第二章：范数理论

§2. 1　向量范数

§2. 2　矩阵范数

§2. 3　范数应用举例

范数应用举例

- 矩阵的谱半径

定义2. 6： 设$A \in \mathbb{C}^{n\times n}$，$\lambda_1, \lambda_2, \cdots, \lambda_n$是$A$的特征值，则称

$$\rho(A) = \max_{1\le i\le n} |\lambda_i|$$

为矩阵A的谱半径.

定理2. 12： 设$A \in \mathbb{C}^{n\times n}$. 则

(1) $\rho(A^k) = [\rho(A)]^k$;

(2) $\rho(A^H A) = \rho(AA^H) = \|A\|_2^2$

(3) 当A是正规矩阵时，$\rho(A) = \|A\|_2$.

谱半径与范数的关系

定理2. 13： $\rho(A) \le \|A\|$, 其中$\|A\|$是A的任一矩阵范数.

证明：

设$A \in \mathbb{C}^{n\times n}$, λ是A的一个特征值, x是A属于λ的特征向量,

则　$Ax = \lambda x$.

对$\mathbb{C}^{n\times n}$中任一矩阵范数$\|\cdot\|$以及与它相容的向量范数$\|\cdot\|_v$，

我们有

$$|\lambda| \cdot \|x\|_v = \|\lambda x\|_v = \|Ax\|_v \le \|A\| \cdot \|x\|_v$$

从而　$|\lambda| \le \|A\|$, 即$\rho(A) \le \|A\|$.

Remark. 正规矩阵的谱范数$\|\cdot\|_2$是最小的矩阵范数.

谱半径与范数的关系

例 2.11　已知 $A = \begin{pmatrix} 0 & 0.2 & 0.1 \\ -0.2 & 0 & 0.2 \\ -0.1 & -0.2 & 0 \end{pmatrix}$，试估计 A 的谱半径.

解　可求得 $\|A\|_1 = \|A\|_\infty = 0.4$，$\|A\|_{m_1} = 1$，$\|A\|_{m_\infty} = 0.6$，$\|A\|_F = \sqrt{0.18} \approx 0.4243$. 于是 $\rho(A) \leqslant 0.4$.

实际计算可知 A 的特征值为 $0, -0.3i, 0.3i$，从而 $\rho(A) = 0.3$. 可见对此矩阵谱半径的估计较精确. 但对多数矩阵来说，估计的结果偏保守.

谱半径与范数的关系

- **定理2. 14：** 设$A \in \mathbb{C}^{n\times n}$, 则 $\forall \varepsilon > 0$,
  必存在$\mathbb{C}^{n\times n}$上矩阵范数$\|\cdot\|_m$使得　$\|A\|_m \le \rho(A) + \varepsilon$.

证明. 设$\lambda_1, \lambda_2, \cdots, \lambda_n$是$A$的特征值. 由Jordan分解定理，存在可逆矩阵$P$使得

$$P^{-1}AP = J = \begin{pmatrix} \lambda_1 & \delta_1 & & \\ & \lambda_2 & \ddots & \\ & & \ddots & \delta_{n-1} \\ & & & \lambda_n \end{pmatrix}, \text{其中}\delta_i\text{=0 or 1.}$$

谱半径与范数的关系

$\forall \varepsilon > 0$, 令

$$D = \begin{pmatrix} \varepsilon & & & \\ & \varepsilon^2 & & \\ & & \ddots & \\ & & & \varepsilon^n \end{pmatrix},$$

$$D^{-1}P^{-1}APD = \begin{pmatrix} \varepsilon^{-1} & & & \\ & \varepsilon^{-2} & & \\ & & \ddots & \\ & & & \varepsilon^{-n} \end{pmatrix} \begin{pmatrix} \lambda_1 & \delta_1 & & \\ & \lambda_2 & \ddots & \\ & & \ddots & \delta_{n-1} \\ & & & \lambda_n \end{pmatrix} \begin{pmatrix} \varepsilon & & & \\ & \varepsilon^2 & & \\ & & \ddots & \\ & & & \varepsilon^n \end{pmatrix}$$

谱半径与范数的关系

$$=\begin{pmatrix}\lambda_1 & \varepsilon\delta_1 & & \\ & \lambda_2 & \ddots & \\ & & \ddots & \varepsilon\delta_{n-1} \\ & & & \lambda_n\end{pmatrix},\text{ 于是 } \left\|D^{-1}P^{-1}APD\right\|_\infty \le \rho(A)+\varepsilon.$$

定义 $\|A\|_m = \left\|D^{-1}P^{-1}APD\right\|_\infty$，可以验证$\|\cdot\|_m$是$C^{n\times n}$上一个矩阵范数，满足$\|A\|_m \le \rho(A)+\varepsilon.$

矩阵的条件数

- 问题的提出

实际问题求的方程组 $A\vec{x}=\vec{b}$ 中的 A 与 $\vec{b}$ 都是测量数据，与真实的数据有误差，我们的问题是：

当有误差 δA 与 $\delta\vec{b}$ 时，

1）对矩阵的逆 A^{-1} 有何影响；

2）方程组的解 $\vec{x}$ 有何影响；

矩阵的条件数

- 引理：设$P\in C^{n\times n}$，若对$C^{n\times n}$上的某一矩阵范数$\|\cdot\|$有$\|P\|<1$，则$I-P$可逆.

证明：反证，若$I-P$ 不可逆，则

$(I-P)\overrightarrow{x_0}=\vec{0}$ 有非零解 x_0 $\Longrightarrow$ $\overrightarrow{x_0}=P\overrightarrow{x_0}$

设$\|\cdot\|_V$ 是与 $\|\cdot\|$ 相容的向量范数，则

$$\left\|\overrightarrow{x_0}\right\|_V=\left\|P\overrightarrow{x_0}\right\|_V\le\|P\|\left\|\vec{x}\right\|_V$$

即$\|P\|\ge1$与$\|P\|<1$矛盾　故$I-P$可逆.

注：可证 $I+P$ 也可逆，即引理2.2可简述如下：

$\|P\|<1$ $\Longrightarrow$ $I\pm P$可逆

矩阵的条件数

- 定理 2.15：设$A\in C_n^{n\times n}$，$\delta A\in C^{n\times n}$. 若对$C^{n\times n}$上的某一矩阵范数$\|\cdot\|$有$\|A^{-1}\delta A\|<1$，则

(1)$A+\delta A$可逆

(2)$\left\|(A+\delta A)^{-1}\right\|\le\dfrac{\|A^{-1}\|}{1-\|A^{-1}\delta A\|}$;

(3)$\dfrac{\left\|A^{-1}-(A+\delta A)^{-1}\right\|}{\|A^{-1}\|}\le\dfrac{\|A^{-1}\delta A\|}{1-\|A^{-1}\delta A\|}$.

(1)$A+\delta A$可逆　　(2)$\left\|(A+\delta A)^{-1}\right\|\le\dfrac{\|A^{-1}\|}{1-\|A^{-1}\delta A\|}$;

证明：（1）由于 $A+\delta A=A(I+A^{-1}\delta A)$　$\because\|A^{1}\delta A\|<1$

由引理知$I+A^{-1}\delta A$可逆，又A可逆　$\therefore A+\delta A$可逆

（2）由$(A+\delta A)(A+\delta A)^{-1}=I$

得$A(A+\delta A)^{-1}=I-\delta A\cdot(A+\delta A)^{-1}$

$(A+\delta A)^{-1}=A^{-1}-A^{-1}\cdot\delta A\cdot(A+\delta A)^{-1}$

$\therefore\left\|(A+\delta A)^{-1}\right\|\le\left\|A^{-1}\right\|+\left\|A^{-1}\delta A\right\|\bullet\left\|(A+\delta A)^{-1}\right\|$

解之得：$\left\|(A+\delta A)^{-1}\right\|\le\dfrac{\|A^{-1}\|}{1-\|A^{-1}\delta A\|}$

(3)$\dfrac{\left\|A^{-1}-(A+\delta A)^{-1}\right\|}{\|A^{-1}\|}\le\dfrac{\|A^{-1}\delta A\|}{1-\|A^{-1}\delta A\|}$.

证明：（3）由矩阵乘法左右分配律

$$A^{-1}-(A+\delta A)^{-1}=A^{-1}[(A+\delta A)-A](A+\delta A)^{-1}$$
$$=A^{-1}\delta A(A+\delta A)^{-1}$$

由（2）得 $\left\|A^{-1}-(A+\delta A)^{-1}\right\|\le\left\|A^{-1}\delta A\right\|\cdot\left\|(A+\delta A)^{-1}\right\|$

$\left(\left\|(A+\delta A)^{-1}\right\|\le\dfrac{\|A^{-1}\|}{1-\|A^{-1}\delta A\|}\right)$　　$\le\left\|A^{-1}\delta A\right\|\cdot\dfrac{\|A^{-1}\|}{1-\|A^{-1}\delta A\|}$

即$\dfrac{\left\|A^{-1}-(A+\delta A)^{-1}\right\|}{\|A^{-1}\|}\le\dfrac{\|A^{-1}\delta A\|}{1-\|A^{-1}\delta A\|}$

矩阵的条件数

- 推论：设$A \in C_n^{n\times n}$，$\delta A \in C^{n\times n}$，若对$C^{n\times n}$上的某一矩阵范数$\|\cdot\|$有$\|A^{-1}\|\|\delta A\| < 1$，则

$$\frac{\left\|A^{-1}-(A+\delta A)^{-1}\right\|}{\left\|A^{-1}\right\|} \le \frac{\|A\|\cdot\left\|A^{-1}\right\|\cdot\frac{\|\delta A\|}{\|A\|}}{1-\|A\|\cdot\left\|A^{-1}\right\|\cdot\frac{\|\delta A\|}{\|A\|}}$$

注：1）$\left\|A^{-1}\delta A\right\| \le \left\|A^{-1}\right\|\|\delta A\| < 1$及$\frac{x}{1-x}$单增，由定理2.15中性质（3）即可完成推论证明。

矩阵的条件数

- 推论：设$A \in C_n^{n\times n}$，$\delta A \in C^{n\times n}$，若对$C^{n\times n}$上的某一矩阵范数$\|\cdot\|$有$\|A^{-1}\|\|\delta A\| < 1$，则

$$\frac{\left\|A^{-1}-(A+\delta A)^{-1}\right\|}{\left\|A^{-1}\right\|} \le \frac{\|A\|\cdot\left\|A^{-1}\right\|\cdot\frac{\|\delta A\|}{\|A\|}}{1-\|A\|\cdot\left\|A^{-1}\right\|\cdot\frac{\|\delta A\|}{\|A\|}}$$

注：2）$\frac{\left\|A^{-1}-(A+\delta A)^{-1}\right\|}{\left\|A^{-1}\right\|}, \frac{\|\delta A\|}{\|A\|}$ 分别是 A^{-1}，与 A 关于矩阵范数的相对误差。

矩阵的条件数

- 定理 2.16：设$A \in C_n^{n\times n}$，$\delta A \in C^{n\times n}$，$b, \delta b \in C^n$，若对$C^{n\times n}$上的某一矩阵范数$\|\cdot\|$有$\|A^{-1}\|\|\delta A\| < 1$，则非齐次线性方程组

$$Ax=b \quad 与 \quad (A+\delta A)(x+\delta x)=b+\delta b$$

的解满足

$$\frac{\|\delta x\|_v}{\|x\|_v} \le \frac{\|A\|\cdot\left\|A^{-1}\right\|}{1-\|A\|\cdot\left\|A^{-1}\right\|\cdot\frac{\|\delta A\|}{\|A\|}}\left(\frac{\|\delta A\|}{\|A\|}+\frac{\|\delta b\|_v}{\|b\|_v}\right),$$

其中$\|\cdot\|_v$是C^n上与矩阵范数$\|\cdot\|$相容的向量范数

矩阵的条件数

- Remark
  - 由前面的推论和定理2.16可知：数据的误差对逆矩阵和线性方程组解的影响与数$\|A\|\|A^{-1}\|$的大小有关
- 定义2.7：设$A \in C_n^{n\times n}$，$\|\cdot\|$是$C^{n\times n}$上的矩阵范数，称

$$cond(A)=\|A\|\cdot\left\|A^{-1}\right\|$$

为矩阵A(关于求逆或求解线性方程组)的条件数

矩阵的条件数

注意：1）常用的条件数有

$$Cond_\infty A=\|A\|_\infty \cdot \left\|A^{-1}\right\|_\infty$$

$$Cond_2 A=\|A\|_2 \cdot \left\|A^{-1}\right\|_2=\sqrt{\frac{\mu_1}{\mu_n}}=\sqrt{\frac{\max A^H A特征值}{\min A^H A特征值}}$$

2）当 $CondA$ 较小时，称 $A\vec{x}=\vec{b}, A\in C_n^{n\times n}$ 是良态。反之，当 $CondA$ 较大时，称 $A\vec{x}=\vec{b}$ 是病态的。

例：设方程组

$$\begin{cases} x_1+10^5 x_2=10^5 \\ x_1+x_2=2 \end{cases}$$

改写为 $A\vec{x}=\vec{b}$ 求条件数 $Cond_\infty A$

解：

$$\because A^{-1}=\frac{\begin{pmatrix} -1 & 10^5 \\ 1 & -1 \end{pmatrix}}{10^5-1}$$

则 $$Cond_\infty A=\|A\|_\infty\left\|A^{-1}\right\|_\infty=\frac{(1+10^5)^2}{10^5-1}\approx 10^5$$

$$\therefore \begin{cases} x_1+10^5 x_2=10^5 \\ x_1+x_2=2 \end{cases}$$ 是病态的

例：设方程组

$$\begin{cases} x_1 + 10^5 x_2 = 10^5 \\ x_1 + x_2 = 2 \end{cases}$$ 改写为 $A\vec{x} = \vec{b}$ 求条件数 $Cond_\infty A$

解：但若方程组改为

$$\begin{cases} 10^{-5} x_1 + x_2 = 1 \\ x_1 + x_2 = 2 \end{cases}$$ 写为 $B\vec{x} = \vec{c}$

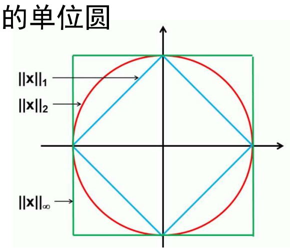
$$B = \begin{pmatrix} 10^{-5} & 1 \\ 1 & 1 \end{pmatrix} \quad B^{-1} = \frac{\begin{pmatrix} 1 & -1 \\ -1 & 10^{-5} \end{pmatrix}}{10^{-5} - 1}$$

$$Cond_\infty B = \|B\|_\infty \|B^{-1}\|_\infty \approx 4$$

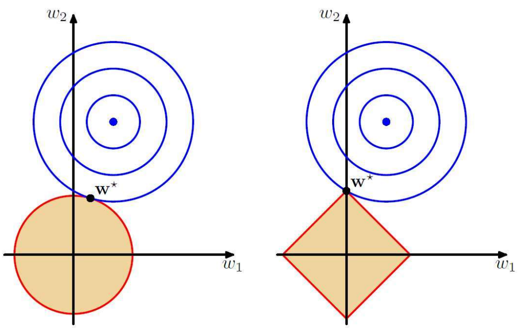
$$\begin{cases} 10^{-5} x_1 + x_2 = 1 \\ x_1 + x_2 = 2 \end{cases}$$ 方程组是良态的

范数理论的应用

- 向量范数的应用
  - 机器学习算法里的距离度量，如k-NN、k-means、正则项（regularization）、度量学习
  - 向量p-范数，$p \geq 1$时$\|x\|_p \leq 1$是凸集
  - 不同范数的单位圆



1范数 vs 2范数



L2 regularization　　L1 regularization

[Bishop, 2006. Pattern recognition and machine learning.]

思考题

- 设λ是可逆矩阵A特征值，证明下列不等式

$$\frac{1}{\|A^{-1}\|_2} \leq |\lambda| \leq \|A\|_2$$

小结

- 矩阵的谱半径：定义、计算、估计
- 矩阵的条件数：定义、应用

Assignment

- 徐仲、张凯院等.《矩阵论简明教程》, 第三版. 科学出版社出版.
  - 习题二：11，12