代数-2.矩阵

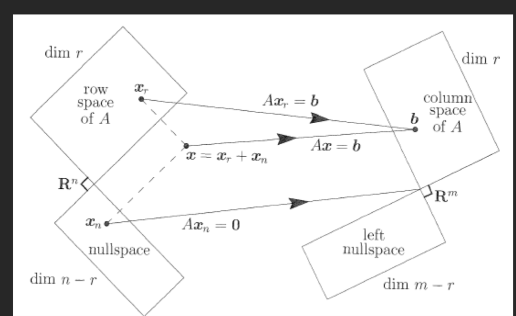
属性

矩阵的秩

- 定义
  - $A \in R^{m\times n}$, 秩 $Rank(A) = \dim ColA$(列秩) $= \dim RowA$(行秩)
- 性质
  - 等价矩阵等秩
    - $A, B \in R^{m\times n}, P$ 可逆$\in R^m, Q$ 可逆$\in R^n$;
    - 称 A 与经任意次初等行变换或 (同时) 初等列变换得到的矩阵 B 等价, 即
    - $A \cong B \iff B = PAQ.$
    - 所有与 A 等价的矩阵均等价于形如 $\begin{bmatrix} I_{rank(A)} & 0 \\ 0 & 0 \end{bmatrix}$ 的唯一等价标准型
  - 秩不等式
    1. $A \in R^{m\times n}, 0 \le r(A) \le \min\{m, n\}$
    2. $A \in R^{m\times n}, B \in R^{n\times s}, r(A) + r(B) - n \le r(AB) \le \min\{r(A), r(B)\}$
    3. $A \in R^{m\times n}, r(A) = r(A^\top) = r(AA^\top) = r(A^\top A)$
    4. $A, B \in R^{m\times n}, r(A+B) = r(B) \le r([A, B]) \le r(A) + r(B)$
    5. 若 $A \cong \Lambda$, 则 $rank(A) = rank(\Lambda) = \#\Lambda$ 非 0 元个数
    6. $A \in R^{m\times n}, B \in R^{n\times k}, r(AB) \le \min\{r(A), r(B)\}$

矩阵的子空间

- 定义
  - 零空间 $NulA$
    - 矩阵 $A \in R^{m\times n}$ 的零空间 $NulA$ 是 $Ax = 0$ 的解空间
    - $NulA = \{x | x \in R^n \wedge Ax = 0\}$
  - 左零空间 $NulA^\top$
    - 矩阵 $A \in R^{m\times n}$ 的左零空间 $NulA^\top$ 是 $A^\top x = 0$ 的解空间
    - $NulA^\top = \{x | x \in R^m \wedge A^\top x = 0\}$
  - 列空间 $ColA$
    - 矩阵 $A = [a_1 \quad a_2 \quad \cdots \quad a_n] \in R^{m\times n}$ 的列空间 $ColA$ 是矩阵各列的线性组合
    - $ColA = Span\{a_1, a_2, \ldots, a_n\} = \{b | b = Ax, \forall x \in R^n\} = range(x \mapsto Ax)$
  - 行空间 $RowA$
    - 矩阵 $A = [a_1^\top \quad a_2^\top \quad \cdots \quad a_n^\top] \in R^{m\times n}$ 的行空间 $RowA = ColA^T$ 是矩阵各行的线性组合
    - $RowA = Span\{a_1^\top, a_2^\top, \ldots, a_n^\top\} = \{b | b = A^\top x, \forall x \in R^m\} = range(x \mapsto A^\top x)$
- 子空间关系



(线性代数基本定理)$A \in R^{m\times n}$,

1[正交角度] $\mathrm{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)^{\perp} = \mathrm{Null}(\boldsymbol{A}), \mathrm{Col}(\boldsymbol{A})^{\perp} = \mathrm{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right)$,

2[扩张角度] $\mathrm{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right) \oplus \mathrm{Null}(\boldsymbol{A}) = \mathbb{R}^n, \mathrm{Col}(\boldsymbol{A}) \oplus \mathrm{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right) = \mathbb{R}^m$,

3[维数角度] $\dim \mathrm{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right) + \dim \mathrm{Null}(\boldsymbol{A}) = n, \dim \mathrm{Col}(\boldsymbol{A}) + \dim \mathrm{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right) = m$

矩阵的特征值与特征向量

- 定义
  - $A \in R^{n\times n}$, 矩阵方程 $Ax = \lambda x$ 的非零解称为矩阵关于特征值 λ 的特征向量。
  - 称 $|A - \lambda I_n| (= 0)$ 为特征多项式 (方程), $Nul(A - \lambda I_n)$ 为矩阵 A 关于特征值 λ 的特征空间
  - 特征值的集合称为矩阵的谱。模最大的特征值称为主特征值，其模称为矩阵的谱半径
  - 理解
    - 当矩阵A作用于其特征向量时，效果是将其进行系数为对应特征值的数乘
    - "一元n次方程有且只有n个根，考虑重数、虚数",而特征多项式是关于特征值的一元n次方程
- 性质
  - 不同特征值对应的特征向量线性无关
  - 同一特征值的所有特征向量线性相关，且与0向量一同构成R^n的线性子空间
  - k重特征值至多有k个线性无关的特征向量

矩阵 (f 为多项式)	$\boldsymbol{A}$	$f(\boldsymbol{A})$	A^{-1}	$A^\top$	$P^{-1}\boldsymbol{AP}$
特征值	λ	$f(\lambda)$	$\frac{1}{\lambda}$	λ	λ
对应的特征向量	$\boldsymbol{\xi}$	$\boldsymbol{\xi}$	ξ	单独计算	$P^{-1}\boldsymbol{\xi}$

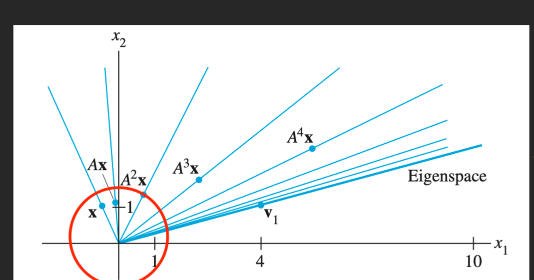
  - $A \in R^{n\times n}; \sum_{i=1}^{n} \lambda_i = Tr(A), \Pi(\lambda_i) = |A|$
  - $A \in R^{n\times n}$, 特征值集合 $\{\lambda_1, \ldots, \lambda_p\}$
    $\forall k \in [p], \lambda_k$ 几何重数 $\dim Nul(A - \lambda_k) \le \lambda_k$ 的代数重数
- 计算
  - 特征方程+线性方程组
    - 1.表示特征多项式,求解特征方程
    - 2.对特征方程的每个根,利用线性方程组求解对应的特征空间
  - 幂法、反幂法
    - 幂法
      - 作用 — 计算矩阵主特征值及其特征向量
      - 算法

Algorithm 1 幂法
Input: $A \in R^{n\times n}$ 可对角化, 主特征值的模严格大
Output: 主特征值 λ, 主特征向量 x
1: $k = 0$; 随机初始化单位向量 $\boldsymbol{x}^{(k)}$
2: repeat
3: $\quad \boldsymbol{y}^{(k+1)} = \boldsymbol{A}\boldsymbol{x}^{(k)}$
4: $\quad \boldsymbol{x}^{(k+1)} = \boldsymbol{y}^{(k+1)} / \left\|\boldsymbol{y}^{(k+1)}\right\|_2$
5: $\quad \lambda^{(k+1)} = \boldsymbol{x}^{(k+1)^T} \boldsymbol{A}\boldsymbol{x}^{(k+1)}$
6: $\quad k = k + 1$
7: until 收敛

Algorithm 2位移加速幂法
Input: $A \in R^{n\times n}$ 可对角化, 主特征值的模严格大, 位移 μ
Output: 主特征值 λ, 主特征向量 x
1: $k = 0$; 随机初始化单位向量 $\boldsymbol{x}^{(k)}$
2: repeat
3: $\quad \boldsymbol{y}^{(k+1)} = (\boldsymbol{A} - \boldsymbol{\mu I})x$
4: $\quad \boldsymbol{x}^{(k+1)} = \boldsymbol{y}^{(k+1)} / \left\|\boldsymbol{y}^{(k+1)}\right\|_2$
5: $\quad \lambda^{(k+1)} = \boldsymbol{x}^{(k+1)^T} \boldsymbol{A}\boldsymbol{x}^{(k+1)}$
6: $\quad k = k + 1$
7: until 收敛

Algorithm 2 瑞丽商加速的幂法 (瑞丽商作位移)
Input: $A \in R^{n\times n}$ 可对角化, 主特征值的模严格大
Output: 主特征值 λ, 主特征向量 x
1: $k = 0$; 随机初始化单位向量 $\boldsymbol{x}^{(k)}$
2: repeat
3: $\quad \lambda_k = \frac{x^{(k)^\top} A x^{(k)}}{x^{(k)^\top} x^{(k)}}$
4: $\quad \boldsymbol{y}^{(k+1)} = (\boldsymbol{A} - \boldsymbol{\lambda_k I})\boldsymbol{x}^{(k)}$
5: $\quad \boldsymbol{x}^{(k+1)} = \boldsymbol{y}^{(k+1)} / \left\|\boldsymbol{y}^{(k+1)}\right\|_2$
6: $\quad k = k + 1$
7: until 收敛

      - 算法原理



设 $A = P\Lambda P^{-1} \in R^{n\times n}, P = [\xi_1, \xi_2, \ldots, \xi_n] \in C^{n\times n}$
$\Lambda = [\lambda_1, \lambda_2, \ldots, \lambda_n] \in C^{n\times n}$, 且 $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n|$

$$\begin{aligned} x^{(k)} &= \boldsymbol{A}^k \boldsymbol{x}^{(0)} = \alpha_1 \boldsymbol{A}^k \boldsymbol{\xi}_1 + \alpha_2 \boldsymbol{A}^k \boldsymbol{\xi}_2 + \cdots + \alpha_n \boldsymbol{A}^k \boldsymbol{\xi}_n \\ &= \alpha_1 \lambda_1^k \boldsymbol{\xi}_1 + \alpha_2 \lambda_2^k \boldsymbol{\xi}_2 + \cdots + \alpha_n \lambda_n^k \boldsymbol{\xi}_n \\ &= \lambda_1^k \left(\alpha_1 \boldsymbol{\xi}_1 + \alpha_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k \boldsymbol{\xi}_2 + \cdots + \alpha_n \left(\frac{\lambda_n}{\lambda_1}\right)^k \boldsymbol{\xi}_n \right) \end{aligned}$$

$$\lim_{k\to\infty} \frac{x^{(k)}}{\lambda_1^k} = \alpha_1 \boldsymbol{\xi}_1 \qquad x^{(k)} \approx \lambda_1^k \alpha_1 \boldsymbol{\xi}_1$$

      - 理解

1. $\alpha_1 \ne 0$(即 $x^{(0)} \notin Span\{\xi_2, \xi_3, \ldots, \xi_n\}$), 否则理论上 $x^{(k)}$ 收敛至 0 。
   实际做法: (1) 随机选取 $x^{(0)}$, 若 $\alpha_1 = 0$, 则重新选择 $x^{(0)}$; (2) 依靠 $x^{(k)} (k > 0)$ 的 $\alpha_1 \ne 0$
2. 主特征值需要严格大, 否则算法不收敛
3. 幂法的收敛速度取决于 $|\frac{\lambda_2}{\lambda_1}|$, 通过移位 $(B = A - \alpha I)$ 降低该比值可加速幂法
4. 标准化是为了避免上溢或下溢, 以及方便从特征向量计算特征值
5. 瑞丽商加速的幂法具有局部二次收敛性质

$Algorithm2$ 位移加速幂法中, 位移 σ 的选取应满足:
(1)$\lambda_1 - \sigma$ 是 $A - \sigma I$ 的模最大的特征值→确保计算模最大特征值 λ_1
(2) $\max_{2\le i\le n} \left| \frac{\lambda_i - \sigma}{\lambda_1 - \sigma} \right|$ 尽可能地小→加速

    - 反幂法(幂法应用)
      - 作用 — 给定特征值估计，计算更精确的特征值及其特征向量
      - 算法

Algorithm 1反幂法
Input: A 可对角化, 待估特征值 μ
Output: μ 的更精确值 λ 及其特征向量 x
1: $k = 0$; 随机初始化单位向量 $\boldsymbol{x}^{(k)}$
2: repeat
3: $\quad (\boldsymbol{A} - \boldsymbol{\mu I})\boldsymbol{y}^{(k+1)} = \boldsymbol{x}^{(k)} \iff \boldsymbol{y}^{(k+1)} = (\boldsymbol{A} - \boldsymbol{\mu I})^{-1}\boldsymbol{x}^{(k)}$
4: $\quad \boldsymbol{x}^{(k+1)} = \boldsymbol{y}^{(k+1)} / \left\|\boldsymbol{y}^{(k+1)}\right\|_2$
5: $\quad \lambda^{(k+1)} = \boldsymbol{x}^{(k+1)^T} \boldsymbol{A}\boldsymbol{x}^{(k+1)}$
6: $\quad k = k + 1$
7: until 收敛

Algorithm 2 瑞丽商加速的反幂法
Input: $A \in R^{n\times n}$ 可对角化, 主特征值的模严格大
Output: 主特征值 λ, 主特征向量 x
1: $k = 0$; 随机初始化单位向量 $\boldsymbol{x}^{(k)}$
2: repeat
3: $\quad \lambda_k = \frac{x^{(k)^\top} A x^{(k)}}{x^{(k)^\top} x^{(k)}}$
4: $\quad (\boldsymbol{A} - \boldsymbol{\lambda_k I})\boldsymbol{y}^{(k+1)} = \boldsymbol{x}^{(k)}$
5: $\quad \boldsymbol{x}^{(k+1)} = \boldsymbol{y}^{(k+1)} / \left\|\boldsymbol{y}^{(k+1)}\right\|_2$
6: $\quad k = k + 1$
7: until 收敛

      - 理解

1. 反幂法通过构造新矩阵 $A - \mu I$ 使得其主特征值接近待估特征值 μ ，进而利用幂法计算该特征值
2. 反幂法的收敛速度取决于估计值 μ 与对应特征值的接近程度
3. 用方程组而不用求逆是因为 LU 分解可以加速该场景下的方程组计算 (系数矩阵固定)
4. 使用瑞丽商加速的反幂法收敛于哪一个特征值并不确定
5. 当 $\mu = 0$ 时, 计算的是 A 最小特征值

  - QR分解法

定理：（实 Schur 分解）
设 A 为 n 阶实矩阵，则存在正交矩阵 Q，使得

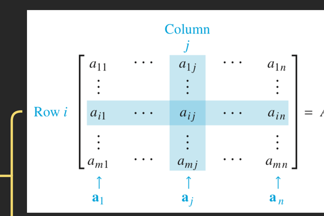
$$Q^T A Q = R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ & R_{22} & \cdots & R_{2m} \\ & & \ddots & \vdots \\ & & & R_{mm} \end{bmatrix} \quad \text{拟上三角矩阵}$$

其中 R_{ii} 是一阶或二阶方阵。

- 若 R_{ii} 是一阶方阵，则它就是 A 的特征值；
- 若 R_{ii} 是二阶方阵，则其特征值为 A 的两个共轭复特征值。

(1) 令 $A_1 = A$
(2) 对 $k = 1, 2, \ldots$,
计算 A_k 的 QR 分解 $A_k = Q_k R_k$
计算 $A_{k+1} = R_k Q_k$
直到 A_{k+1} 收敛到一个 拟上三角阵

定义



- 分块矩阵
  - 定义 — 同矩阵定义，只不过元素 a_{ij} 为矩阵
  - 计算
    - 块间、块内都要满足运算条件
    - 乘法的书写顺序
- a_{ij} 为 m 行 n 列的矩阵 $A = (a_{ij})_{m\times n}$ 的 (i, j) 元

运算

基本运算: 加法、数乘乘法、求幂、求逆

加法、数乘

- 定义/计算
  - $A, B \in R^{m\times n}, \lambda \in R$
  - $(A + B)_{ij} = a_{ij} + b_{ij}, \forall i \in [m] \forall j \in [n]$
  - $(\lambda A)_{ij} = \lambda a_{ij}, \forall i \in [m] \forall j \in [n]$
  - $A = B \iff \forall i \in [m] \forall j \in [n], a_{ij} = b_{ij}$
- 性质
  - $A, B \in R^{m\times n}, r, s \in R$
  - a. $A + B = B + A$ d. $r(A + B) = rA + rB$
  - b. $(A + B) + C = A + (B + C)$ e. $(r + s)A = rA + sA$
  - c. $A + 0 = A$ f. $r(sA) = (rs)A$

乘法、幂

- 定义/计算
  - $A \in R^{m\times p}, B \in R^{p\times n}$
  - $(AB)_{ij}^{m\times n} = \sum_{k=1}^{p} a_{ik} b_{kj}, \forall i \in [m] \forall j \in [n]$ — $A \in R^{n\times n}, k \in Z^+$; $A^k = k$ 个 A 相乘
  - 理解 — 矩阵乘法的定义源于线性映射的复合
  - 特例 — (列行展开)$AB = (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_p) \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_p \end{pmatrix} = \mathbf{a}_1\mathbf{b}_1 + \mathbf{a}_2\mathbf{b}_2 + \cdots + \mathbf{a}_p\mathbf{b}_p$
    其中 $\mathbf{a}_i\mathbf{b}_i, i \in [p]$ 是一个 $m \times n$ 矩阵.
- 性质
  - 设以下矩阵可乘、可加
  - a. $A(BC) = (AB)C$ b. $A(B + C) = AB + AC$
  - c. $(B + C)A = BA + CA$ d. $r(AB) = (rA)B = A(rB)$
  - e. $I_m A = A = A I_n$

逆

- 定义
  - $A \in R^{n\times n}$ 可逆/非奇异 $\iff \exists A^{-1} \in R^{n\times n}, A^{-1}A = I_n \vee AA^{-1} = I_n$
  - A^{-1} 称为 A 的唯一可逆矩阵
- 性质
  - $A, B \in R^{n\times n}$
  - 1.A 可逆，则 A^{-1} 可逆，且 $(A^{-1})^{-1} = A$
  - 2.A, B 可逆, 则 AB 可逆, 且 $(AB)^{-1} = B^{-1}A^{-1}$
  - 3.A 可逆, 则 $A^\top$ 可逆, 且 $(A^\top)^{-1} = (A^{-1})^\top$
  - 4.A 可逆, 则 $kA (k \ne 0 \in R)$ 可逆, 且 $(kA)^{-1} = \frac{1}{k}A^{-1}$
  - 5. 若对角矩阵 Λ 对角元均非 0(可逆), 则 $\Lambda^{-1} = (\frac{1}{\Lambda_i})_{i\in[n]}$
- 计算
  - 行变换法
    - 原理
      1. (若 A 可逆, 则)$A^{-1}A = I$
      2. (若 A 可逆, 则)$\exists E_n E_{n-1} \cdots E_1 A = I$ — 初等矩阵 E :初等行列变换 (互换, 倍乘, 倍加) 作用一次在单位阵 E 上；左行右列：对 $A \in R^{m\times n}$ 做初等行 / 列变换后的矩阵$= EA/AE$；其中, E 是相同的变换作用在单位阵得到的初等矩阵；初等矩阵的逆就是使 E 变为 I 的初等矩阵
      3. $E_n E_{n-1} \cdots E_1 I = A^{-1}$
    - 算法 — $[A \quad I] \xrightarrow{\text{初等行变换}} [I \quad A^{-1}][I \quad A^{-1}]$
  - 伴随矩阵法 — $A \in R^{n\times n}$, A 的伴随矩阵 $A^* := \begin{bmatrix} A_{11} & A_{21} & \ldots & A_{n1} \\ A_{12} & A_{22} & \ldots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \ldots & A_{nn} \end{bmatrix}$，且 $AA^* = A^*A = |A|I$
  - 待定系数法(矩阵方程法)
    - $A, X \in R^{n\times n}$
    - $AX = I \iff A[x_1 \quad x_2 \quad \cdots \quad x_n] = [e_1 \quad e_2 \quad \cdots \quad e_n] \iff Ax_i = e_i, \forall i \in [n]$
    - 这是一组系数矩阵相同而增广列不同的线性方程组, 可以一起进行行变换
- 应用
  - 等价命题
  - a. A是可逆矩阵 b. A与 I_n等价
  - c. A有 n个主元位 d. $A\mathbf{x} = \mathbf{0}$只有0解
  - e. A的列构成一个线性无关组 f. 线性变换 $\mathbf{x} \mapsto A\mathbf{x}$是单射
  - g. $\forall \mathbf{b} \in \mathbb{R}^n, A\mathbf{x} = \mathbf{b}$有唯一解 h. A的列张成$\mathbb{R}^n$
  - i. 线性变换 $\mathbf{x} \in \mathbb{R}^n \mapsto A\mathbf{x} \in \mathbb{R}^n$ 是满射 j. $\exists C \in R^{n\times n}, CA = I$
  - k. $\exists D \in R^{n\times n}, AD = I$ l. A^T是可逆矩阵
  - m. A的列构成$\mathbb{R}^n$的一个基 n. $\mathrm{Col}\, A = \mathbb{R}^n$
  - o. $\dim \mathrm{Col}\, A = n$ p. $\mathrm{rank}\, A = n$
  - q. $\mathrm{Nul}\, A = \{0\}$ r. $\dim \mathrm{Nul}\, A = 0$
  - s. 0不是 A的特征值 t. $|A| \ne 0$

高级运算：转置、行列式、迹

转置

- 定义/计算 — $A = (a_{ij})_{i\in[m], j\in[n]} \in R^{m\times n}, A^T := (a_{ji})_{j\in[n], i\in[m]} \in R^{n\times m}$
- 性质
  - 以下矩阵可乘可加
  - $(A^\top)^\top = A$ $(A + B)^\top = A^\top + B^\top$
  - $(rA)^\top = rA^\top, \forall r \in R$ $(AB)^\top = B^\top A^\top$

迹

- 定义/计算 — 方阵 $A \in R^{n\times n}$ 的迹 $\mathrm{Tr}(A) := \sum_{i=0}^{n} a_{ii}$
- 性质
  1. 线性: $\mathrm{Tr}(\alpha A + \beta B) = \alpha \mathrm{Tr}(A) + \beta \mathrm{Tr}(B), \forall \alpha, \beta \in R$
  2. 循环不变性: $\mathrm{Tr}(A_1 A_2 \cdots A_n) = \mathrm{Tr}(A_n A_1 \cdots A_{n-1}) = \cdots = \mathrm{Tr}(A_2 A_3 \cdots A_1)$, $A_1 A_2 \cdots A_n$ 为可乘矩阵序列
  3. $\mathrm{Tr}(A^T B) =< A, B >= \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ij} \quad A, B \in R^{n\times m}$
  4. $\mathrm{Tr}(A^T) = \mathrm{Tr}(A)$
  5. (微分运算和迹运算可交换), 即设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ ，则 $d\,\mathrm{Tr}(\boldsymbol{A}) = \mathrm{Tr}(d\boldsymbol{A})$

行列式

- 定义/计算

$$|\boldsymbol{A}| = \begin{cases} a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} = \sum_{j=1}^{n} a_{ij}A_{ij} (i \in [n]), (\sum_{j=1}^{n} a_{ij}A_{i'j} (i \ne i' \in [n]) = 0) \\ a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj} = \sum_{i=1}^{n} a_{ij}A_{ij} (j \in [n]), (\sum_{i=1}^{n} a_{ij}A_{ij'} (j \ne j' \in [n]) = 0). \end{cases}$$

其中, $A_{ij} = a_{ij}$ 的代数余子式$= (-1)^{i+j} \mid A[\{1, 2, \ldots, i-1, i+1, \ldots, n\}, \{1, 2, \ldots, j-1, j+1, \ldots, n\}]|, |a_{ij}| = a_{ij}$

- 性质
  - $A, B \in R^{n\times n}$
  - a. 行倍加后, $|A|$ 不变 b. 行互换后 $|A|$ 取反
  - c. $|AB| = |A||B|$ d. $|A^\top| = |A|, |A^{-1}| = \frac{1}{|A|}$
  - e. $|A|$ 关于某行 / 列具有线性性质 即 $|A| = f(\alpha_1, \ldots, \alpha_n)$ 关于 α_i 为线性函数
  - f. 行列式表示列 / 行向量 (起点为原点有向线段) 围成的 “有向”超平行多面体体积
  - g. 行列式表示表示 A 导出的坐标变换下 超平行多面体体积的伸缩因子