a_{yx}，后构造 M_{xx}

③ Adjoint matrix

a. Premise: Square

Square ⟶ Det ⟶ Cofactor ⟶ Adjoint matrix

b. Definition:

$$\operatorname{adj}(\mathbf{A}) = \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & & & \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}$$

☆: Transpose!!!

防止出错：先写出完整的adjoint matrix

c. Compute inverse by adjoint matrix

$$\mathrm{adj}(\mathbf{A})\mathbf{A}=\begin{bmatrix}C_{11}&C_{21}&\cdots&C_{n1}\\C_{12}&C_{22}&\cdots&C_{n2}\\\vdots&&&\\C_{1n}&C_{2n}&\cdots&C_{nn}\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&&&\\a_{n1}&a_{2n}&\cdots&a_{nn}\end{bmatrix}=\begin{bmatrix}\det(\mathbf{A})&0&\cdots&0\\0&\det(\mathbf{A})&\cdots&0\\\vdots&&&\\0&0&\cdots&\det(\mathbf{A})\end{bmatrix}=\det(\mathbf{A})\mathbf{I}$$

Why "Transpose"?
To satisfy the order to make cofactor expansion

The last equality is derived from (See next page)
$$\sum_{k=1}^{n}a_{kj}C_{ki}=\begin{cases}\det(\mathbf{A})&\text{if } i=j\\0&\text{if } i\neq j\end{cases}$$

$$\mathrm{adj}(\mathbf{A})\mathbf{A}=\det(\mathbf{A})\mathbf{I}\Rightarrow\boxed{\mathbf{A}^{-1}=\frac{1}{\det(\mathbf{A})}\mathrm{adj}(\mathbf{A})}$$

The (i,j) entry of $\mathrm{adj}(\mathbf{A})\mathbf{A}$ is

$$\boxed{(\mathbf{A}^{-1})_{ij}=\frac{C_{ji}}{\det(\mathbf{A})}=\frac{(-1)^{i+j}\det(\mathbf{M}_{ji})}{\det(\mathbf{A})}}$$

注意与adj(A)的对应位置

proof:

相当于把cofactor expansion中的j换成i

$$\sum_{k=1}^{n}a_{ki}C_{kj}=\begin{cases}\det(\mathbf{A})&\text{if } i=j\\0&\text{if } i\neq j\end{cases}$$
$$=\det\underbrace{[\mathbf{a}_1\ \cdots\ \mathbf{a}_{j-1}\ \mathbf{a}_i\ \mathbf{a}_{j+1}\ \cdots\ \mathbf{a}_n]}_{\mathbf{B}}$$

$i=j$: $\mathbf{A}=\mathbf{B}$ ⇒ $\det(\mathbf{A})$

$i\neq j$: Two columns are identical ⇒ $\det(\mathbf{B})=0$

Example $n=3, i=1, j=2$ $\mathbf{A}=\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}$

L.H.S=

$$a_{11}C_{12}+a_{21}C_{22}+a_{31}C_{32}=-a_{11}\det\begin{bmatrix}a_{21}&a_{23}\\a_{31}&a_{33}\end{bmatrix}+a_{21}\det\begin{bmatrix}a_{11}&a_{13}\\a_{31}&a_{33}\end{bmatrix}-a_{31}\det\begin{bmatrix}a_{11}&a_{13}\\a_{21}&a_{23}\end{bmatrix}$$

R.H.S=

$$\det\begin{bmatrix}a_{11}&a_{11}&a_{13}\\a_{21}&a_{21}&a_{23}\\a_{31}&a_{31}&a_{33}\end{bmatrix}=-a_{11}\det\begin{bmatrix}a_{21}&a_{23}\\a_{31}&a_{33}\end{bmatrix}+a_{21}\det\begin{bmatrix}a_{11}&a_{13}\\a_{31}&a_{33}\end{bmatrix}-a_{31}\det\begin{bmatrix}a_{11}&a_{13}\\a_{21}&a_{23}\end{bmatrix}$$

d. Compute variables by adjoint matrix: Cramer's Rule

⚠ Key: $x=A^{-1}b=\frac{1}{\det(A)}\mathrm{adj}(A)b$

$\longrightarrow x_i=\frac{1}{\det(A)}(b_1A_{1i}+\cdots+b_nA_{ni})=\frac{1}{\det(A)}\det(A_i)$, where $A_i=[a_1\cdots a_{i-1}\ b\ a_{i+1}\cdots a_n]$

Example
$$\underbrace{\begin{bmatrix}1&1\\-1&1\end{bmatrix}}_{\mathbf{A}}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}3\\1\end{bmatrix}\Rightarrow\begin{bmatrix}x_1\\x_2\end{bmatrix}=\underbrace{\frac{1}{2}\begin{bmatrix}1&-1\\1&1\end{bmatrix}}_{\mathbf{A}^{-1}}\begin{bmatrix}3\\1\end{bmatrix}=\begin{bmatrix}1\\2\end{bmatrix}$$
(Recall from P.9)

Using the Cramer's rule

$$x_1=\frac{1}{2}\det\begin{bmatrix}3&1\\1&1\end{bmatrix}=\frac{3-1}{2}=1\qquad x_2=\frac{1}{2}\det\begin{bmatrix}1&3\\-1&1\end{bmatrix}=\frac{3+1}{2}=2$$

⚠ Explanation

$$\underbrace{\begin{bmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\end{bmatrix}}_{\mathbf{A}}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}b_1\\b_2\end{bmatrix}$$

$$\begin{bmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\end{bmatrix}\begin{bmatrix}x_1&0\\x_2&1\end{bmatrix}=\begin{bmatrix}b_1&a_{12}\\b_2&a_{22}\end{bmatrix}$$

$$\det(\mathbf{A})\cdot\det\begin{bmatrix}x_1&0\\x_2&1\end{bmatrix}=\det\begin{bmatrix}b_1&a_{12}\\b_2&a_{22}\end{bmatrix}\qquad x_1=\frac{\det\begin{bmatrix}b_1&a_{12}\\b_2&a_{22}\end{bmatrix}}{\det(\mathbf{A})}$$

$$\begin{bmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\end{bmatrix}\begin{bmatrix}1&x_1\\0&x_2\end{bmatrix}=\begin{bmatrix}a_{11}&b_1\\a_{21}&b_2\end{bmatrix}\qquad x_2=\frac{\det\begin{bmatrix}a_{11}&b_1\\a_{21}&b_2\end{bmatrix}}{\det(\mathbf{A})}$$

2. Leibniz formula (Definition of Det)

$$\det(A) = \sum (-1)^r a_{1k_1} a_{2k_2} \cdots a_{nk_n} = \sum (-1)^r a_{k_1 1} a_{k_2 2} \cdots a_{k_n n}$$

$\downarrow$ $\qquad\qquad$ $\rightarrow$

where r: $k_1, k_2, \cdots k_n$ 的逆序数

Remark: $k_1、k_2、\cdots k_n$ 要取遍所有可能，故展开后总项数为 $n!$

3. Linear transformation

① Definition: Linear Transformation

$L(\alpha v_1 + \beta v_2) = \alpha L(v_1) + \beta L(v_2)$

Use this property to determine whether a mapping is linear transformation

$$\Rightarrow \begin{cases} L(v_1+v_2) = L(v_1) + L(v_2) \\ L(\alpha v) = \alpha L(v) \end{cases}$$

A mapping from a vector space to another vector space, denoted by $L: V \to W$

If $L: V \to V$, then L is called a linear operator

② Independence after linear transformation

Let $L: V \longrightarrow W$

a. v_is are linearly dependent $\xrightarrow{✓}$ $L(v_i)$s are linearly dependent (a1)

independent $\xrightarrow{✗}$ independent (a2) 见③b.△

b. $L(v_i)$s are linearly independent $\xrightarrow{✓}$ v_is are linearly independent (b1)

dependent $\xrightarrow{✓}$ dependent (b2)

proof:

a1: Since v_is are linearly dependent, then $\exists c_i \neq 0$ s.t. $\sum c_i v_i = 0$

$\Rightarrow \sum c_i L(v_i) = L(\sum c_i v_i) = L(0) = 0 \Rightarrow \exists c_i \neq 0$ s.t. $\sum c_i L(v_i) = 0 \Rightarrow L(v_i)$s are linearly dependent

a2: [Counterexample]

$$\text{Let } v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \; L(v) = \begin{bmatrix} v_1+v_2 \\ v_1+v_2 \end{bmatrix} \Rightarrow \underbrace{\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}}_{\text{linearly independent}} \text{ but } \left\{ L\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right), L\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \right\} = \underbrace{\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}}_{\text{linearly dependent}}$$

b1: Let $\sum c_i v_i = 0$, then $L(\sum c_i v_i) = \sum c_i L(v_i) = 0 \Rightarrow c_i$s are all zero

$\Rightarrow v_i$s are linearly independent

b2: $\sim\sim = 0$, $\sim\sim = 0 \Rightarrow c_i$s are not all zero $\Rightarrow v_i$s are linearly dependent

都要先设 $\sum c_i v_i = 0$，再具体分析

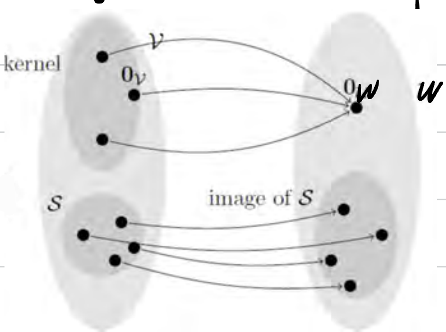
③ Kernel & Image

a. Definition

Kernel: $\ker(L) = \{v \in V \mid L(v) = 0_W\}$

Image: $L(S) = \{w \in W \mid w = L(v)$ for certain $v \in S\}$, $L(V)$ (V is the entire vector space) is called the range of L

$L(V) = \text{range}(L)$



b. Theorems

△ Kernel & Image are both subspaces

Let $L: V \to W$, $\begin{cases} \ker(L) \text{ is a subspace of } V \\ L(S) \text{ is a subspace of } W \end{cases}$

△ Let $L: V \to W$, for any $x, v \in V$, $x \in v + \ker(L)$, $L(x) = L(v)$

proof:

Let $x = v + w \to L(x) = L(v+w) = L(v) + L(w) = L(v)$

△ Let $L: V \to W$, for any linearly independent $v, v' \in V$, $L(v)$ and $L(v')$ are linearly independent iff $w \in \text{Span}\{v, v'\} \cap \ker(L) = \{0_V\}$

proof:

i.e. $(\alpha v + \beta v') = 0_V$

$\alpha L(v) + \beta L(v') = 0$ iff $\alpha = \beta = 0 \Rightarrow L(\alpha v + \beta v') \neq 0$ iff $\alpha = \beta = 0 \Rightarrow L(\alpha v + \beta v') = 0$ iff $w = \alpha v + \beta v'$, $w = 0 \longrightarrow w \in \text{Span}\{v, v'\} \cap \ker(L) = 0_V$

△ For each $\underbrace{L: R^n \to R^m}_{\text{vectors}}$, there exists an $m \times n$ matrix A s.t. $L(x) = Ax$ for each $x \in R^n$, where the jth col of A is $L(e_j)$

What's more: i) $\ker(L) = \text{Null}(A)$; ii) $\text{range}(L) = \text{Col}(A)$ → 连接了 properties of linear transformation 和 matrix representation 之间的关系。

Vec → Vec linear transformation ⇐⇒ pre-multiply a matrix

proof:

$x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n \longrightarrow L(x) = x_1 L(e_1) + x_2 L(e_2) + \cdots + x_n L(e_n) = [L(e_1)\ L(e_2) \cdots L(e_n)] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = Ax$

Slide 8

1. Matrix representation for vector spaces

From last slide, we know that $R^n \xrightarrow{A} R^m$, so we can represent the linear transformation $L: R^n \to R^m$ by pre-multiplying a matrix $A_{m\times n}$.

① Matrix representation for vector spaces

a. Process

△ Denote the vector space & the basis w.r.t. its vector space.

Vector Space	$\mathcal{V}$ $\xrightarrow{L:\mathcal{V}\to\mathcal{W}}$	$\mathcal{W}$
Basis	$\mathcal{B}=\{\mathbf{v}_1,\dots,\mathbf{v}_n\}$	$\mathcal{C}=\{\mathbf{w}_1,\dots,\mathbf{w}_m\}$

△ Let $a_j = \underbrace{[L(v_j)]_C}_{=C^{-1}L(v_j)}$, then we get A, $[L(u)]_C = A\cdot[u]_B$　　a_j的个数取决于v_j

实际上有两步运算：① 线性变换 ②求坐标

proof:

For any $u\in V$, we denote $x=[u]_B$, then $u = x_1v_1+\cdots+x_nv_n$

$L(u) = L(x_1v_1+\cdots+x_nv_n) = x_1L(v_1)+\cdots+x_nL(v_n)$

$$[L(u)]_C = x_1[L(v_1)]_C+\cdots+x_n[L(v_n)]_C = [\,[L(v_1)]_C \cdots [L(v_n)]_C\,]\cdot\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} = [a_1\ a_2\cdots a_n]\cdot x = A\cdot[u]_B$$

Example

❑ Let $L:\mathbb{R}^3\to\mathbb{R}^2$ be a linear transformation defined by

$L(\mathbf{x}) = x_1\mathbf{b}_1 + (x_2+x_3)\mathbf{b}_2$ where $\mathbf{b}_1=\begin{bmatrix}1\\1\end{bmatrix}$, $\mathbf{b}_2=\begin{bmatrix}-1\\1\end{bmatrix}$

Find the matrix $\mathbf{A}$ representing L with respect to the standard basis of $\mathbb{R}^3$ and the basis $\{\mathbf{b}_1,\mathbf{b}_2\}$ of $\mathbb{R}^2$.

Solution:

Step a.

Vector Space	$\mathcal{V}$	$\mathcal{W}$
Basis	$\mathcal{B}=\left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix},\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$	$\mathcal{C}=\left\{\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}-1\\1\end{bmatrix}\right\}$

Step b.

$\mathbf{a}_j = [L(\mathbf{v}_j)]_C$

$[L(\mathbf{x})]_C = [x_1\mathbf{b}_1+(x_2+x_3)\mathbf{b}_2]_C = \begin{bmatrix}x_1\\x_2+x_3\end{bmatrix}$

$[L(\mathbf{e}_1)]_C = \begin{bmatrix}1\\0\end{bmatrix}$　　$[L(\mathbf{e}_2)]_C = [L(\mathbf{e}_3)]_C = \begin{bmatrix}0\\1\end{bmatrix}$

$[L(\mathbf{u})]_C = \mathbf{A}\cdot[\mathbf{u}]_B$

$$\begin{bmatrix}x_1\\x_2+x_3\end{bmatrix} = \begin{bmatrix}1&0&0\\0&1&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$$

后 = A · 前

b. Rank-Nullity Theorem Revisited

Linear Transformation　　$[L(x)]_C = A\cdot[x]_B$　　Matrix representation

都是以V作为对象

$$\begin{pmatrix}\mathrm{Ker}(L)\ (=\mathrm{Ker}[L]_x)\\ [L(V)]_C\\ \dim(V)\end{pmatrix} = \begin{pmatrix}\mathrm{Null}(A)\\ \mathrm{Col}(A)\\ n\end{pmatrix}$$

$\dim(\mathrm{Ker}(L))+\dim([L(V)]_C)=\dim(V)$　　$\dim(\mathrm{Null}(A))+\dim(\mathrm{Col}(A))=n$

注意与前页的 $L(x)=Ax$ 有细微差别

② Change of basis

a. Comparison

We compare
- Matrix representation for vector spaces — Linear transformation: $V \longrightarrow W$
- Change of basis — Linear operator: $V \longrightarrow V$ — Essentially, it is a special case.

Recall

$$\begin{cases} a_j = [L(v_j)]_C \longrightarrow a_j = [v_j]_C \\ [L(u)]_C = A \cdot [u]_B \longrightarrow [u]_C = A[u]_B \end{cases}$$

b. Linear transformation wrt basis B + Transition matrix from basis C to basis B $\longrightarrow$ Linear transformation wrt basis C

从未知到已知的!!!

i.e. $\left.\begin{array}{l} [L(v)]_B = B[v]_B \\ [v]_B = S[v]_C \end{array}\right\} \longrightarrow C = S^{-1}BS$, where $[L(v)]_C = C[v]_C$

四个向量都同行 $\longrightarrow$ B、C、S 都是square

proof:

Since $[v]_B = S[v]_C$, then $[L(v)]_B = S[L(v)]_C$ 关键在于构造出$[L(v)]_C$

$[L(v)]_C = S^{-1}B[v]_B = S^{-1}BS[v]_C$

c. Similarity

Two $n \times n$ matrices (matrix representation) A & B are similar if there exists an invertible matrix S s.t. $B = S^{-1}AS$

Since for $\begin{cases} [L(x)]_A = A \cdot [x]_A \\ [L(x)]_B = B \cdot [x]_B \end{cases}$, the vectors are at the same size, so $\begin{cases} A \text{ \& } B \text{ are square} \\ \text{The dimension of } A \text{ \& } B \text{ is the same} \end{cases}$

③ Coordinate mapping

a. Definition

$[\cdot]_B : V \longrightarrow \mathbb{R}^n$, where $[\cdot]_B$ is the unique vector $x \in \mathbb{R}^n$ s.t. $v = x_1v_1 + \cdots + x_nv_n$

$v \in \mathbb{R}^n$ That is, $v \to [v]_B$ (a kind of linear operator)

b. Properties

△ Coordinate mapping is linear (i.e. it's a kind of linear transformation)

☆ What is "linear"?

"Linear" means first order functional relationship:
- Addition
- Scalar multiplication

- $Ker(L) = \{0_v\}$ 非常特殊，这意味着 $L(v)=0_w \longrightarrow v=0_v$!
- $Range(L) = R^n$
- Coordinate mapping has the identity matrix representation

Vector Space	$\mathbb{R}^n$ $\xrightarrow{L(\mathbf{u}) = [\mathbf{u}]_B}$	$\mathbb{R}^n$
Basis	$B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$	$\mathcal{C} = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$!

$$\mathbf{a}_j = [L(\mathbf{v}_j)]_{\mathcal{C}} = [\mathbf{v}_j]_B = \left[0 \times \sum_{i \neq j} \mathbf{v}_i + 1 \times \mathbf{v}_j\right]_B = \mathbf{e}_j$$

$$\boxed{[L(\mathbf{u})]_{\mathcal{C}}} = \boxed{\mathbf{A}} \cdot [\mathbf{u}]_B \Rightarrow \boxed{[\mathbf{u}]_B} = \boxed{[\mathbf{e}_1 \ \cdots \ \mathbf{e}_n]} \cdot [\mathbf{u}]_B = \boxed{\mathbf{I}} \cdot [\mathbf{u}]_B$$

前后都是坐标！(而且是一样的)

△ Coordinate mapping is one to one

c. Summary

$$\begin{tikzcd}
\mathcal{V} \arrow[r, "L", "\mathbf{v}_j \mapsto L(\mathbf{v}_j)"'] \arrow[d, dashed, "\text{basis } \mathcal{B}"', "{[\cdot]_{\mathcal{B}}}"] & \mathcal{W} \arrow[d, dashed, "\text{basis } \mathcal{C}"', "{[\cdot]_{\mathcal{C}}}"] \\
\mathbb{R}^n \arrow[r, "\mathbf{A} : \mathbf{a}_j = {[L(\mathbf{v}_j)]_{\mathcal{C}}}", "{[L(\mathbf{x})]_{\mathcal{C}} = \mathbf{A} \cdot [\mathbf{x}]_{\mathcal{B}}}"'] & \mathbb{R}^m
\end{tikzcd}$$

注意 matrix representation 的对象是坐标
↓
R^n

Slide 9

1. Inner product

① Inner product

a. Definition   inner product = scalar product

$\langle x,y\rangle = x^T y = x_1y_1+\cdots+x_ny_n$, $x,y\in\mathbb{R}^n$ the same size!

b. Properties

- $\langle x,y\rangle = \langle y,x\rangle$
- $\langle x,x\rangle \geq 0$, equal iff $x=0$
- $\langle ax+by,z\rangle = a\langle x,z\rangle + b\langle y,z\rangle$ $a\langle x,y\rangle = \langle ax,y\rangle = \langle x,ay\rangle$
- $\langle x+y,x+y\rangle = \langle x,x\rangle+\langle y,y\rangle+2\langle x,y\rangle$ (binomial expansion)

c. Relevant concepts & Relationships

△ Relevant concepts

- Length: $\|x\| = \sqrt{\langle x,x\rangle} = \sqrt{x^Tx} = \sqrt{x_1^2+x_2^2+\cdots+x_n^2}$
- Distance: $\|x-y\| = \sqrt{(x-y)^T(x-y)} = \sqrt{(x_1-y_1)^2+\cdots+(x_n-y_n)^2}$

△ Relationships

i) Cauchy-Schwarz Inequality

$|\langle x,y\rangle| \leq \|x\|\|y\|$, equal iff x & y are linearly dependent   连结了内积与模长

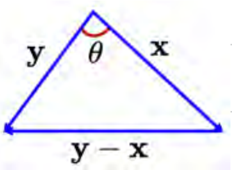
proof: (i) 构造unit vector ; ii) 构造 $\langle x-y,x-y\rangle$)

Let $\|\tilde{x}\| = \|\tilde{y}\| = 1$

$\|\tilde{x}-\tilde{y}\|^2 \geq 0 \longrightarrow \langle \tilde{x}-\tilde{y},\tilde{x}-\tilde{y}\rangle \geq 0 \longrightarrow \|\tilde{x}\|^2+\|\tilde{y}\|^2-2\langle\tilde{x},\tilde{y}\rangle \geq 0 \longrightarrow 2\langle\tilde{x},\tilde{y}\rangle \leq \|\tilde{x}\|^2+\|\tilde{y}\|^2 \longrightarrow \langle\tilde{x},\tilde{y}\rangle \leq 1$

$\longrightarrow \langle \underbrace{\frac{x}{\|x\|}}_{\text{unit vector}}, \frac{y}{\|y\|}\rangle \leq 1 \longrightarrow \langle x,y\rangle \leq \|x\|\|y\|$ (equal when $\frac{x}{\|x\|} = \frac{y}{\|y\|} \longrightarrow x = \alpha y \longrightarrow x$ & y are linearly dependent)

ii) Angle between two vectors



$\|y-x\|^2 = \begin{cases} \Delta: & \|x\|^2+\|y\|^2-2\cos\theta\|x\|\|y\| \\ \langle y-x,y-x\rangle: & \|x\|^2+\|y\|^2-2\langle x,y\rangle \end{cases} \Rightarrow \langle x,y\rangle = \|x\|\|y\|\cos\theta \ / \ \cos\theta = \frac{\langle x,y\rangle}{\|x\|\|y\|}$

When $\cos\theta=0$, $\langle x,y\rangle = x^Ty = 0$, denoted by $x\perp y$, then we say x & y are orthogonal

iii) Pythagorean Law

if $x\perp y$, then $\|x+y\|^2 = \|x\|^2+\|y\|^2$

$\uparrow$

$\langle x,y\rangle = 0$ ($2\langle x,y\rangle = 0$)

② Inner product space

a. More general definition of inner product

Vec-Vec multiplication $\longrightarrow$ A mapping

△ Definition

Inner product is a mapping satisfying the following 3 conditions: (对应着 inner product 的四个性质)

- △ Positive-definite: $\langle x,x\rangle \geq 0$ with equality iff $x=0$ 只须验证 $x \perp y$，因为倘若其为 inner product，它会满足对称性
- △ Symmetry: $\langle x,y\rangle = \langle y,x\rangle$
- △ Linearity: $\langle \alpha x+\beta y, z\rangle = \alpha\langle x,z\rangle + \beta\langle y,z\rangle$

△ Examples (Using 3 conditions to judge whether it is an inner product)

Example i)

For any $n \times n$ diagonal matrix $\mathbf{D}$ with positive diagonal entries, $\mathbf{x}^T\mathbf{D}\mathbf{y}$ is an inner product on $\mathbb{R}^n$.

$\langle x,x\rangle = x^TDx = \Sigma d_{ii}\,\Sigma x_i^2 \geq 0$

$\langle x,y\rangle = x^TDy = \Sigma d_{ii}(x_1y_1+\cdots+x_ny_n) = y^TDx = \langle y,x\rangle$

$\langle \alpha x+\beta y, z\rangle = (\alpha x+\beta y)^TDz = (\alpha x)^TDz + (\beta y)^TDz = \alpha\langle x,z\rangle + \beta\langle y,z\rangle$

z: 由自己构造

Example ii)

$\langle f,g\rangle = \int_a^b w(x)f(x)g(x)dx$ is an inner product on $C[a,b]$ where $w(x) \in C[a,b]$ is a positive function.

$\langle f,f\rangle = \int_a^b w(x)f^2(x)\,dx \geq 0$

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

$\langle \alpha f+\beta g, k\rangle = \int_a^b w(x)[\alpha f(x)+\beta g(x)]k(x)dx = \int_a^b \alpha w(x)f(x)k(x)dx + \int_a^b \beta w(x)g(x)k(x)dx = \alpha\langle f,k\rangle + \beta\langle g,k\rangle$

b. Inner product space

A vector space in which the elements (广义的vector) satisfy the inner product operation

Inner product space V_1 $\xrightarrow{\text{Inner product operation}}$ Scalar V_2

2. Norm

① Definition:

$\|v\| = \sqrt{\langle v, v\rangle}$ (2-norm)

② Normed linear space

$\{v \in V \mid \|v\|$ exists and satisfies the following 3 conditions$\}$

△ $\|v\| \geq 0$ equal iff $v = 0$ — positive-definite

△ $\|\alpha v\| = |\alpha| \|v\|$ — linearity

△ $\|v+w\| \leq \|v\| + \|w\|$

相比于 inner product space, 减少了 symmetry 的要求

proof:

△ $\|\alpha v\| = \sqrt{\langle \alpha v, \alpha v\rangle} = \sqrt{\alpha^2 v^T v} = |\alpha| \sqrt{\langle v, v\rangle} = |\alpha| \|v\|$

△ $\|v+w\|^2 = \langle v+w, v+w\rangle = \|v\|^2 + \|w\|^2 + 2\langle v, w\rangle \leq \|v\|^2 + \|w\|^2 + 2\|v\|\|w\| = (\|v\| + \|w\|)^2$

③ Types

1-norm $\quad \|X\|_1 = \sum_{i=1}^{n} |x_i|$

p-norm $\quad \|x\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$

infinity norm $\quad \|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$

1-norm 求和

∞-norm 取最大值

3. Orthogonality

① Orthogonal

a. Definition

$\langle x, y\rangle = 0$ / $\cos\theta = 0$ $\longrightarrow$ x & y are orthogonal, denoted by $x \perp y$

b. Property

x & y are orthogonal $\longrightarrow$ x & y are linearly independent (if x & y are both not 0)

proof:

Let $\alpha x + \beta y = 0$

$\alpha \|x\|^2 = \alpha \|x\|^2 + \beta \langle x, y\rangle$

$= \langle \alpha x, x\rangle + \langle x, \beta y\rangle$

$= \langle x, \alpha x\rangle + \langle x, \beta y\rangle$ $\Rightarrow$ the same as β, $\beta = 0$ $\Rightarrow$ x & y are linearly independent

$= \langle x, \alpha x + \beta y\rangle$

$= \langle x, 0\rangle$

$= 0$

$\Rightarrow \alpha = 0$ ($\|x\| > 0$)

② Orthogonal subspaces

a. Definition

$x \perp y$ for every $x \in X, y \in Y \longrightarrow X$ & Y are orthogonal, denoted by $X \perp Y$

b. Properties

From definition

△ X & Y are orthogonal $\longrightarrow X \cap Y = \{0\}$ $\quad x = y \longrightarrow x = y = 0$

零向量与任何向量都垂直；
除了零向量之外无其他交集。
(零向量是任意向量的linear combination)

proof:

for any $k \in X \cap Y$, $\langle k, k \rangle = 0 \longrightarrow k = 0$

From properties of orthogonal vectors

△ B_1 is a basis of X, B_2 is a basis of Y $\longrightarrow (B_1 + B_2)$ is a basis of $(X+Y)$

即B_1中的所有向量与B_2中的所有向量都独立，两组基底无交集。

proof:

$X + Y = \text{Span}\{x_1, \cdots, x_n, y_1, \cdots, y_m\}$

Let $\sum_{i=1}^{n} a_i x_i + \sum_{j=1}^{m} \beta_j y_j = 0 \longrightarrow \sum_{i=1}^{n} a_i x_i = -\sum_{j=1}^{m} \beta_j y_j \longrightarrow \sum_{i=1}^{n} a_i x_i = \sum_{j=1}^{m} \beta_j y_j$

Since $\begin{cases} \sum_{i=1}^{n} a_i x_i \in X, \sum_{j=1}^{m} \beta_j y_j \in Y \\ X \cap Y = \{0\} \end{cases} \longrightarrow \sum_{i=1}^{n} a_i x_i = \sum_{j=1}^{m} \beta_j y_j = 0$

Since $\begin{cases} x_1, \cdots, x_n \text{ are linearly independent} \\ y_1, \cdots, y_n \text{ are linearly independent} \end{cases} \Rightarrow \begin{cases} a_i = 0 \\ \beta_j = 0 \end{cases} \longrightarrow \{x_1, \cdots, x_n, y_1, \cdots, y_m\}$ are linearly independent

$\longrightarrow (B_1 + B_2)$ is a basis of $X + Y$

△ $\dim(X+Y) = \dim(X) + \dim(Y)$

△ $v \in X + Y$, then $v = x_i + y_j$ $(x_i \in X, y_j \in Y)$, and here x_i, y_j are unique

proof: $\quad = \sum_k c_k b_{1k} + \sum_p d_p b_{2p} \longrightarrow$ Key: the coefficient in linear combination is unique

Assume $v = x + y = x' + y' \Rightarrow x - x' = y' - y$

Since $X \cap Y = 0$, then $x - x' = y' - y = 0 \longrightarrow x = x'$ & $y = y'$ $\longrightarrow$ the sum is unique

③ Orthogonal complement

a. Definition

$Y^{\perp} = \{x \in V \mid \langle x, y \rangle = 0 \text{ for every } y \in Y\}$, $Y \in V$

all vectors in V

Complement

b. Example

Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be the standard basis of $\mathbb{R}^3$. $\longrightarrow V$

Let $\mathcal{X} = \mathrm{Span}\{\mathbf{e}_1\}$ and $\mathcal{Y} = \mathrm{Span}\{\mathbf{e}_2\}$.

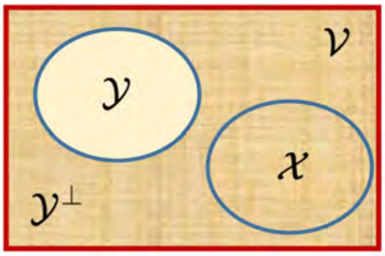
- $\mathcal{X}$ and $\mathcal{Y}$ are orthogonal $\boxed{\text{but they are NOT orthogonal complements to each other.}}$
- $\mathcal{X}^{\perp} \Rightarrow \mathrm{Span}\{\mathbf{e}_2, \mathbf{e}_3\}$
- $\mathcal{Y}^{\perp} \Rightarrow \mathrm{Span}\{\mathbf{e}_1, \mathbf{e}_3\}$

因为不满足包含所有向量

垂直空间+所有

↓

最大orthogonal subspace



b. Properties

△ If Y is a subset of V, then $Y^{\perp}$ is also a subset of V. 向量同维

△ $\mathrm{Null}(A^T) = \mathrm{Col}(A)^{\perp}$

proof:

i) $\mathrm{Null}(A^T) \subset \mathrm{Col}(A)^{\perp}$

For any $x \in \mathrm{Null}(A^T)$, $a_i^T x = 0$, for $i = 1, \cdots n$, then $a_i \perp x$, so $\mathrm{Null}(A^T) \subset \mathrm{Col}(A)^{\perp}$;

ii) $\mathrm{Col}(A)^{\perp} \subset \mathrm{Null}(A^T)$

For any $y \in \mathrm{Col}(A)^{\perp}$, $y \perp a_i$, for $i = 1, \cdots, n$, then $a_i^T y = 0$, so $\mathrm{Col}(A)^{\perp} \subset \mathrm{Null}(A^T)$

△ S & $S^{\perp}$ are orthogonal: $S \perp S^{\perp}$

△ $(S^{\perp})^{\perp} = S$

proof:

i) $S \subset (S^{\perp})^{\perp}$

For any $x \in S$, $\{x\} \perp S^{\perp} \longrightarrow S \subset (S^{\perp})^{\perp}$

ii) $(S^{\perp})^{\perp} \subset S$

在R^n内,向量要么在S中,要么在$S^{\perp}$中。

For any $y \in (S^{\perp})^{\perp}$, $\underbrace{y = u + v \; (u \in S \;\&\; v \in S^{\perp})}_{\text{uniquely}}$, $\xrightarrow{\subset R^n}$

$(S^{\perp})^{\perp} \perp S^{\perp} \longrightarrow \langle y, v \rangle = \langle v, v \rangle + \langle u, v \rangle = 0 \longrightarrow \langle v, v \rangle = 0 \longrightarrow v = 0 \longrightarrow y = u \in S \longrightarrow (S^{\perp})^{\perp} \subset S$

$(S)^{\perp} \perp S \longrightarrow \langle v, u \rangle = 0$

$\longrightarrow (S^{\perp})^{\perp} = S$

证明S中包含所有满足条件的向量

⚠ $R^n = S + S^\perp$

proof:

$$\begin{cases} S+S^\perp \subset R^n \\ \dim(S+S^\perp) = \dim(S) + \dim(S^\perp) \end{cases}$$

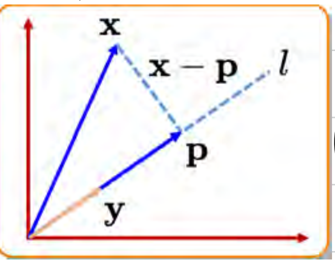
Since $\begin{cases} S = \{0\},\ S^\perp = R^n \\ S \neq \{0\},\ \text{let } \{a_1, \cdots, a_r\} \text{ be the basis for } S,\ \longrightarrow S^\perp = Col(A)^\perp = Null(A^T) \longrightarrow \dim(S^\perp) = \dim(R^n) - rank(A^T) = n - r \end{cases}$

$\Rightarrow \dim(S+S^\perp) = \dim(S) + \dim(S^\perp) = n$

$\Rightarrow$ n independant vectors span $R^n \longrightarrow R^n = S + S^\perp$

4. Projection onto a line

① Definition: 点线垂直



Find a point p on l s.t. it's closest to x.

$\begin{cases} \text{Vector projection} & p = a^* y \\ \text{Scalar projection} & a^* = \dfrac{\langle x, y\rangle}{\|y\|^2} = \dfrac{\langle x, y\rangle}{\langle y, y\rangle} \end{cases}$ ☆

关于张成的系数，分母就是那

proof:

$\|x\|^2 = \|p\|^2 + \|x-p\|^2 = a^{*2}\|y\|^2 + \langle x-p, x-p\rangle = a^{*2}\|y\|^2 + \|x\|^2 + \|p\|^2 - 2\langle x, p\rangle = a^{*2}\|y\|^2 + \|x\|^2 - 2a^*\langle x, y\rangle + a^{*2}\|y\|^2$

$\Rightarrow 2a^{*2}\|y\|^2 = 2a^*\langle x, y\rangle \Rightarrow a^* = \dfrac{\langle x, y\rangle}{\|y\|^2}$

② Example

In $\mathbb{R}^2$, find the point Q on $y = \frac{1}{3}x$ that is closest to the point $(1, 4)$

Solution:

$\mathbf{x} = \begin{bmatrix}1\\4\end{bmatrix}$ $\mathbf{y} = \begin{bmatrix}3\\1\end{bmatrix}$ ➡ $\alpha^* = \dfrac{\langle \mathbf{x}, \mathbf{y}\rangle}{\|\mathbf{y}\|^2} = \dfrac{7}{10}$ ➡ $\mathbf{p} = \dfrac{7}{10}\begin{bmatrix}3\\1\end{bmatrix}$

x, y 根据条件自己设出

Slide 10

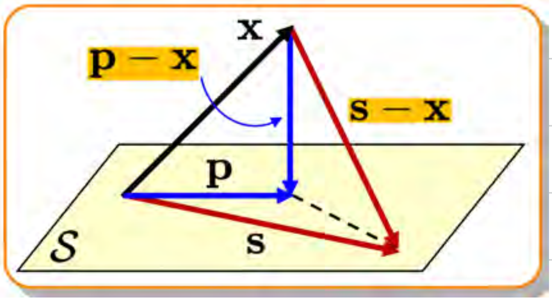
1. Projection onto a subspace of R^n

① Definition

The unique vector p, $p \in S$, that is closest to x, i.e $\|p-x\|$ takes minimum ⟶ ⋯ a subspace

p is on the line L ⟶ projection onto a line

$x = \underbrace{(x-p)}_{S^\perp} + \underbrace{p}_{S}$ ⟶ $(x-p)$ & p are linearly independent ⟶ p & $(p-x)$ are both unique w.r.t. x



② Computation

Let $\{a_1, \cdots, a_r\}$ be a basis of S and let $A = [a_1 \cdots a_r]$

Use matrix to represent a subspace, that is, subspace = Col(A)

Then, the following three statements $\triangleq$ $p = A\underset{\text{unique}}{c}$ is the projection of x onto S

$$\begin{cases} (x-Ac) \in S^\perp \\ a_i^T(x-Ac) = 0 \quad \text{for } i=1,\cdots,r \\ A^T(x-Ac) = 0 \end{cases}$$

⟶ i.e. $(x-Ac) \perp S \Rightarrow p = Ac$ is the projection onto S

⇒ We summerize a necessary and sufficient condition that $p = Ac$ is the projection of x onto S:

$A^TAc = A^Tx \quad (A^Tp = A^Tx)$ when A is a random matrix

A is the "basis matrix" of the subspace S

Since A is a matrix formed by a basis, then $\dim(Null(A)) = 0$

Since $Null(A) = Null(A^TA)$, then $\dim(Null(A^TA)) = 0$

So, (A^TA) is invertible

$$\Rightarrow \begin{cases} P = A(A^TA)^{-1}A^T \text{ named "projection matrix"} \\ p = Px \end{cases}$$

（当A满秩时才有所谓的投影矩阵）

$p = A(A^TA)^{-1}A^Tx$ when A is of full rank

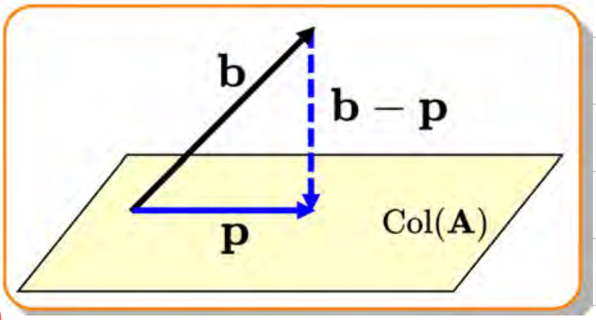
2. Least square solutions

① Introduction & Definition

$Ax=b$ may not have a solution ⟶ we can find a value of x that $\underset{p}{\underline{Ax}}$ and b is close
$b \notin Col(A)$

⟶ the value of x that minimizes $\|Ax-b\|$ is called a least squares solution of $Ax=b$ (at this time, $b \to p$)

② Computation



Main ideas
- a. compute the projection p of b onto $Col(A)$
- b. solve $Ax_{LS}=p$

$A^TAc=A^Tx \xrightarrow[x \text{ here is } b]{Ac=p=Ax} A^TAx=A^Tb$

$\Rightarrow \underline{A^TAx=A^Tb} \xrightarrow{\text{if be of full rank}} \underline{\hat{x}_{LS}=(A^TA)^{-1}A^Tb}$ (直接求 $\hat{x}_{LS}$)

The solution of $A^TAx=A^Tb$ may not be unique (it depends on whether A is of full rank or not), but $p=Ax_{LS}$ must be unique! ⟶ Least square solution may not be unique; Projection is unique

(p & b−p are both unique)

example

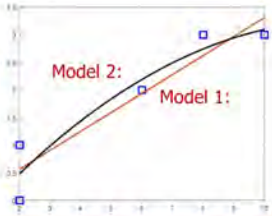
Rank=1 $\mathbf{A}=\begin{bmatrix}1&2\\1&2\\1&2\end{bmatrix}$ $\mathbf{b}=\begin{bmatrix}1\\1\\2\end{bmatrix} \notin \text{Col}(\mathbf{A})$ ➡ $\mathbf{A}^T\mathbf{A}=\begin{bmatrix}1&1&1\\2&2&2\end{bmatrix}\begin{bmatrix}1&2\\1&2\\1&2\end{bmatrix}=\begin{bmatrix}3&6\\6&12\end{bmatrix}$ Rank=1

$\begin{bmatrix}3&6\\6&12\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}1&1&1\\2&2&2\end{bmatrix}\begin{bmatrix}1\\1\\2\end{bmatrix}=\begin{bmatrix}4\\8\end{bmatrix}$ ➡ $x_1+2x_2=\frac{4}{3}$ ➡ $\mathbf{A}\hat{\mathbf{x}}=\mathbf{A}\hat{\mathbf{x}}'=\frac{4}{3}\begin{bmatrix}1\\1\\1\end{bmatrix}$ unique

$\hat{\mathbf{x}}=\begin{bmatrix}\frac{1}{3}\\\frac{1}{2}\end{bmatrix}$ $\hat{\mathbf{x}}'=\begin{bmatrix}1\\\frac{1}{6}\end{bmatrix}$ not unique

③ Examples

x	2	2	6	8	10
y	0	1	2	3	3



Model 1: $y=kx+c$

$\underbrace{\begin{bmatrix}2&1\\2&1\\6&1\\8&1\\10&1\end{bmatrix}}_{\mathbf{A}}\underbrace{\begin{bmatrix}k\\c\end{bmatrix}}_{\mathbf{x}}=\underbrace{\begin{bmatrix}0\\1\\2\\3\\3\end{bmatrix}}_{\mathbf{w}}$

➡ $\hat{\mathbf{x}}=(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{w}$

➡ $\mathbf{x}_{LS}=\begin{bmatrix}0.3438\\-0.1250\end{bmatrix}$

or $y=0.3438x-0.1250$

Model 2: $y=ax^2+bx+c$

$\underbrace{\begin{bmatrix}4&2&1\\4&2&1\\36&6&1\\64&8&1\\100&10&1\end{bmatrix}}_{\text{Coefficient matrix}}\underbrace{\begin{bmatrix}a\\b\\c\end{bmatrix}}_{\text{Unknown vector}}=\begin{bmatrix}0\\1\\2\\3\\3\end{bmatrix}$ ➡ $\mathbf{x}_{LS}=\begin{bmatrix}-0.0251\\0.6279\\-0.6712\end{bmatrix}$

$y=-0.0251x^2+0.6279x-0.6712$

3. Orthonormal sets

① Definition of Orthogonal set

In the orthogonal set, $\langle v_i, v_j \rangle = 0$ when $i \neq j$

② Definition of Orthonormal set

An orthonormal set is an orthogonal set of unit vectors (orthonormal = orthogonal + unit)

i.e. $\langle u_i, u_j \rangle = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$

③ Conversion

$u_j = \dfrac{v_j}{\|v_j\|}$ (unitize)

④ Example

Check that the following set of vectors in $\mathbb{R}^3$ is orthogonal:

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}, \begin{bmatrix} 4 \\ -5 \\ 1 \end{bmatrix} \right\}$$

Form an orthonormal set using the above set.

Solution:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 1 & -5 \\ 1 & -3 & 1 \end{bmatrix} \Rightarrow \mathbf{A}^T\mathbf{A} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 42 \end{bmatrix} \Rightarrow \text{Orthogonal}$$

Orthonormal set $\left\{ \dfrac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \dfrac{1}{\sqrt{14}} \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}, \dfrac{1}{\sqrt{42}} \begin{bmatrix} 4 \\ -5 \\ 1 \end{bmatrix} \right\}$

判断 orthogonality:

i> 向量数目少: by definition;

ii> 向量数目多: write down (A^TA), if (A^TA) is diagonal, then vectors are orthogonal

when $i \neq j$, $a_i^T a_j = 0$

判断完orthogonal后直接转化为orthonormal set

⑤ Orthonormal basis

a. Definition

$B = \{u_1, \cdots, u_n\}$ is an orthonormal set in vector space V, then B is an orthonormal basis of V

orthonormal $\longrightarrow$ independent $\longrightarrow$ basis

b. Properties:

△ If $x = \sum c_i u_i$, then $c_i = \langle x, u_i \rangle$

$x \longrightarrow c_1, c_2, c_3 \cdots c_n$

proof:

$\langle x, u_i \rangle = \langle \sum c_j u_j, u_i \rangle = \sum c_j \langle u_j, u_i \rangle = c_i$

△ If $x = \sum a_i u_i$, $y = \sum b_i u_i$, then $\sum_i a_i b_i = \langle x, y \rangle$

proof:

$\langle x, y \rangle = \sum_j \sum_i \langle a_i u_i, b_j u_j \rangle = \sum_j \sum_i a_i b_j \langle u_i, u_j \rangle = \sum_i a_i b_i$

△ If $x=\sum c_i u_i$, then $\sum_1 c_i^2 = \|x\|^2$

proof:

$\sum_1 c_i^2 = \langle x, x \rangle$

C. Projection with orthonormal basis

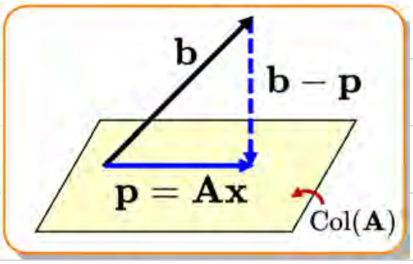
△ Let A be the matrix formed by an orthonormal basis, then $A^TA=I$

B orthogonal set (B^TB) is diagonal

△ Let $\{a_1, \cdots, a_r\}$ be an orthonormal basis of S,

then $p = A(A^TA)^{-1}A^Tb = AA^Tb = A\begin{bmatrix} a_1^T \\ \vdots \\ a_r^T \end{bmatrix} b = [a_1 \cdots a_r]\underbrace{\begin{bmatrix} \langle a_1, b\rangle \\ \vdots \\ \langle a_r, b\rangle \end{bmatrix}}_{\text{"inner product"}} = \underbrace{\sum_{i=1}^{r} \langle a_i, b\rangle a_i}_{\text{linear combination of } a_i\text{'s}}$

取至所有基向量



proof:

For any $a_j \in S$, $\langle b-p, a_j\rangle = \langle b, a_j\rangle - \langle p, a_j\rangle = \langle b, a_j\rangle - \langle \sum \langle a_i, b\rangle a_i, a_j\rangle = \langle b, a_j\rangle - \sum \langle a_i, b\rangle\langle a_i, a_j\rangle = \langle b, a_j\rangle - \langle a_j, b\rangle \times 1 = 0$

$\longrightarrow x-p \in S^\perp$

4. Orthogonal matrix

① Definition

An $n \times n$ matrix Q is an orthogonal matrix if its column vectors form an orthonormal set.

② Theorem

Q is orthogonal $\longleftrightarrow Q^TQ=I$

③ Properties

a. The cols of Q form an orthonormal basis of R^n <可以 span R^n>

b. $Q^T = Q^{-1}$

proof:

$\left\{\begin{array}{l} Q^TQ=I \\ Q \,\&\, Q^T \text{ are both square} \end{array}\right. \longrightarrow Q^T = Q^{-1}$

c. Q^T is orthogonal

proof:

$(Q^T)^TQ^T = QQ^T = QQ^{-1} = I$

d. $\langle Qx, Qy\rangle = \langle x, y\rangle$

proof:

$\langle Qx, Qy\rangle = (Qx)^T Qy = x^T Q^T Q y = x^T y = \langle x, y\rangle$

e. $\|Qx\| = \|x\|$

proof:

$\langle Qx, Qx\rangle = \langle x, x\rangle \xrightarrow{\|x\| \text{ is non-negative}} \|Qx\| = \|x\|$

5. Gram-Schmidt orthogonalization (process)

① Introduction

$\to$ Col(A) $\to$ a_i's are from a matrix A

We want to construct an orthogonormal set (basis) $\{u_1, \cdots, u_n\}$ s.t. $\text{Span}\{u_1, \cdots, u_k\} = \text{Span}\{a_1, \cdots, a_k\}$ for $k = 1, \cdots, n$

orthogonalization $\longrightarrow$ orthonormal set

one by one

② Lemma

Let $\{u_1, \cdots, u_k\}$ be an orthonormal set, $\{u_1, \cdots, u_k\} \subset$ inner product space V

Let x be a vector, $x \in$ inner product space V, but x is not in $\text{Span}\{u_1, \cdots, u_k\}$

Let $\begin{cases} v = x - \sum_{i=1}^{k} \langle x, u_i\rangle u_i \\ u = \dfrac{v}{\|v\|} \end{cases} \longrightarrow \begin{cases} \{u_1, \cdots, u_k, u\} \text{ is an orthonormal set} \quad u \text{ 与 } \text{Span}\{u_1, \cdots, u_k\} \perp \\ \text{Span}\{u_1, \cdots, u_k, u\} = \text{Span}\{u_1, \cdots, u_k, x\} \end{cases}$

proof:

i) $\{u_1, \cdots, u_k, u\}$ is an orthonormal set

$\begin{cases} \text{For any } u_j \in \{u_1, \cdots, u_k\},\ \langle v, u_j\rangle = \langle x - \sum_{i=1}^{k} \langle x, u_i\rangle u_i, u_j\rangle = \langle x, u_j\rangle - \sum_{i=1}^{k} \langle x, u_i\rangle\langle u_i, u_j\rangle = \langle x, u_j\rangle - \langle x, u_j\rangle = 0 \longrightarrow v \perp u_j \\ u = \dfrac{v}{\|v\|} \longrightarrow \|u\| = 1 \end{cases} \longrightarrow$ orthonormal

ii) $x = \sum_{i=1}^{k} \langle x, u_i\rangle u_i + v = \sum_{i=1}^{k} \langle x, u_i\rangle u_i + \|v\| u \longrightarrow x$ is a linear combination of $\{u_1, \cdots, u_k, u\}$

$\longrightarrow \text{Span}\{u_1, \cdots, u_k, u\} = \text{Span}\{u_1, \cdots, u_k, x\}$

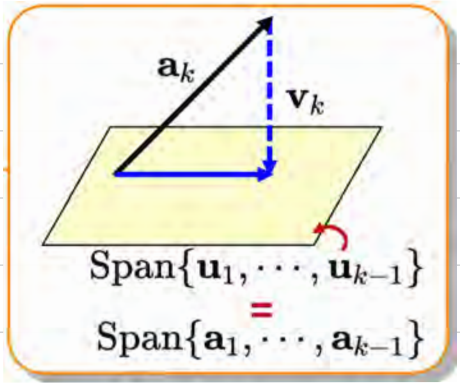
③ Process （一步步地找 u_i，直至与 a_i 个数相同时） avu

a. Let $v_1 = a_1$

b. For $k = 2, \cdots, n$, $v_k = a_k - \underbrace{\sum_{i=1}^{k-1} \langle a_k, u_i \rangle u_i}_{\text{max } i \text{ 比 } k \text{ 小 } 1} = a_k - \sum_{i=1}^{k-1} \langle a_k, \frac{v_i}{\|v_i\|} \rangle \frac{v_i}{\|v_i\|} = a_k - \sum_{i=1}^{k-1} \langle a_k, v_i \rangle \cdot \frac{v_i}{\|v_i\|^2}$

c. $u_k = \frac{v_k}{\|v_k\|}$ $\sqrt{\langle v_k, v_k \rangle}$!

$a_j = \sum_{k=1}^{j-1} \langle a_j, u_k \rangle u_k + \|v_j\| u_j \longrightarrow a_j \in \text{Span}\{u_1, \cdots, u_{j-1}, u_j\}$



④ Example

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{bmatrix}$$

$k = 1$

$$\mathbf{v}_1 = \mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

$k = 2$

$$\mathbf{a}_2 = \begin{bmatrix} -1 \\ 4 \\ 4 \\ -1 \end{bmatrix} \qquad \mathbf{v}_2 = \mathbf{a}_2 - \langle \mathbf{a}_2, \mathbf{u}_1 \rangle \mathbf{u}_1 = \begin{bmatrix} -\frac{5}{2} \\ \frac{5}{2} \\ \frac{5}{2} \\ -\frac{5}{2} \end{bmatrix} \qquad \mathbf{u}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$$

$k = 3$

$$\mathbf{a}_3 = \begin{bmatrix} 4 \\ -2 \\ 2 \\ 0 \end{bmatrix} \qquad \mathbf{v}_3 = \mathbf{a}_3 - \sum_{i=1}^{2} \langle \mathbf{a}_3, \mathbf{u}_i \rangle \mathbf{u}_i = \begin{bmatrix} 2 \\ -2 \\ 2 \\ -2 \end{bmatrix} \qquad \mathbf{u}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$$

Therefore, we fund the following orthonormal basis for $\mathbf{A}$

$$\left\{ \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}, \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}, \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} \right\}$$

u_i 的数目与 a_i 一致

6. QR decomposition

① Definition

公式应记

From the Gran-Schmidt process, $\begin{cases} a_1 = v_1 = \|v_1\| u_1 \\ \text{For } j=2,\cdots,n,\ a_j = \sum_{k=1}^{j-1} \langle a_j, u_k \rangle u_k + \|v_j\| u_j \end{cases}$

then

$$\underset{m\times n}{A} = \underset{m\times n}{Q}\ \underset{n\times n}{R} \qquad A=[a_1,\cdots,a_n]$$

$Q=[u_1,\cdots,u_n]$ (Q由GS process得来，和A同size，不一定是方阵，故非orthogonal matrix)

R is an $n\times n$ upper triangular matrix with $\begin{cases} r_{ii} = \langle a_i, u_i \rangle = \|v_i\| \\ \text{for } j>i,\ r_{ij} = \langle a_j, u_i \rangle \end{cases}$

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{bmatrix} \quad \mathbf{Q} = \begin{array}{c} \begin{array}{ccc} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{array} \\ \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} \end{array} \quad \mathbf{R} = \begin{bmatrix} \langle \mathbf{a}_1, \mathbf{u}_1 \rangle & \langle \mathbf{a}_2, \mathbf{u}_1 \rangle & \langle \mathbf{a}_3, \mathbf{u}_1 \rangle \\ 0 & \langle \mathbf{a}_2, \mathbf{u}_2 \rangle & \langle \mathbf{a}_3, \mathbf{u}_2 \rangle \\ 0 & 0 & \langle \mathbf{a}_3, \mathbf{u}_3 \rangle \end{bmatrix}$$

三个矩阵都由a_i、u_i表示

② Solving least squares

$A^TAx = A^Tb$

$\xrightarrow{A=QR} (QR)^T QRx = (QR)^Tb$

$\longrightarrow R^TQ^TQRx = R^TQ^Tb$

$\longrightarrow R^TRx = R^TQ^Tb$

Since $\|v_i\| \neq 0$, then $\det(R) \neq 0$, then R is invertible

$\longrightarrow x = (R^TR)^{-1}R^TQ^Tb$

$= R^{-1}(R^T)^{-1}R^TQ^Tb$

$= R^{-1}Q^Tb$

Slide 11

1. Eigenvalues and Eigenvectors

① Definition

$Ax = \lambda x$
- A is an n×n matrix — matrix representation of linear operator
- if x is nonzero ⟶ x is an eigenvector belonging to λ
- λ is an eigenvalue of A

Examples — Find eigenvalues & eigenvectors

Example 1

❑ Find the eigenvalues and eigenvectors of $\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix}$

i) solve characteristic equation to find eigenvalues

ii) for each λi, find the corresponding eigenvectors

Solution: $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ $(A-\lambda I)x = 0$ Non-zero → $(A-\lambda I)$ is singular

$$\det\begin{bmatrix} 3-\lambda & 2 \\ 3 & -2-\lambda \end{bmatrix} = (3-\lambda)(-2-\lambda) - 6 = \lambda^2 - \lambda - 12 = 0$$

Thus, the eigenvalues of $\mathbf{A}$ are $\lambda_1 = 4$ and $\lambda_2 = -3$

1 - 1 mapping

$$\lambda_1 = 4 \qquad (\mathbf{A} - 4\mathbf{I})\mathbf{x} = \mathbf{0} \Rightarrow \begin{bmatrix} -1 & 2 \\ 3 & -6 \end{bmatrix}\begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} = \mathbf{0} \Rightarrow \mathbf{x}_1 = \alpha\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\lambda_2 = -3 \qquad (\mathbf{A} + 3\mathbf{I})\mathbf{x} = \mathbf{0} \Rightarrow \begin{bmatrix} 6 & 2 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix} = \mathbf{0} \Rightarrow \mathbf{x}_2 = \beta\begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

Example 2

❑ Find the eigenvalues and eigenvectors of $\mathbf{A} = \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix}$

Solution: $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$

$$\det\begin{bmatrix} 2-\lambda & -3 & 1 \\ 1 & -2-\lambda & 1 \\ 1 & -3 & 2-\lambda \end{bmatrix} = 0 \Rightarrow -\lambda(\lambda-1)^2 = 0$$

Thus, the eigenvalues of $\mathbf{A}$ are $\lambda_1 = 0$ and $\lambda_2 = \lambda_3 = 1$

$$\lambda_1 = 0 \qquad \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix}\begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \end{bmatrix} = 0 \Rightarrow \mathbf{x}_1 = \alpha\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\lambda_2 = \lambda_3 = 1 \qquad \begin{bmatrix} 1 & -3 & 1 \\ 1 & -3 & 1 \\ 1 & -3 & 1 \end{bmatrix}\begin{bmatrix} x_{21} \\ x_{22} \\ x_{23} \end{bmatrix} = 0 \Rightarrow \mathbf{x}_2 = \mathbf{x}_3 = \alpha\begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + \beta\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

用 linear combination of free variables 来表示!

② Properties of eigenvalue & eigenvector

a. x & Ax have the same direction

b. If x is an eigenvector, so is (cx) for $c \neq 0$

c. If λ is an eigenvalue of A, then λ^n is an eigenvalue of A^n

proof:

$Ax = \lambda x \longrightarrow A^2x = A(Ax) = \lambda(Ax) = \lambda^2 x \longrightarrow \cdots$

d. For an invertible matrix A, if λ is an eigenvalue of A, then λ^{-1} is an eigenvalue of A^{-1}, vice versa

proof:

$Ax = \lambda x \longrightarrow A^{-1}Ax = A^{-1}\lambda x \longrightarrow x = \lambda A^{-1}x \longrightarrow \frac{1}{\lambda}x = A^{-1}x \longrightarrow A^{-1}x = \lambda^{-1}x$

the same eigenvector!

e. Equivalent characterization:

$Ax = \lambda x \longleftrightarrow (A - \lambda I)x = 0$

△ $(A - \lambda I)x = 0$ has infinite solutions (the # of eigenvectors is ∞)

△ $(A - \lambda I)$ is singular

△ Characteristic polynomial

i) Definition

$\det(A - \lambda I) = p(\lambda) = \underset{=1}{c}\,(\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda)$

$p(\lambda)$ is characteristic polynomial of A

$p(\lambda) = 0$ is characteristic equation of A

ii) Theorem

If A & B are $n \times n$ matrices and similar to each other, then they have the same characteristic polynomial and the same eigenvalues.

Similar $\longrightarrow$ the same $p(\lambda)$ $\longrightarrow$ the same $\lambda_1, \cdots, \lambda_n$

proof:

Let $B = S^{-1}AS$

then $p_B(\lambda) = \det(B - \lambda I) = \det(S^{-1}AS - \lambda I)$ (*)

Since $\lambda I = \lambda S^{-1}S = \lambda S^{-1}IS = S^{-1}(\lambda I)S$

then (*) $= \det(S^{-1}(A - \lambda I)S) = \det(S^{-1})\det(A - \lambda I)\det(S) = p_A(\lambda)$

Since $p_A(\lambda) = p_B(\lambda)$, then A & B also have the same eigenvalues

f. product & sum of eigenvalues

Let A be an n×n matrix with eigenvalues

$$\begin{cases} \prod_{i=1}^{n} \lambda_i = \det(A) \\ \sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii} = \operatorname{trace}(A) \end{cases}$$ the relationship between eigenvalues & its matrix

proof:

$$p(\lambda) = \det(A-\lambda I) = \begin{bmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{bmatrix} = \underbrace{(a_{11}-\lambda)\det(M_{11})}_{\text{Containing term of } \lambda^n} + \underbrace{\sum_{i=2}^{n} a_{i1}(-1)^{i+1}\det(M_{i1})}_{\text{Containing terms up to } \lambda^{n-2}}$$

$\longrightarrow (a_{11}-\lambda)(a_{22}-\lambda)\cdots(a_{nn}-\lambda)$ is the only term containing $\overset{\text{i)}}{\lambda^n}$ & $\overset{\text{ii)}}{\lambda^{n-1}}$

i) prove the equation about the product (focus on the coefficient of λ^n)

$\begin{cases} \text{The coefficient of } \lambda^n \text{ is } (-1)^n \\ \det(A-\lambda I) = p(\lambda) = c(\lambda_1-\lambda)(\lambda_2-\lambda)\cdots(\lambda_n-\lambda) \end{cases}$ $\longrightarrow c=1 \longrightarrow p(\lambda) = (\lambda_1-\lambda)\cdots(\lambda_n-\lambda)$

$p(0) = \det(A) = \prod_{i=1}^{n} \lambda_i$

ii) prove the equation about the sum (focus on the coefficient of $(-\lambda)^{n-1}$)

$\begin{cases} (a_{11}-\lambda)\cdots(a_{nn}-\lambda) \longrightarrow \sum_{i=1}^{n} a_{ii} \\ p(\lambda) = (\lambda_1-\lambda)\cdots(\lambda_n-\lambda) \longrightarrow \sum_{i=1}^{n} \lambda_i \end{cases}$ $\longrightarrow \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i$

③ Eigenspace

$Null(A-\lambda I)$ is called eigenspace.

All nonzero vectors in $Null(A-\lambda I)$ are eigenvectors corresponding to λ.

i.e. $Null(A-\lambda I) = \{0\} + \{\text{eigenvectors}\}$

④ Complex eigenvalues

a. Concepts

Complex number : $a+bi$ $(i^2=-1)$

Complex conjugate: $\overline{a+bi}=a-bi$ $\longrightarrow$ $\bar{A}=(\bar{a}_{ij})$ (A matrix is real if $\bar{A}=A$)

b. Theorems

For real matrix:

⚠ Complex eigenvalues occur in conjugate pairs: if λ is an eigenvalue, so is $\bar{\lambda}$

⚠ Complex eigenvectors occur in conjugate pairs: if z is an eigenvector belonging to λ, then $\bar{z}$ is an eigenvector belonging to $\bar{\lambda}$

$Az=\lambda z \longrightarrow \overline{Az}=\overline{\lambda z} \longrightarrow \bar{A}\bar{z}=\bar{\lambda}\bar{z} \longrightarrow A\bar{z}=\bar{\lambda}\bar{z}$

(此处证明略)

$$\mathbf{A}=\begin{bmatrix}1 & 2\\ -2 & 1\end{bmatrix} \Rightarrow \begin{bmatrix}1-\lambda & 2\\ -2 & 1-\lambda\end{bmatrix}=(1-\lambda)^2+4=0$$

⚠①

Thus, the eigenvalues of $\mathbf{A}$ are $\lambda_1=1+2i$ and $\lambda_1=1-2i$

⚠②

The corresponding eigenvectors are $\mathbf{x}_1=\alpha\begin{bmatrix}1\\ i\end{bmatrix}$ and $\mathbf{x}_2=\beta\begin{bmatrix}1\\ -i\end{bmatrix}$

⑤ Complex inner product

a. Definition

A mapping: Complex inner product space $\longrightarrow$ complex numbers

Complex inner product should satisfy the following 3 conditions

- Positive—definite: $\langle x,x\rangle \geq 0$ with equality iff $x=0$
- Conjugate symmetry: $\langle x,y\rangle=\overline{\langle y,x\rangle}$
- Linearity: $\langle \alpha x+\beta y,z\rangle=\alpha\langle x,z\rangle+\beta\langle y,z\rangle$ 注意：仅在前部分才满足该等式!

b. Theorem

If x & y are vectors in complex inner product space, then $|\langle x,y\rangle|\leq\|x\|\|y\|$ with equality iff x & y are linearly dependent

Proof:

$\|x-\alpha y\|^2=\langle x-\alpha y,x-\alpha y\rangle=\langle x,x-\alpha y\rangle-\alpha\langle y,x-\alpha y\rangle=\overline{\langle x-\alpha y,x\rangle}-\alpha\overline{\langle x-\alpha y,y\rangle}=\overline{\langle x,x\rangle-\alpha\langle y,x\rangle}-\alpha\overline{\langle x,y\rangle-\alpha\langle y,y\rangle}=\|x\|^2-\bar{\alpha}\langle x,y\rangle-\alpha\langle y,x\rangle+\alpha\bar{\alpha}\|y\|^2$

Let $\bar{\alpha}=\frac{\langle x,y\rangle}{\|y\|^2} \longrightarrow \|x\|^2-\frac{|\langle x,y\rangle|^2}{\|y\|^2}\geq 0 \longrightarrow |\langle x,y\rangle|\leq\|x\|\|y\|$

c. Hermitian

$z^H=\bar{z}^T$, and define an inner product by $\langle x,y\rangle=y^Hx=\bar{y}^Tx$

- $\langle y,x\rangle=x^Hy=x^H(\bar{y}^H)^T=\bar{x}^T\bar{y}^{H^T}=(\bar{y}^H\bar{x})^T=\overline{y^Hx}=\overline{\langle x,y\rangle}$
- $\|y\|=\sqrt{\langle y,y\rangle}=\sqrt{y^Hy}$

$Z\in$ complex inner product space

2. Diagonalization

① Definition

An $n\times n$ matrix A is diagonalizable if A is similar to a diagonal matrix.

can be converted into diagonal matrix

i.e. $A = X^{-1}DX$ / $D = X^{-1}AX$

diagonalizing matrix

② Theorems

a. Theorem 1: About Diagonalization

An $n\times n$ matrix A is diagonalizable $\longleftrightarrow$ A has n linearly independent eigenvectors

proof:

Let $\{\lambda_1,\cdots,\lambda_n\}$ be the eigenvalues and $\{x_1,\cdots,x_n\}$ be their corresponding eigenvectors of A

$Ax_i = \lambda_i x_i \longrightarrow AX = XD$ where $\begin{cases} X = [x_1,\cdots,x_n] \\ D = \mathrm{Diag}\{\lambda_1,\cdots,\lambda_n\} \end{cases}$ $\longrightarrow A = XDX^{-1}$ (vice versa)

每一列相等

b. Theorem 2: Distinct eigenvalues $\longrightarrow$ Linearly independent eigenvectors $\longrightarrow$ diagonalizable

Let $\lambda_1,\cdots,\lambda_k$ be distinct eigenvalues of an $n\times n$ matrix, then the corresponding eigenvectors $x_1,\cdots,x_n$ are linearly independent.

proof:

Suppose that no all of the k eigenvectors are linearly independent.

Assume $\{x_1,\cdots,x_i\}$ are the $i<k$ linearly independent eigenvectors, then x_{i+1} can be written as a linear combination of $\{x_1,\cdots,x_i\}$

$x_{i+1} = \sum_{j=1}^{i} c_j x_j \longrightarrow Ax_{i+1} = \sum_{j=1}^{i} c_j Ax_j \longrightarrow \lambda_{i+1}x_{i+1} = \sum_{j=1}^{i} c_j\lambda_j x_j \longrightarrow \lambda_{i+1}\sum_{j=1}^{i} c_j x_j = \sum_{j=1}^{i} c_j\lambda_j x_j \longrightarrow \sum_{j=1}^{i} c_j(\lambda_j-\lambda_{i+1})x_j = 0$

Since $\lambda_j \neq \lambda_{i+1}$, $x_1\cdots x_i$ are linearly independent $\longrightarrow c_j = 0,\ j=1,2,\cdots,i \longrightarrow x_{i+1} = \sum_{j=1}^{i} c_j x_j = 0$

Since eigenvectors are nonzero, then x_{i+1} is not the linear combination of $\{x_1,\cdots,x_i\}$ $\longrightarrow x_1,\cdots,x_k$ are linearly independent.

c. Theorem 3:

Let A be an $n\times n$ matrix with distinct eigenvalues $\lambda_1,\cdots,\lambda_p$

△ For $1\le k\le p$, dim (eigenspace corresponding to λ_k) $\le$ the multiplicity of λ_k

△ A is diagonalizable $\longleftrightarrow$ Σ dim (eigenspaces) $= n$

△ If A is diagonalizable and B_k is a basis for eigenspace corresponding to λ_k for each k,

then the total collections of vectors in the sets $B_1,\cdots,B_p$ forms an eigenvector basis for R^n

if a square matrix is with distinct eigenvalues:

i) eigenvectors are linearly independent $\longrightarrow$ diagonalizable

ii) dim (eigenspace corresponding to λ_k) $\le$ the multiplicity of λ_k

③ Applications

a. Compute the power of diagonalizable matrix

❑ Let $\mathbf{A}=\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}$. Compute $\mathbf{A}^8$.

Recall $\mathbf{A}=\mathbf{XDX}^{-1}$ where $\underset{\text{eigenvectors}}{\mathbf{X}=[\mathbf{x}_1,\cdots,\mathbf{x}_n]}$ & $\underset{\text{eigenvalues}}{\mathbf{D}=\mathrm{Diag}\{\lambda_1,\cdots,\lambda_n\}}$ ← $Ax_i=\lambda_i x_i$ $\Downarrow$ $AX=XD$

$$\Rightarrow \mathbf{A}^n=\underbrace{(\mathbf{XDX}^{-1})(\mathbf{XDX}^{-1})\cdots(\mathbf{XDX}^{-1})}_{n \text{ terms}}=\mathbf{XD}^n\mathbf{X}^{-1}$$

Solution: We can find

$$\mathbf{X}=\begin{bmatrix}0.5257 & -0.8507\\ -0.8507 & -0.5257\end{bmatrix} \qquad \mathbf{D}=\begin{bmatrix}\frac{1-\sqrt{5}}{2} & 0\\ 0 & \frac{1+\sqrt{5}}{2}\end{bmatrix}$$

取后检验可逆

$$\mathbf{XD}^8\mathbf{X}^{-1}=\begin{bmatrix}0.5257 & -0.8507\\ -0.8507 & -0.5257\end{bmatrix}\begin{bmatrix}\left(\frac{1-\sqrt{5}}{2}\right)^8 & 0\\ 0 & \left(\frac{1+\sqrt{5}}{2}\right)^8\end{bmatrix}\begin{bmatrix}0.5257 & -0.8507\\ -0.8507 & -0.5257\end{bmatrix}^{-1}$$

$$\Rightarrow \mathbf{A}^8=\begin{bmatrix}34 & 21\\ 21 & 13\end{bmatrix}$$

b. Fibonacci numbers (其实只是Application a的拓展,即对角化只用于求方阵的高次方)

❑ The Fibonacci sequence $\{F_n\}$ is defined as

$$F_0=0, F_1=1 \quad \text{and} \quad F_n=F_{n-1}+F_{n-2}, n\geq 2$$

❑ We can write a recursive matrix equation

$$\begin{bmatrix}F_n\\ F_{n-1}\end{bmatrix}=\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}F_{n-1}\\ F_{n-2}\end{bmatrix}=\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}^{n-1}\begin{bmatrix}1\\ 0\end{bmatrix}$$

$F_n=F_{n-1}+F_{n-2}$

$F_{n-1}=F_{n-1}$

❑ We have the diagonalization $\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}=S\begin{bmatrix}\lambda & 0\\ 0 & -\lambda^{-1}\end{bmatrix}S^{-1}$

where $\lambda=\dfrac{1+\sqrt{5}}{2}$, $S=\begin{bmatrix}1 & -1\\ \lambda^{-1} & \lambda\end{bmatrix}$ and

$Ax=\lambda x$

$$S^{-1}=\frac{1}{\lambda+\lambda^{-1}}\begin{bmatrix}\lambda & 1\\ -\lambda^{-1} & 1\end{bmatrix}$$

Slide 12

1. Singular Value Decomposition (SVD)

① Two types

a. For real symmetric matrix A — Hessian / Quadratic form

(见P55. Property No.2) A is symmetric then X is orthogonal

$A \overset{E_1}{=} UDU^T$
- U: Orthogonal matrix formed by eigenvectors
- D: Diagonal matrix formed by eigenvalues

实际上，这是Diagonalization的变形，$A = XDX^{-1} \Rightarrow A = UDU^T$

外积！

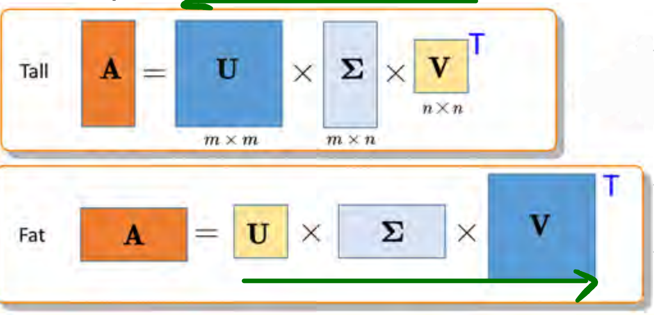
$$\overset{E_2}{=} [u_1 \cdots u_n] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix} = [u_1 \cdots u_n] \begin{bmatrix} \lambda_1 u_1^T \\ \vdots \\ \lambda_n u_n^T \end{bmatrix} = \sum_{i=1}^{n} \lambda_i u_i u_i^T$$

multiplication of partitioned matrix

Linear combination of outer product of u_i

即：实对称矩阵可直接由关于λ_i、x_i的式子表示！

b. For any real matrix A (SVD)



Σ: i) is in the same size as A

ii) diagonal entries are σ_i, $\sigma_i = \sqrt{\lambda_i}$ (singular value) ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{min}$, $\lambda_i \geq 0$) !!有顺序要求!!!

off-diagonal entries are 0s

(U和V只代表顺序，无论是fat/tall，都是$A = U\Sigma V^T$)

$$A = \underbrace{[\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_r}_{\mathbf{U}_r}\ \underbrace{\mathbf{u}_{r+1}\ \cdots\ \mathbf{u}_m]}_{\mathbf{U}_2} \begin{bmatrix} \boxed{\begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{matrix}}_{\Sigma_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \underbrace{[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_r}_{\mathbf{V}_r}\ \underbrace{\mathbf{v}_{r+1}\ \cdots\ \mathbf{v}_n]^T}_{\mathbf{V}_2}$$

$m \times r \quad m \times (m-r) \quad m \times n \quad n \times r \quad n \times (n-r)$

$= \mathbf{U}_r \cdot \boldsymbol{\Sigma}_r \cdot \mathbf{V}_r^T$ (非零特征值对应的表达式)

$= \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \Rightarrow$
- $\mathbf{A}\mathbf{v}_j = \left(\sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T\right)\mathbf{v}_j = \sigma_j \mathbf{u}_j \Rightarrow \frac{1}{\sigma_j}\mathbf{A}\mathbf{v}_j = \mathbf{u}_j$
- $\mathbf{A}^T\mathbf{u}_j = \sigma_j \mathbf{v}_j \Rightarrow \frac{1}{\sigma_j}\mathbf{A}^T\mathbf{u}_j = \mathbf{v}_j$ for $j = 1, 2, \cdots, r$

σ_i即是singular value

$u_j = \frac{1}{\sigma_j} A v_j$

$v_j = \frac{1}{\sigma_j} A^T u_j$

△ Process

i) Compute A^TA & AA^T

ii) Compute eigenvalues & eigenvectors
- eigenvalues $\longrightarrow \Sigma$ — λ_j of bigger matrix = 共同的λ_i + 0s
- eigenvectors $\longrightarrow U$ & V (from the smaller one to the bigger one)

iii) Check the sign of the bigger matrix: $u_n = \frac{1}{\sigma_n} A v_n$

* An alternative process:

i) Compute A^TA & AA^T

ii) Compute eigenvalues & eigenvectors of the smaller matrix (才可以辨认出共同的λ_i)
- eigenvalues $\longrightarrow \Sigma$
- eigenvectors $\longrightarrow V$

iii) Compute U by $u_n = \frac{1}{\sigma_n} A v_n$

⚠ Example

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{where} \quad m = 3, n = 2$$

Step 1: $\mathbf{A}\mathbf{A}^T = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and $\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$

Step 2: $\det(\mathbf{A}^T\mathbf{A} - \lambda\mathbf{I}) = 0 \Rightarrow \lambda_1 = 4, \lambda_2 = 0$

Same dimension as $\mathbf{A}$

$$\lambda_1 = 4 \quad \begin{bmatrix} 2-4 & 2 \\ 2 & 2-4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{0} \Rightarrow \mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\lambda_2 = 0 \quad \begin{bmatrix} 2-0 & 2 \\ 2 & 2-0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{0} \Rightarrow \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\mathbf{\Sigma} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad \mathbf{V} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

Step 3: $\mathbf{A}\mathbf{A}^T = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \lambda_1 = 4$ and $\lambda_2 = \lambda_3 = 0$

$$\lambda_1 = 4: \begin{bmatrix} 2-4 & 2 & 0 \\ 2 & 2-4 & 0 \\ 0 & 0 & 0-4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{0} \Rightarrow \begin{matrix} x_1 = x_2 \\ x_3 = 0 \end{matrix} \Rightarrow \mathbf{u}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$$

$$\lambda_2 = \lambda_3 = 0: \begin{bmatrix} 2-0 & 2 & 0 \\ 2 & 2-0 & 0 \\ 0 & 0 & 0-0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{0} \Rightarrow \begin{matrix} x_1 = -x_2 \\ x_3 \text{ arbitrary} \end{matrix}$$

$$\Rightarrow \mathbf{u}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \text{ and } \mathbf{u}_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \Rightarrow \mathbf{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix}$$

Step 4: Determining the sign of each vector corresponding to $\sigma_n \neq 0$ → 若$\sigma_n = 0$，则无法用 $u_n = \frac{1}{\sigma_n} A v_n$ 检验

$$\mathbf{u}_n = \frac{1}{\sigma_n} \mathbf{A}\mathbf{v}_n$$

Recall $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}$ $\quad \mathbf{V} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$ and

$$\mathbf{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix}$$

or

$$\mathbf{U} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix}$$

$$\sigma_1 = 2 \qquad \mathbf{A}\mathbf{v}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \Rightarrow 1 \times \frac{1}{\sqrt{2}} + 1 \times \frac{1}{\sqrt{2}} > 0$$

⇒ The first entry of $\mathbf{u}_1$ is positive ⇒ $\mathbf{u}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$ and $\mathbf{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix}$

$$\underset{\mathbf{A}}{\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}} = \underset{\mathbf{U}}{\begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix}} \times \underset{\mathbf{\Sigma}}{\begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}} \times \underset{\mathbf{V}^T}{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^T}$$

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T \Rightarrow \begin{cases} \mathbf{A}\mathbf{A}^T = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \boxed{\mathbf{V}^T \cdot \mathbf{V}} \cdot \mathbf{\Sigma}^T \cdot \mathbf{U}^T = \mathbf{U} \cdot \boxed{\mathbf{\Lambda}} \cdot \mathbf{U}^T \\ \mathbf{A}^T\mathbf{A} = \mathbf{V} \cdot \mathbf{\Sigma}^T \cdot \overset{I}{\boxed{\mathbf{U}^T \cdot \mathbf{U}}} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T = \mathbf{V} \cdot \boxed{\tilde{\mathbf{\Lambda}}} \cdot \mathbf{V}^T \end{cases}$$

$D = \Sigma\Sigma^T$

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \quad \begin{bmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix}^T$$

$$\mathbf{A}^T\mathbf{A} = \mathbf{V}\tilde{\mathbf{\Lambda}}\mathbf{V}^T \quad \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \times \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^T$$

② Properties

a. Properties of real symmetric matrix

△ The eigenvalues of a real symmetric matrix are all real.

△ Eigenvectors belonging to distinct eigenvalues are orthogonal. $\longrightarrow A = XDX^{-1} \Rightarrow A = UDU^T$

proof:

$$x_i^TAx_j = \begin{cases} \lambda_j x_i^T x_j \\ x_i^TA^Tx_j = (Ax_i)^Tx_j = (\lambda_i x_i)^T x_j = \lambda_i x_i^T x_j \end{cases} \xrightarrow{\lambda_i \neq \lambda_j} x_i^Tx_j = 0 \longrightarrow x_i \perp x_j$$

b. Properties of A^TA

△ A^TA is symmetric

△ The eigenvalues of A^TA are non-negative

proof:

$$A^TAx = \lambda x \longrightarrow x^TA^TAx = \lambda x^Tx \longrightarrow \|Ax\|^2 = \lambda \|x\|^2 \longrightarrow \lambda = \frac{\|Ax\|^2}{\|x\|^2} \geq 0$$

c. Properties of A (A → SVD)

$$A = [\underbrace{\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_r}_{\substack{\mathbf{U}_r \\ m\times r}}\ \underbrace{\mathbf{u}_{r+1}\ \cdots\ \mathbf{u}_m}_{\substack{\mathbf{U}_2 \\ m\times(m-r)}}]\ \underbrace{\begin{bmatrix} \boxed{\begin{matrix}\sigma_1 & & \mathbf{\Sigma}_r \\ & \ddots & \\ & & \sigma_r\end{matrix}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}}_{m\times n}\ [\underbrace{\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_r}_{\substack{\mathbf{V}_r \\ n\times r}}\ \underbrace{\mathbf{v}_{r+1}\ \cdots\ \mathbf{v}_n}_{\substack{\mathbf{V}_2 \\ n\times(n-r)}}]^T$$

$$= \mathbf{U}_r\cdot\mathbf{\Sigma}_r\cdot\mathbf{V}_r^T$$

$$= \sum_{i=1}^{r}\sigma_i\mathbf{u}_i\mathbf{v}_i^T \Rightarrow \begin{cases} \mathbf{A}\mathbf{v}_j = \left(\sum_{i=1}^{r}\sigma_i\mathbf{u}_i\mathbf{v}_i^T\right)\mathbf{v}_j = \sigma_j\mathbf{u}_j \Rightarrow \dfrac{1}{\sigma_j}\mathbf{A}\mathbf{v}_j = \mathbf{u}_j \\ \mathbf{A}^T\mathbf{u}_j = \sigma_j\mathbf{v}_j \Rightarrow \dfrac{1}{\sigma_j}\mathbf{A}^T\mathbf{u}_j = \mathbf{v}_j \quad \text{for } j = 1,2,\cdots,r \end{cases}$$

△ rank(A) = r, where r is the # of nonzero singular values

proof:

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

$\text{rank}(AB) = \begin{cases} \text{rank}(A), \text{ if } B \text{ is of full rank} \\ \text{rank}(B), \text{ if } A \text{ is of full rank} \end{cases}$

△ $u_1, \cdots, u_r$ form an orthonormal basis of Col(A)

proof:

$y \in Col(A) \longrightarrow y = Ax = U_r\cdot\Sigma_r\cdot V_r^T\cdot x = U_r\cdot c = \sum_{i=1}^{r} c_iu_i \longrightarrow y = \text{Span}\{u_1, \cdots, u_r\}$

△ $u_{r+1}, \cdots, u_m$ form an orthonormal basis of $Null(A^T)$

proof:

$A^Tu_j = (U_r\cdot\Sigma_r\cdot V_r^T)^T\cdot u_j = 0 \longrightarrow u_j \in Null(A^T),\ j = r+1, \cdots, m$

$\dim(Null(A^T)) = m - \text{rank}(A^T) = m - r$

△ $v_1, \cdots, v_r$ form an orthonormal basis of Row(A) $(Col(A^T))$

proof:

$y^T \in Row(A) \longrightarrow y^T = x^TA = x^TU_r\cdot\Sigma_r\cdot V_r^T \longrightarrow y^T \in \text{Span}\{v_1^T, \cdots, v_r^T\}$

△ $v_{r+1}, v_{r+2}, \cdots, v_n$ form an orthonormal basis of Null(A)

proof:

$Av_j = (U_r\cdot\Sigma_r\cdot V_r^T)\cdot v_j = 0 \longrightarrow v_j \in Null(A),\ j = r+1, \cdots, n \longrightarrow \dim(Null(A)) = n - \text{rank}(A) = n - r$

2. Frobenius norm

① Definition

$\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2}$ (Define the norm: vector $\longrightarrow$ matrix)

② Properties

a. $\|A\|_F = \|A^T\|_F$

b. If Q is orthogonal, then $\|QA\|_F = \|A\|_F$ (类比 $\langle Qx, Qx\rangle = \langle x, x\rangle$)

proof:

$\|QA\|_F^2 = \|[Qa_1 \cdots Qa_n]\|_F^2 = \sum_{i=1}^{n} \|\underbrace{Qa_i}_{\text{vector}}\|_F^2 = \sum_{i=1}^{n} a_i^T Q^T Q a_i = \sum_{i=1}^{n} \|a_i\|_F^2 = \|A\|_F^2$

c. If $A = U\Sigma V^T$, then $\|A\|_F^2 = \sum_{i=1}^{n} \sigma_i^2$

proof:

$\|A\|_F^2 = \|U\Sigma V^T\|_F^2 = \|\Sigma V^T\|_F^2 = \|V\Sigma^T\|_F^2 = \|\Sigma\|_F^2 = \sum_{i=1}^{m} \sigma_i^2$

③ Low-rank approximation

For a fixed $m \times n$ matrix A and an integer k, solve $\min_{\text{rank}(S) \le k} \|A - S\|_F$

Solution: $\mathbf{A} = \mathbf{U}_r \cdot \boldsymbol{\Sigma}_r \cdot \mathbf{V}_r^T$

$$\mathbf{A} = [\underbrace{\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_k}_{\mathbf{U}_k} \ \underbrace{\mathbf{u}_{k+1} \ \cdots \ \mathbf{u}_r}_{\mathbf{U}_2}] \begin{bmatrix} \boxed{\begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{matrix}}^{\,\boldsymbol{\Sigma}_k} & \mathbf{0} \\ \mathbf{0} & \begin{matrix} \sigma_{k+1} & & \\ & \ddots & \\ & & \sigma_r \end{matrix} \end{bmatrix} [\underbrace{\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_k}_{\mathbf{V}_k} \ \underbrace{\mathbf{v}_{k+1} \ \cdots \ \mathbf{v}_r}_{\mathbf{V}_2}]^T$$

$$\min_{\text{rank}(\mathbf{s}) \le k} \|\mathbf{A} - \mathbf{S}\|_F = \sqrt{\sum_{i=k+1}^{n} \sigma_i^2}$$

Slide 13

1. Quadratic form

① Definition

A function $R^n \to R$: $Q(x) = x^T A x$, where A is a symmetric matrix

② Examples

$n=1 \quad Q(x) = ax^2$

$n=2 \quad Q(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2$

$n=3 \quad Q(x_1,x_2,x_3) = ax_1^2 + bx_2^2 + cx_3^2 + 2dx_1x_2 + 2ex_1x_3 + 2fx_2x_3$

$$Q(x_1,x_2,x_3) = 5x_1^2 + 3x_2^2 + 4x_3^2 - 2x_1x_2 + 6x_2x_3$$

$$\Rightarrow \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a\,5 & d\,{-1} & e\,0 \\ -1 & b\,3 & f\,3 \\ 0 & 3 & c\,4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

③ Change of variables

(Change x into y)

a. Definition

旧元→新元

$x = Py$ (P is invertible) $\longrightarrow x^TAx = (Py)^TA(Py) = y^TP^TAPy = y^TBy$

y can be regarded as the coordinate vector of x under the basis formed by cols of P

New quadratic form

b. Theorem

If A is real symmetric matrix, then there is a change of variables $x = Uy$ s.t. $x^TAx = y^TDy$

$\begin{cases} A = UDU^T \\ x = Uy \end{cases} \longrightarrow x^TAx = y^TU^TAUy = y^TDy$

④ Types

a. For a quadratic form $Q(x) = x^TAx$

- positive definite : $\forall x \neq 0$, $Q(x) > 0$
- negative definite : $\forall x \neq 0$, $Q(x) < 0$
- indefinite : if $Q(x)$ takes both positive and negative values

b. For a real symmetric matrix A

- positive definite : $Q_A(x)$ is positive definite $\longrightarrow$ λ's are positive
- negative definite : $Q_A(x)$ is negative definite $\longrightarrow$ λ's are negative
- indefinite : $Q_A(x)$ is indefinite $\longrightarrow$ λ's are positive or negative

proof:

A is positive definite $\longrightarrow x^TAx > 0, \forall x \neq 0 \longrightarrow \lambda x^Tx > 0, \forall x \neq 0 \longrightarrow \lambda > 0$