统计计算

2016 年 3 月 24 日

目录

目录

例

设

$$\mathbf{A}_1 = \begin{bmatrix} 2 & 6 \\ 2 & 6.00001 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 2 & 6 \\ 2 & 5.99999 \end{bmatrix},$$

则

$$\mathbf{A}_1^{-1} = \begin{bmatrix} 300000.5 & -300000 \\ -100000 & 100000 \end{bmatrix},$$

$$\mathbf{A}_2^{-1} = \begin{bmatrix} -299999.5 & 300000 \\ 100000 & -100000 \end{bmatrix}.$$

关于线性方程组解的计算问题 (续)

设

$$\mathbf{b}_1 = \begin{bmatrix} 8 \\ 8.00001 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} 8 \\ 8.00002 \end{bmatrix},$$

则

1. 线性方程组 $\mathbf{A}_1\mathbf{x} = \mathbf{b}_1$ 的解为 $\mathbf{x} = [1, 1]^{\mathrm{T}}$;
2. 线性方程组 $\mathbf{A}_1\mathbf{x} = \mathbf{b}_2$ 的解为 $\mathbf{x} = [-2, 2]^{\mathrm{T}}$;
3. 线性方程组 $\mathbf{A}_2\mathbf{x} = \mathbf{b}_1$ 的解为 $\mathbf{x} = [7, -1]^{\mathrm{T}}$;
4. 线性方程组 $\mathbf{A}_2\mathbf{x} = \mathbf{b}_2$ 的解为 $\mathbf{x} = [10, -2]^{\mathrm{T}}$.

关于线性方程组解的计算问题 (续)

从以上结果可以看出:

1. 虽然矩阵 $\mathbf{A}_1$ 经过很小的扰动变成 $\mathbf{A}_2$, 但是 $\mathbf{A}_1$ 在扰动前后的逆矩阵却产生了很大的偏差;
2. 线性方程组 $\mathbf{A}_1\mathbf{x} = \mathbf{b}_1$ 在系数矩阵 $\mathbf{A}_1$ 或向量 $\mathbf{b}_1$ 经过很小的扰动前后的解 $\mathbf{x}$ 也有很大的偏差.

所以, 非常有必要研究非奇异矩阵的逆以及线性方程组解的稳定性.

关于线性方程组解的计算问题 (续)

例

设

$$\mathbf{A}_1 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 10 & -10 \\ -1 & 1.001 \end{bmatrix},$$

$$\mathbf{b}_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \qquad \mathbf{b}_2 = \begin{bmatrix} 0 \\ 0.001 \end{bmatrix},$$

以及扰动矩阵

$$\Delta\mathbf{A} = \begin{bmatrix} 0 & 0 \\ 0 & 0.0005 \end{bmatrix}.$$

关于线性方程组解的计算问题 (续)

1. 线性方程组 $\mathbf{A}_1\mathbf{x}=\mathbf{b}_1$ 的解为 $\mathbf{x}=[1,1]^{\mathrm{T}}$, 而 $(\mathbf{A}_1+\Delta\mathbf{A})\mathbf{x}=\mathbf{b}_1$ 的解为 $\mathbf{x}=[2/2.0005,2/2.0005]^{\mathrm{T}}$, 这时系数矩阵的偏差对线性方程组解的影响不大;
2. 线性方程组 $\mathbf{A}_2\mathbf{x}=\mathbf{b}_2$ 的解也为 $\mathbf{x}=[1,1]^{\mathrm{T}}$, 但是 $(\mathbf{A}_2+\Delta\mathbf{A})\mathbf{x}=\mathbf{b}_2$ 的解却为 $\mathbf{x}=[2/3,2/3]^{\mathrm{T}}$, 这时系数矩阵的偏差对线性方程组解的影响是很大的.

可见, 相同的扰动对于不同的线性方程组可以导致完全不同精度的解. 这就有必要揭示系数矩阵本身的性质对解的影响.

关于线性方程组解的计算问题 (续)

为了研究线性方程组的解和最小二乘解的求解算法的精度以及算法的稳定性, 有必要了解向量和矩阵范数, 以及矩阵条件数方面的知识.

目录

定义

设 V 为数域 F ($\mathbb{C}$ 或 $\mathbb{R}$) 上的线性空间. 称函数 $\|\cdot\|: V \to \mathbb{R}$ 为 V 上的向量范数, 如果对任意 $\mathbf{x}, \mathbf{y} \in V$, 有

1. (非负性) $\|\mathbf{x}\| \geq 0$;
2. (正定性) $\|\mathbf{x}\| = 0$ 当且仅当 $\mathbf{x} = \mathbf{0}$.
3. (齐次性) $\|a\mathbf{x}\| = |a|\|\mathbf{x}\|$ 对所有纯量 $a \in F$ 均成立;
4. (三角不等式) $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

常用向量范数

例

由

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|, \quad \mathbf{x} = [x_1, \cdots, x_n]^{\mathrm{T}}$$

定义的 $\|\cdot\|_1$ 是 $\mathbb{C}^n$ 上的向量范数，称为 1-范数或 l_1 范数.

例

由

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}, \quad \mathbf{x} = [x_1, \cdots, x_n]^{\mathrm{T}}$$

定义的 $\|\cdot\|_2$ 是 $\mathbb{C}^n$ 上的向量范数, 称为 2-范数或 l_2 范数, 也称为 Euclid 范数. 容易看到, Euclid 范数是酉不变的.

例

由

$$\|\mathbf{x}\|_\infty = \max_{1\le i\le n} |x_i|, \quad \mathbf{x} = [x_1, \cdots, x_n]^{\mathrm{T}}$$

定义的 $\|\cdot\|_\infty$ 是 $\mathbb{C}^n$ 上的向量范数, 称为 ∞-范数或 l_∞ 范数.

例

设 $1 \le p < \infty$, 则由

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}, \quad \mathbf{x} = [x_1, \cdots, x_n]^{\mathrm{T}}$$

定义的 $\|\cdot\|_p$ 是 $\mathbb{C}^n$ 上的向量范数, 称为 p-范数或 l_p 范数.

矩阵范数

定义

称函数 $\|\cdot\|:\mathbb{C}^{m\times n}\to\mathbb{R}$ 为 $\mathbb{C}^{m\times n}$ 上的矩阵范数, 如果对任意 $\mathbf{A},\mathbf{B}\in\mathbb{C}^{m\times n}$, 有

1. (非负性) $\|\mathbf{A}\|\geq 0$;
2. (正定性) $\|\mathbf{A}\|=0$ 当且仅当 $\mathbf{A}=\mathbf{O}$;
3. (齐次性) $\|a\mathbf{A}\|=|a|\|\mathbf{A}\|$ 对所有纯量 $a\in\mathbb{C}$ 均成立;
4. (三角不等式) $\|\mathbf{A}+\mathbf{B}\|\leq\|\mathbf{A}\|+\|\mathbf{B}\|$.

例

由

$$\|\mathbf{A}\|_{m_1} = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|, \quad \mathbf{A} = (a_{ij}) \in \mathbb{C}^{m\times n}$$

定义的 $\|\cdot\|_{m_1}$ 是 $\mathbb{C}^{m\times n}$ 上的矩阵范数, 称为 l_1 范数.

常用的矩阵范数 (续)

例

由

$$\|\mathbf{A}\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}, \quad \mathbf{A} = (a_{ij}) \in \mathbb{C}^{m\times n}$$

定义的 $\|\cdot\|_F$ 是 $\mathbb{C}^{m\times n}$ 上的矩阵范数, 称为 l_2 范数, Euclid 范数或 Frobenius 范数.

常用的矩阵范数 (续)

例

由

$$\|\mathbf{A}\|_{m_\infty} = \max_{\substack{1\le i\le m\\ 1\le j\le n}} |a_{ij}|, \quad \mathbf{A} = (a_{ij}) \in \mathbb{C}^{m\times n}$$

定义的 $\|\cdot\|_{m_\infty}$ 是 $\mathbb{C}^{m\times n}$ 上的广义矩阵范数，称为 l_∞ 范数.

例

若 $\|\cdot\|$ 为 $\mathbb{C}^m$ 和 $\mathbb{C}^n$ 上的同类向量范数, 则由

$$\|\mathbf{A}\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\| = \max_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|}, \quad \mathbf{A} \in \mathbb{C}^{m\times n}$$

定义的函数 $\|\cdot\|$ 是 $\mathbb{C}^{m\times n}$ 上的矩阵范数, 称为由向量范数诱导的矩阵范数.

常用的矩阵范数 (续)

定理

由向量范数 $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$ 诱导的矩阵范数分别为:

1. $\|\mathbf{A}\|_1 = \max\limits_{1\le j\le n} \sum\limits_{i=1}^{m} |a_{ij}|$;
2. $\|\mathbf{A}\|_2 = \sqrt{\lambda_1}$, λ_1 为 $\mathbf{A}^{\mathrm{H}}\mathbf{A}$ 的最大特征值;
3. $\|\mathbf{A}\|_\infty = \max\limits_{1\le i\le m} \sum\limits_{j=1}^{n} |a_{ij}|$,

其中 $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{m\times n}$.
称矩阵范数 $\|\cdot\|_1, \|\cdot\|_2$ 及 $\|\cdot\|_\infty$ 分别为列和范数, 谱范数及行和范数.

目录

线性方程组解的扰动

下面讨论线性方程组 $\mathbf{Ax}=\mathbf{b}$ 的最小二乘解在 $\mathbf{b}$ 有一个小的扰动 $\Delta\mathbf{b}$ 时线性方程组

$$\mathbf{A}(\mathbf{x}+\Delta\mathbf{x})=\mathbf{b}+\Delta\mathbf{b}$$

的解 $\mathbf{x}+\Delta\mathbf{x}$ 的变化，其中 $\Delta\mathbf{x}=\mathbf{A}^{\dagger}\Delta\mathbf{b}$.

线性方程组解的扰动 (续)

定理

设 $\mathbf{A} \in \mathbb{C}^{n\times m}$ 满列秩且 $\mathbf{b}_1 = \mathbf{A}\mathbf{A}^\dagger \mathbf{b} \neq \mathbf{0}$, 则有

$$\frac{\|\mathbf{A}^\dagger \Delta\mathbf{b}\|}{\|\mathbf{A}^\dagger \mathbf{b}\|} \leq \|\mathbf{A}^\dagger\| \|\mathbf{A}\| \frac{\|\Delta\mathbf{b}_1\|}{\|\mathbf{b}_1\|}$$

其中 $\Delta\mathbf{b}_1 = \mathbf{A}\mathbf{A}^\dagger \Delta\mathbf{b}$.

证明　由于

$$\|\mathbf{A}^{\dagger}\Delta\mathbf{b}\| = \|\mathbf{A}^{\dagger}\mathbf{A}\mathbf{A}^{\dagger}\Delta\mathbf{b}\| = \|\mathbf{A}^{\dagger}\Delta\mathbf{b}_1\| \le \|\mathbf{A}^{\dagger}\|\,\|\Delta\mathbf{b}_1\|$$

$$\|\mathbf{b}_1\| = \|\mathbf{A}\mathbf{A}^{\dagger}\mathbf{b}\| \le \|\mathbf{A}\|\,\|\mathbf{A}^{\dagger}\mathbf{b}\|,$$

所以

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{\|\mathbf{A}^{\dagger}\Delta\mathbf{b}\|}{\|\mathbf{A}^{\dagger}\mathbf{b}\|} \le \|\mathbf{A}^{\dagger}\|\,\|\mathbf{A}\|\frac{\|\Delta\mathbf{b}_1\|}{\|\mathbf{b}_1\|}. \qquad \square$$

记 $\kappa = \|\mathbf{A}^{\dagger}\|\|\mathbf{A}\|$. 上述定理的解释为:

1. 解的相对误差的一个上界是 $\kappa\frac{\|\Delta\mathbf{b}_1\|}{\|\mathbf{b}_1\|}$. 当 $\mathbf{b}$ 的扰动固定时, 这个上界的大小就取决于 κ;
2. κ 越小, 解的相对误差的变化范围就越小;
3. 反之, κ 越大, 解的相对误差的变化范围就越大;
4. 因此, κ 可以用来衡量一个方程的解是否对扰动敏感.

引理

设 $\|\cdot\|$ 为 $\mathbb{C}^{n\times n}$ 上的矩阵范数, 且 $\|\mathbf{I}\|=1$. 若 $\mathbf{A}\in\mathbb{C}^{n\times n}$ 为非奇异矩阵, $\mathbf{E}\in\mathbb{C}^{n\times n}$ 满足 $\|\mathbf{A}^{-1}\|\|\mathbf{E}\|<1$, 则 $\mathbf{B}=\mathbf{A}+\mathbf{E}$ 为非奇异矩阵, 且

$$\|\mathbf{B}^{-1}\|\le\frac{1}{\gamma}\|\mathbf{A}^{-1}\|,\qquad \frac{\|\mathbf{B}^{-1}-\mathbf{A}^{-1}\|}{\|\mathbf{A}^{-1}\|}\le\frac{\kappa}{\gamma}\frac{\|\mathbf{E}\|}{\|\mathbf{A}\|},$$

其中

$$\kappa=\|\mathbf{A}\|\|\mathbf{A}^{-1}\|,\qquad \gamma=1-\kappa\frac{\|\mathbf{E}\|}{\|\mathbf{A}\|}.$$

定理

设 $\mathbf{A} \in \mathbb{C}^{n\times n}$ 非奇异, $\mathbf{b} \in \mathbb{C}^n$ 非零, $\mathbf{x}$ 为方程 $\mathbf{Ax} = \mathbf{b}$ 的解. 若 $\|\cdot\|$ 为 $\mathbb{C}^{n\times n}$ 上的矩阵范数且满足 $\|\mathbf{I}\| = 1, \|\mathbf{A}^{-1}\|\|\Delta\mathbf{A}\| < 1$ $(\Delta\mathbf{A} \in \mathbb{C}^{n\times n})$, 则方程 $(\mathbf{A} + \Delta\mathbf{A})(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} + \Delta\mathbf{b}$ 有唯一解 $\mathbf{x} + \Delta\mathbf{x}$, 且

$$\|\Delta\mathbf{x}\| \le \frac{\|\mathbf{A}^{-1}\|}{1 - \|\mathbf{A}^{-1}\|\|\Delta\mathbf{A}\|}(\|\Delta\mathbf{b}\| + \|\Delta\mathbf{A}\|\|\mathbf{x}\|)$$

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \le \frac{\kappa}{1 - \kappa\frac{\|\Delta\mathbf{A}\|}{\|\mathbf{A}\|}}\left(\frac{\|\Delta\mathbf{A}\|}{\|\mathbf{A}\|} + \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}\right).$$

证明　　见教材第 277 页.

线性方程组解的扰动 (续)

1. 上一定理给出了线性方程组 $\mathbf{Ax}=\mathbf{b}$ 解的绝对误差的估计式. 可以看出: $\|\mathbf{A}^{-1}\|$ 越大, 绝对误差 $\|\Delta\mathbf{x}\|$ 也越大;
2. 一般更关心线性方程组解的相对误差. κ 反映了方程组的解 $\mathbf{x}$ 的相对误差对于 $\mathbf{A}$ 和 $\mathbf{b}$ 的相对误差的依赖程度. 可以看出, 在 $\|\Delta\mathbf{A}\|/\|\mathbf{A}\|$ 和 $\|\Delta\mathbf{b}\|/\|\mathbf{b}\|$ 均相同的条件下, 若 κ 越大, 则方程组解的相对误差也越大.

定义

设 $\mathbf{A} \in \mathbb{C}^{n\times m}$, 称 $\|\mathbf{A}^{\dagger}\|\|\mathbf{A}\|$ 为矩阵 $\mathbf{A}$ 的条件数, 记为 $\operatorname{cond}(\mathbf{A})$.

需要注意的是：条件数与所选择的矩阵范数有关, 当需要强调该范数时, 通常加下标注明, 例如 $\operatorname{cond}_2(\mathbf{A}) = \|\mathbf{A}\|_2\|\mathbf{A}^{-1}\|_2$.

条件数的性质

设 $\mathbf{A} \in \mathbb{C}^{n\times n}$ 为非奇异矩阵.

1. $\mathrm{cond}(\alpha \mathbf{A}) = \mathrm{cond}(\mathbf{A}) = \mathrm{cond}(\mathbf{A}^{-1})$, 其中 α 为任意非零常数;
2. $\mathrm{cond}_p(\mathbf{A}) \geq 1$, 其中 $\|\cdot\|_p$ 为向量 p-范数诱导的矩阵范数 $(p \geq 1)$;
3. 对 2-条件数, 有
$$\mathrm{cond}_2(\mathbf{A}) = \left(\frac{\lambda_{\max}(\mathbf{A}^{\mathrm{H}}\mathbf{A})}{\lambda_{\min}(\mathbf{A}^{\mathrm{H}}\mathbf{A})}\right)^{1/2}.$$
特别地, 若 $\mathbf{A}$ 为 Hermite 矩阵, 则
$$\mathrm{cond}_2(\mathbf{A}) = \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})};$$
4. $\mathrm{cond}_2^2(\mathbf{A}) = \mathrm{cond}_2(\mathbf{A}^{\mathrm{H}}\mathbf{A})$.

当条件数 $\mathrm{cond}(\mathbf{A})$ 的值大时，我们称 $\mathbf{A}$ 为病态的；反之，将条件数相对小的矩阵称为良态的.

对于前述线性方程组解两个例子中：

1. 例 1: $\mathrm{cond}_2(\mathbf{A}_1) \approx 4.0 \times 10^6$ 很大，所以 $\mathbf{A}$ 和 $\mathbf{b}$ 的很小扰动可以导致 $\mathbf{A}$ 的逆和线性方程组 $\mathbf{Ax} = \mathbf{b}$ 的解出现很大的偏差.
2. 例 2: $\mathrm{cond}_2(\mathbf{A}_1 + \Delta\mathbf{A}) \approx 1$ 很小，而 $\mathrm{cond}_2(\mathbf{A}_2 + \Delta\mathbf{A}) \approx 13467$ 很大，这就说明了相同的偏差 $\Delta\mathbf{A}$ 为什么对不同线性方程组的解有很不相同的影响.

矩阵条件数与线性方程组的解 (续)

在实际应用中求矩阵的逆或者解线性方程组时，往往需要考虑进行某些线性变换，降低矩阵的条件数，以便提高计算精度.

值得注意的是矩阵行列式不能刻划矩阵的条件数. 详见教材第 276 页的例.

n 阶 Hilbert 矩阵定义为

$$\mathbf{H}_n = \left(\frac{1}{i+j-1}\right)_{i,j=1,\cdots,n}$$
$$= \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n+1} \\ \vdots & \vdots & & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \frac{1}{2n-1} \end{bmatrix}.$$

Hilbert 矩阵 (续)

对 Hilbert 矩阵感兴趣主要是因为它是严重病态的矩阵, 即 $\mathbf{H}_n$ 的条件数很大. 下面是几个 n 值所对应 Hilbert 矩阵的条件数.

n	$\mathrm{cond}_2(\mathbf{H}_n)$	n	$\mathrm{cond}_2(\mathbf{H}_n)$
3	5.24×10^2	4	1.55×10^4
5	4.77×10^5	6	1.50×10^7
7	4.75×10^8	8	1.53×10^{10}
9	4.93×10^{11}	10	1.60×10^{13}

可以看出 Hilbert 矩阵的条件数随着其阶数 n 的增大而快速增大. Hilbert 矩阵的这种严重病态性, 常常用来检验与矩阵相关的算法稳定性.

算法对误差的影响

见教材第 278 页.

问　题?

目录

目录

1. 矩阵的广义逆及其计算
    - 线性方程组解的稳定性
    - 向量和矩阵范数与条件数
    - 线性方程组解的扰动
2. 消去变换
    - 引言
    - 消去变换的定义
    - 消去变换的基本性质
    - 消去变换的应用

引言

见教材第 278 页.

目录

例

用消去法求矩阵

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

的逆矩阵.

解: 构造矩阵 $[\mathbf{A} \quad \mathbf{I}]$. 用消去法求 $\mathbf{A}^{-1}$ 的过程就是对矩阵 $[\mathbf{A} \quad \mathbf{I}]$ 实行一系列初等行变换, 使其中的子阵 $\mathbf{A}$ 通过变换变为 $\mathbf{I}$, 当这一过程结束时, 原来的 $\mathbf{I}$ 就变成 $\mathbf{A}^{-1}$.

记 $\mathbf{e}_i$ 为第 i 个元素为 1, 其余元素为 0 的列向量.

消去变换的定义 (续)

求 $\mathbf{A}^{-1}$ 的具体过程为

1. 将第 1 列变为 $\mathbf{e}_1$.

$$[\mathbf{A} \quad \mathbf{I}] = \begin{bmatrix} 2 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{3}{2} & -\frac{1}{2} & 0 & 1 \end{bmatrix}$$

2. 将第 2 列变为 $\mathbf{e}_2$. (已完成)
3. 将第 3 列变为 $\mathbf{e}_3$.

$$\to \begin{bmatrix} 1 & 0 & 0 & \frac{2}{3} & 0 & -\frac{1}{3} \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -\frac{1}{3} & 0 & \frac{2}{3} \end{bmatrix}$$

消去变换的定义 (续)

上述计算过程有以下特点:

1. 每一步工作只是用初等行变换将矩阵的某一列变成单位向量;
2. 将某列变为 $\mathbf{e}_i$ 所作的所有变换不影响矩阵中其它的单位向量 $\mathbf{e}_j (j \neq i)$.
   例如: 第 1 步的变换不影响第 2, 5, 6 列; 第 3 步的变换不影响第 1, 2, 5 列.

消去变换的定义 (续)

上述计算过程有以下特点:

1. 每一步工作只是用初等行变换将矩阵的某一列变成单位向量;
2. 将某列变为 $\mathbf{e}_i$ 所作的所有变换不影响矩阵中其它的单位向量 $\mathbf{e}_j (j \neq i)$.
   例如: 第 1 步的变换不影响第 2, 5, 6 列; 第 3 步的变换不影响第 1, 2, 5 列.

消去变换的定义 (续)

变换的上述两个特点可以使我们将上述计算过程中的矩阵形式写得更紧凑些.

对于上述例子而言：第 1 步的变换是将第 1 列变为 $\mathbf{e}_1$，第 5, 6 列不变，于是不必写出这两列；第 1 列变为 $\mathbf{e}_1$ 之后就不会再变了，所以可以将已变换的第 4 列替换第 1 列而不保留 $\mathbf{e}_1$，即用该初等行变换对 $\mathbf{e}_1$ 作用的结果替换 $\mathbf{A}$ 中的第 1 列.

于是第 1 步可写成

$$\mathbf{A} \to \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{3}{2} \end{bmatrix}$$

消去变换的定义 (续)

同样地, 第 2, 3 步也可写成类似的紧凑形式. 整个计算过程可写成

$$
\mathbf{A} \to \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{3}{2} \end{bmatrix} \to \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{3}{2} \end{bmatrix}
$$

$$
\to \begin{bmatrix} \frac{2}{3} & 0 & -\frac{1}{3} \\ 0 & 1 & 0 \\ -\frac{1}{3} & 0 & \frac{2}{3} \end{bmatrix} = \mathbf{A}^{-1}.
$$

上述每一步所作的紧凑形式的变换就称为消去变换.

消去变换的定义 (续)

称 $\mathbf{T}_{ij}$ 是对 $\mathbf{A}$ 实行的 (i,j) 消去变换, 如果

1. 对 $\mathbf{A}$ 做初等行变换, 使其第 j 列变为 $\mathbf{e}_i$;
2. 将上面初等行变换对 $\mathbf{e}_i$ 的作用结果替换当前矩阵的第 j 列.

消去变换的定义 (续)

教材第 281 页定义 6.2.

1. G-J 消去变换
2. 消去变换

两个变换之间的关系:

$$\mathbf{T}_{ii}\mathbf{A} = \mathbf{G}_{ii}\begin{bmatrix}\mathbf{A} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{I}\\ \mathbf{I}\end{bmatrix} - \mathbf{I}_n$$
$$= \mathbf{G}_{ii}\mathbf{A} + \mathbf{G}_{ii} - \mathbf{I}_n.$$

设 $\mathbf{A}_{n\times m}=(a_{\alpha\beta})_{n\times m}$. $\mathbf{T}_{ij}$ 对 $\mathbf{A}$ 的变换结果记为 $\mathbf{T}_{ij}\mathbf{A}=(a^*_{\alpha\beta})_{n\times m}$. 构造矩阵 $[\mathbf{A}\quad \mathbf{e}_i]$, 用初等行变换将 $\mathbf{A}$ 的第 j 列变为 $\mathbf{e}_i$ 的过程为

$$
[\mathbf{A}\quad \mathbf{e}_i]
$$

$$
\xrightarrow{\text{第 } i \text{ 行除以 } a_{ij}}
\begin{bmatrix}
* & \cdots & a_{1j} & \cdots & * & 0 \\
\cdots & & \cdots & & \cdots & \cdots \\
\frac{a_{i1}}{a_{ij}} & \cdots & 1 & \cdots & \frac{a_{im}}{a_{ij}} & \frac{1}{a_{ij}} \\
\cdots & & \cdots & & \cdots & \cdots \\
* & \cdots & a_{nj} & \cdots & * & 0
\end{bmatrix}
$$

$$
\xrightarrow[\text{的 } a_{kj} \text{ 倍 } (k\neq i)]{\text{第 } k \text{ 行减第 } i \text{ 行}}
\begin{bmatrix}
a_{11}-\frac{a_{i1}a_{1j}}{a_{ij}} & \cdots & 0 & \cdots & a_{1m}-\frac{a_{im}a_{1j}}{a_{ij}} & -\frac{a_{1j}}{a_{ij}} \\
\cdots & & \cdots & & \cdots & \cdots \\
\frac{a_{i1}}{a_{ij}} & \cdots & 1 & \cdots & \frac{a_{im}}{a_{ij}} & \frac{1}{a_{ij}} \\
\cdots & & \cdots & & \cdots & \cdots \\
a_{n1}-\frac{a_{i1}a_{nj}}{a_{ij}} & \cdots & 0 & \cdots & a_{nm}-\frac{a_{im}a_{nj}}{a_{ij}} & -\frac{a_{nj}}{a_{ij}}
\end{bmatrix}
$$

消去变换的算法 (续)

再由定义中的第 2 步将最后的矩阵中最后一列替换到第 j 列, 则知此前 m 列组成的子阵就是 $\mathbf{T}_{ij}\mathbf{A}$.

矩阵 $\mathbf{T}_{ij}\mathbf{A}$ 的元素只有四种情形.

消去变换的算法 (续)

对任意 $\alpha = 1, \cdots, n; \beta = 1, \cdots, m$, 有:

$$a^*_{\alpha\beta} = \begin{cases} \dfrac{1}{a_{ij}} & \text{当 } \alpha = i, \beta = j \text{ 时} \\ -\dfrac{a_{\alpha j}}{a_{ij}} & \text{当 } \alpha \neq i, \beta = j \text{ 时} \\ \dfrac{a_{i\beta}}{a_{ij}} & \text{当 } \alpha = i, \beta \neq j \text{ 时} \\ a_{\alpha\beta} - \dfrac{a_{i\beta} a_{\alpha j}}{a_{ij}} & \text{当 } \alpha \neq i, \beta \neq j \text{ 时} \end{cases}$$

注意: 当 $a_{ij} = 0$ 时 $\mathbf{T}_{ij}$ 无定义.

问　题?

目录

消去变换的基本性质

1. 反身性:
   $\mathbf{T}_{ij}\mathbf{T}_{ij}\mathbf{A} = \mathbf{A}$;
2. 可交换性:
   若 $i \neq k, j \neq l$, 则 $\mathbf{T}_{ij}\mathbf{T}_{kl}\mathbf{A} = \mathbf{T}_{kl}\mathbf{T}_{ij}\mathbf{A}$;
3. 若 $\mathbf{A}$ 为对称阵, $\mathbf{B} = \mathbf{T}_{kk}\mathbf{A} = (b_{\alpha\beta})$, 则

$$\begin{aligned} b_{k\beta} &= -b_{\beta k}, & \beta &\neq k, \\ b_{\alpha\beta} &= b_{\beta\alpha}, & \alpha &\neq k,\ \beta \neq k. \end{aligned}$$

4. 行列置换与消去变换的关系:
设 $\mathbf{P}_{ip}$ 为第 i 行和 p 行交换的行置换矩阵,
$\mathbf{Q}_{jq}$ 为第 j 列和 q 列交换的列置换矩阵, 则
    1. $\mathbf{T}_{ij}(\mathbf{A}\mathbf{Q}_{jq}) = (\mathbf{T}_{iq}\mathbf{A})\mathbf{Q}_{jq}$;
    2. $\mathbf{T}_{ij}(\mathbf{P}_{ip}\mathbf{A}) = \mathbf{P}_{ip}(\mathbf{T}_{pj}\mathbf{A})$;
    3. $\mathbf{T}_{ij}(\mathbf{P}_{ip}\mathbf{A}\mathbf{Q}_{jq}) = \mathbf{P}_{ip}(\mathbf{T}_{pq}\mathbf{A})\mathbf{Q}_{jq}$.

消去变换的基本性质 (续)

5. 设

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix},$$

且 $\mathbf{A}_{11}$ 为 r 阶可逆矩阵, 则

$$\mathbf{T}_{rr} \cdots \mathbf{T}_{11}\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix}.$$

6. 若 $\mathbf{A} \in \mathbb{R}^{n\times n}$ 非奇异, 且每次均能进行消去变换 $\mathbf{T}_{ii}(i=1,\cdots,n)$, 简记为 $\mathbf{T}_i$, 则

$$\mathbf{A}^{-1} = \mathbf{T}_n \cdots \mathbf{T}_1 \mathbf{A}.$$

特别地, 若 $\mathbf{A}$ 为对称正定矩阵, 则消去变换 $\mathbf{T}_i(i=1,\cdots,n)$ 的主元均非 0, 从而

$$\mathbf{A}^{-1} = \mathbf{T}_n \cdots \mathbf{T}_1 \mathbf{A}.$$

目录

计算非奇异矩阵的逆

若 $\mathbf{A}$ 的主元非零, 则

$$\mathbf{A}^{-1} = \mathbf{T}_n \cdots \mathbf{T}_1 \mathbf{A}.$$

若 $\mathbf{A}$ 的主元为零.

1. 置换法

先将 $\mathbf{A}$ 的行和列分别做置换, 使其置换后的 $\mathbf{PAQ}$ 的主元均非零. 记 $\mathbf{A}$ 的 (i_p, j_q) 元素对应于 $\mathbf{PAQ}$ 的 (i,j) 元素, 于是

$$\begin{aligned} \mathbf{A}^{-1} &= \mathbf{Q}(\mathbf{PAQ})^{-1}\mathbf{P} \\ &= \mathbf{QT}_n \cdots \mathbf{T}_1(\mathbf{PAQ})\mathbf{P} \\ &= \mathbf{QP}(\mathbf{T}_{n_p n_q} \cdots \mathbf{T}_{1_p 1_q})\mathbf{AQP}. \end{aligned}$$

例: 教材第 284 页例 6.2.

例 6.2 进一步的分析. 见教材第 284 页最后一段至第 285 页.

计算非奇异矩阵的逆 (续)

2. 选主元法

   设 $\mathbf{A} = (a_{ij}), \mathbf{B} = (b_{ij}), C = \{1, \cdots, n\}$.

   - 对 $i = 1, 2, \cdots, n$ 做

     取 $a_{ij} = \max_{k \in C} |a_{ik}|$;

     令 $L(i) = j, C = C - \{j\}$;

     计算 $\mathbf{A} = \mathbf{T}_{ij}\mathbf{A}$.
   - 对 $i = 1, 2, \cdots, n$; $j = 1, 2, \cdots, n$, 令 $b_{L(i)j} = a_{iL(j)}$.

注: 在第 1 步中, 将各主元的列号记录下来, 以便在第 2 步中进行行列置换, 输出时 $\mathbf{B}$ 中存放的就是 $\mathbf{A}^{-1}$.

计算非奇异矩阵的逆 (续)

例: 求矩阵 $\mathbf{A} = \begin{bmatrix} 0 & 4 \\ 2 & 0 \end{bmatrix}$ 的逆.

解: 第 1 步, 先按行选主元并作消去变换

$$\begin{bmatrix} 0 & 4 \\ 2 & 0 \end{bmatrix} \xrightarrow[L(1)=2]{\mathbf{T}_{12}} \begin{bmatrix} 0 & 1/4 \\ 2 & 0 \end{bmatrix} \xrightarrow[L(2)=1]{\mathbf{T}_{21}} \begin{bmatrix} 0 & 1/4 \\ 1/2 & 0 \end{bmatrix} = (a_{ij}).$$

再由第 2 步立即得到 $\mathbf{A}^{-1} = (b_{ij}) = \begin{bmatrix} 0 & 1/2 \\ 1/4 & 0 \end{bmatrix}$,

其中

$$\begin{aligned} b_{11} &= b_{L(2)1} = a_{2L(1)} = a_{22} = 0, \\ b_{12} &= b_{L(2)2} = a_{2L(2)} = a_{21} = 1/2, \\ b_{21} &= b_{L(1)1} = a_{1L(1)} = a_{12} = 1/4, \\ b_{22} &= b_{L(1)2} = a_{1L(2)} = a_{11} = 0. \end{aligned}$$

问　题?