范数及敏度分析

教师: 胡俊

北京大学数学科学学院

November 9, 2019

目录

目录

Definition 1.1

一个从$\mathbb{R}^n$到$\mathbb{R}$的非负函数$\|\cdot\|$称做$\mathbb{R}^n$上的向量范数. 如果有

(1) 正定性: $\|x\| \geq 0$, 且$\|x\| = 0 \Leftrightarrow x = 0$;

(2) 齐次性: $\|\alpha x\| = |\alpha|\|x\|, \forall \alpha \in \mathbb{R}, x \in \mathbb{R}^n$;

(3) 三角不等式: $\|x + y\| \leq \|x\| + \|y\|, \forall x, y \in \mathbb{R}^n$.

$$\Big|\|x\| - \|y\|\Big| \leq \|x - y\| \leq \max_{1\leq i\leq n} \|e_i\| \sum_{i=1}^{n} |x_i - y_i|.$$

由此知, $\|\cdot\|$是$\mathbb{R}^n$上的连续函数.

定理:

设V是数域R上n维线性空间, $\|\cdot\|_\alpha$和$\|\cdot\|_\beta$是V上的范数, 则存在正常数C_1和C_2使得

$$C_1\|v\|_\beta \le \|v\|_\alpha \le C_2\|v\|_\beta \qquad \forall v \in V.$$

证明: 设$\xi = [\xi_1, \xi_2, \cdots, \xi_n]$是$V$的一个基. 对$V$上的任意范数$\|\cdot\|$, 定义$\phi : R^n \to R$

$$\phi(x) = \|\xi x\| \qquad \forall x \in R^n.$$

易证$\phi(x)$是R^n上非负连续函数，且当$\phi(x) = 0$时有$x = 0$. 因而$\phi(x)$在单位球面$Y = \{x : \|x\|_2 = 1, x \in R^n\}$上有最大最小值. 特别地, 对范数$\|\cdot\|_\alpha$, 最大和最小值分别为$D^\alpha_{max}$ 和D^α_{min}; 对范数$\|\cdot\|_\beta$, 最大和最小值分别为D^β_{max} 和D^β_{min}. 对任意非零向量v, 存在$x \in R^n$使得$v = \xi x = \|x\|_2\xi\frac{x}{\|x\|_2}$. 因此,

$$C_1 := \frac{D^\alpha_{min}}{D^\beta_{max}} \le \frac{\|v\|_\alpha}{\|v\|_\beta} = \frac{\|\xi\frac{x}{\|x\|_2}\|_\alpha}{\|\xi\frac{x}{\|x\|_2}\|_\beta} \le C_2 := \frac{D^\alpha_{max}}{D^\beta_{min}}$$

常用向量范数是p范数(Hölder范数)

$$\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}}, \ p \geq 1$$

$$\|x\|_1 = (|x_1| + \cdots + |x_n|) \quad \rightarrow 1\text{范数}$$
$$\|x\|_2 = (|x_1|^2 + \cdots + |x_n|^2)^{\frac{1}{2}} = \sqrt{x^T x} \quad \rightarrow 2\ \text{范数}$$
$$\|x\|_\infty = \max_{1\leq i\leq n} |x_i|, \quad \rightarrow \infty\ \text{范数}$$

范数等价(作业)

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2$$
$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty$$
$$\|x\|_\infty \leq \|x\|_1 \leq n\|x\|_\infty$$

Theorem 1.2 (杨不等式)

假设a和b为非负实数，$p>1$，且$\frac{1}{p}+\frac{1}{q}=1$，则

$$ab \le \frac{a^p}{p}+\frac{b^q}{q},$$

等号成立当且仅当$a^p=b^q$.

证明： 若$b=0$, 则不等式显然成立. 若$b\neq 0$, 令$t=\frac{a^p}{b^q}$, $\gamma=\frac{1}{p}$, 则不等式变成

$$t^\gamma \le \gamma t+1-\gamma.$$

令$f(t)=t^\gamma-\gamma t$, 则$f'(t)=\gamma t^{\gamma-1}-\gamma$. 当$t>1$时，$f'(t)<0, f(t)$单调递减; 当$t<1$时，$f'(t)>0, f(t)$单调递增. 因此，

$$f(t)\le f(1)=1-\gamma$$

即

$$t^\gamma \le \gamma t+1-\gamma.$$

证毕.

Theorem 1.3 (Hölder不等式)

$$|x^T y| \le \|x\|_p \|y\|_q \quad \frac{1}{p} + \frac{1}{q} = 1, \ \ p \ge 1.$$

证明： 不妨设$x_i \ge 0$, $y_i \ge 0$($i = 1, \cdots, n$). 设$X = \sum_{i=1}^{n} x_i^p \ge 0$, $Y = \sum_{i=1}^{n} y_i^q \ge 0$. 当$XY = 0$时, 不等式显然成立. 因此假设$XY \ne 0$.
在杨不等式令$a = \frac{x_i}{X^{1/p}}$, $b = \frac{y_i}{Y^{1/q}}$, 得

$$\frac{x_i}{X^{1/p}} \frac{y_i}{Y^{1/q}} \le \frac{x_i^p}{Xp} + \frac{y_i^q}{Yq}.$$

对i求和得

$$\sum_{i=1}^{n} x_i y_i \le X^{1/p} Y^{1/q} \Big(\sum_{i=1}^{n} \frac{x_i^p}{Xp} + \sum_{i=1}^{n} \frac{y_i^q}{Yq} \Big) = X^{1/p} Y^{1/q} = \|x\|_p \|y\|_q.$$

证毕.

目录

Definition 2.1

一个从$\mathbb{R}^{n\times n}$到$\mathbb{R}$的非负函数$\|\cdot\|$称做$\mathbb{R}^{n\times n}$上的矩阵范数. 如果它满足

1. 正定性: 对$\forall A\in\mathbb{R}^{n\times n}$, 有$\|A\|\geq 0$, 且$\|A\|=0 \Leftrightarrow A=0$.
2. 齐次性: 对$\forall A\in\mathbb{R}^{n\times n}$, $\alpha\in\mathbb{R}$, 有$\|\alpha A\|=|\alpha|\|A\|$.
3. 三角不等式: 对$\forall A,B\in\mathbb{R}^{n\times n}$, 有$\|A+B\|\leq\|A\|+\|B\|$.
4. 相容性: 对$\forall A,B\in\mathbb{R}^{n\times n}$, 有$\|AB\|\leq\|A\|\|B\|$.

Definition 2.2

若矩阵范数$\|\cdot\|_M$和向量范数$\|\cdot\|_v$满足

$$\|Ax\|_v \le \|A\|_M\|x\|_v,\ A \in \mathbb{R}^{n\times n}, x \in \mathbb{R}^n.$$

则称$\|\cdot\|_M$与$\|\cdot\|_v$是相容的.

$\|A\| = \max\limits_{\|x\|=1} \|Ax\|,\ A \in \mathbb{R}^{n\times n}$ 是$\mathbb{R}^{n\times n}$上的一个矩阵范数, 其中$\|\cdot\|$ 是$\mathbb{R}^n$上的一个范数. 称$|||\cdot|||$为矩阵的算子范数.

$$\|A\|_p = \max_{\|x\|_p=1} \|Ax\|_p,\ A \in \mathbb{R}^{n\times n}$$

作业：证明$\|A\| = \max\limits_{\|x\|=1} \|Ax\|,\ A \in \mathbb{R}^{n\times n}$ 是$\mathbb{R}^{n\times n}$上的一个矩阵范数.

Theorem 2.3

设$A = [a_{ij}] \in \mathbb{R}^{n\times n}$, 则有

$$
\begin{aligned}
\|A\|_1 &= \max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}| \longrightarrow \text{列和范数}\\
\|A\|_\infty &= \max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}| \longrightarrow \text{行和范数}\\
\|A\|_2 &= \sqrt{\lambda_{\max}(A^TA)}
\end{aligned}
$$

证明:

$A=0$时, 定理显然成立. 下面设$A\neq 0$. 对于1范数, 将$A\in\mathbb{R}^{n\times n}$按列分块为$A=[a_1,\cdots,a_n]$, 并设

$$\delta=\|a_{j_0}\|_1=\max_{1\le j\le n}\|a_j\|_1$$

则对任意$x\in\mathbb{R}^n$满足$\|x\|_1=\sum_{i=1}^{n}|x_i|=1$有

$$\begin{aligned}\|Ax\|_1&=\left\|\sum_{j=1}^{n}x_ja_j\right\|_1\le\sum_{j=1}^{n}|x_j|\|a_j\|_1\\&\le\left(\sum_{j=1}^{n}|x_j|\right)\max_{1\le j\le n}\|a_j\|_1=\|a_{j_0}\|_1=\delta.\end{aligned}$$

另一方面, 若取e_{j_0}为n阶单位矩阵的第j_0列, 则有$\|e_{j_0}\|_1=1$, 且

$$\|Ae_{j_0}\|_1=\|a_{j_0}\|_1=\delta.$$

因此, 有

$$\|A\|_1 = \max_{\|x\|_1=1} \|Ax\|_1 = \delta = \max_{1\le j\le n} \|a_j\|_1 = \max_{1\le j\le n} \sum_{i=1}^{n} |a_{ij}|.$$

对于∞范数, 记$\eta = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|$, 则对任一$x \in \mathbb{R}^n$满足$\|x\|_\infty = 1$有

$$\begin{aligned}\|Ax\|_\infty &= \max_{1\le i\le n} \left|\sum_{j=1}^{n} a_{ij}x_j\right| \le \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}||x_j| \\ &\le \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}| = \eta.\end{aligned}$$

设A的第k行的∞范数最大, 即$\eta = \sum\limits_{j=1}^{n} |a_{kj}|$. 令

$$\tilde{x} = (\mathrm{sgn}(a_{k1}), \cdots, \mathrm{sgn}(a_{kn}))^T,$$

则$A \neq 0$蕴含着$\|\tilde{x}\|_\infty = 1$, 而且有$\|A\tilde{x}\|_\infty = \eta$. 这样, 我们就已经证明了

$$\|A\|_\infty = \eta = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|.$$

对于2范数, 应有

$$\begin{aligned}\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 &= \max_{\|x\|_2=1} [(Ax)^T Ax]^{\frac{1}{2}} \\ &= \max_{\|x\|_2=1} [x^T (A^T A) x]^{\frac{1}{2}}.\end{aligned}$$

注意, A^TA是半正定的对称阵, 设其特征值为

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0.$$

以及其对应的正交规范特征向量为$v_1, \cdots, v_n \in \mathbb{R}^n$, 则对任一满足$\|x\|_2 = 1$的向量$x \in \mathbb{R}^n$有

$$x = \sum_{i=1}^{n} \alpha_i v_i \text{ 和 } \sum_{i=1}^{n} \alpha_i^2 = 1.$$

于是, 有

$$x^T A^T A x = \sum_{i=1}^{n} \lambda_i \alpha_i^2 \leq \lambda_1.$$

另一方面, 若取$x = v_1$, 则有

$$x^T A^T A x = v_1^T A^T A v_1 = v_1^T \lambda_1 v_1 = \lambda_1.$$

所以

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \sqrt{\lambda_1} = \sqrt{\lambda_{\max}(A^T A)}.$$

证明结束.

谱半径

Definition 2.4

设$A \in C^{n\times n}$, 则称

$$\rho(A) = \max\{|\lambda| : \lambda \in \lambda(A)\}$$

为A的谱半径, 这里$\lambda(A)$表示A的特征值的全体.

谱半径与矩阵范数之间有如下关系:

Theorem 2.5

设$A \in C^{n\times n}$, 则有

(1) 对$C^{n\times n}$上的任意矩阵范数$\|\cdot\|$, 有

$$\rho(A) \le \|A\|;$$

(2) 对任意给定的$\epsilon > 0$, 存在$C^{n\times n}$上的算子范数$\|\cdot\|$, 使得

$$\|A\| \le \rho(A) + \epsilon.$$

证明:

(1) 设$x \in C^n$满足

$$x \neq 0,\ Ax = \lambda x,\ |\lambda| = \rho(A)$$

$$\lambda x e_1^T = A x e_1^T,\ 且 0 \neq x e_1^T \in C^{n\times n}.$$

这样

$$\rho(A)\|xe_1^T\| = \|\lambda x e_1^T\| \leq \|A\|\|xe_1^T\|.$$

因此

$$\rho(A) \leq \|A\|.$$

(2)由Jordan分解定理知, 存在非奇异矩阵$X \in C^{n\times n}$, 使得

$$X^{-1}AX = \begin{bmatrix} \lambda_1 & \delta_1 & & & \\ & \lambda_2 & \delta_2 & & \\ & & \ddots & \ddots & \\ & & & \lambda_{n-1} & \delta_{n-1} \\ & & & & \lambda_n \end{bmatrix}$$

其中$\delta_i = 1$或0.

对于任意给足的$\epsilon > 0$, 令

$$D_\epsilon = \mathrm{diag}(1, \epsilon, \epsilon^2, \cdots, \epsilon^{n-1}).$$

则有

$$D_\epsilon^{-1} X^{-1} A X D_\epsilon = \begin{bmatrix} \lambda_1 & \epsilon\delta_1 & & & \\ & \lambda_2 & \epsilon\delta_2 & & \\ & & \ddots & \ddots & \\ & & & \lambda_{n-1} & \epsilon\delta_{n-1} \\ & & & & \lambda_n \end{bmatrix}$$

现在定义

$$\|G\|_\epsilon = \|D_\epsilon^{-1} X^{-1} G X D_\epsilon\|_\infty, \quad G \in C^{n\times n}.$$

则容易验证这样定义的函数$\|\cdot\|_\epsilon$是由如下的向量范数

$$\|x\|_{XD_\epsilon} = \|(XD_\epsilon)^{-1}x\|_\infty, \quad x \in C^n$$

诱导出的算子范数, 而且有

$$\begin{aligned}\max_{x\neq 0}\frac{\|Ax\|_{XD_\epsilon}}{\|x\|_{XD_\epsilon}} &= \max_{x\neq 0}\frac{\|(XD_\epsilon)^{-1}Ax\|_\infty}{\|(XD_\epsilon)^{-1}x\|_\infty}\\ &= \max_{y\neq 0}\frac{\|(XD_\epsilon)^{-1}AXD_\epsilon y\|_\infty}{\|y\|_\infty} = \|D_\epsilon^{-1}X^{-1}AXD_\epsilon\|_\infty.\end{aligned}$$

因此

$$\|A\|_\epsilon = \|D_\epsilon^{-1}X^{-1}AXD_\epsilon\|_\infty = \max_{1\leq i\leq n}(|\lambda_i| + |\epsilon\delta_i|) \leq \rho(A) + \epsilon.$$

其中假定$\delta_n = 0$. 证明结束.

Theorem 2.6

设$A \in C^{n\times n}$, 则

$$\lim_{k\to\infty} A^k = 0 \iff \rho(A) < 1.$$

证明:

必要性. 设$\lim_{k\to\infty} A^k = 0$, 并假定$\lambda \in \lambda(A)$满足$\rho(A) = |\lambda|$. 由于对任意的$k$有$\lambda^k \in \lambda(A^k)$, 故由上述定理有

$$\rho(A)^k = |\lambda|^k \le \rho(A^k) \le \|A^k\|$$

对一切k成立, 从而必有$\rho(A) < 1$.

充分性. 设$\rho(A) < 1$, 则由上述定理, 必有算子范数$\|\cdot\|$使$\|A\| < 1$, 从而

$$0 \le \|A^k\| \le \|A\|^k \to 0,\ k \to \infty$$

于是$\lim_{k\to\infty} A^k = 0$. 证明结束.

Theorem 2.7

设$A \in C^{n\times n}$, 则有

1. $\sum\limits_{k=0}^{\infty} A^k$收敛的充分与必要条件是$\rho(A) < 1$;
2. 当$\sum\limits_{k=0}^{\infty} A^k$收敛时, 有

$$\sum_{k=0}^{\infty} A^k = (I - A)^{-1},$$

而且存在$C^{n\times n}$上的算子范数$\|\cdot\|$, 使得

$$\|(I - A)^{-1} - \sum_{k=0}^{m} A^k\| \leq \frac{\|A\|^{m+1}}{1 - \|A\|}.$$

Corollary 2.8

设$\|\cdot\|$是$C^{n\times n}$上的一个满足条件$\|I\| = 1$的矩阵范数, 并假定$A \in C^{n\times n}$ 满足$\|A\| < 1$, 则$I - A$可逆且有

$$\|(I - A)^{-1}\| \le \frac{1}{1 - \|A\|}.$$

目录

矩阵条件数

$\kappa(A) = \|A\|\|A^{-1}\|$称为矩阵$A$的条件数.

$$Ax = b, \quad (A + \delta A)(x + \delta x) = b + \delta b, \quad (\|A^{-1}\|\|\delta A\| < 1)$$

则

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\kappa(A)}{1 - \kappa(A)\frac{\|\delta A\|}{\|A\|}} \left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|} \right).$$

当$\frac{\|\delta A\|}{\|A\|}$较小时, 有

$$\frac{\kappa(A)}{1 - \kappa(A)\frac{\|\delta A\|}{\|A\|}} \approx \kappa(A).$$

从而

$$\frac{\|\delta x\|}{\|x\|} \lessapprox \kappa(A) \left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|} \right).$$

定理： 设$A \in \mathbb{R}^{n\times n}$非奇异, 则

$$\min\left\{\frac{\|\delta A\|_2}{\|A\|_2} : A+\delta A \quad \text{奇异} \quad \right\} = \frac{1}{\|A\|_2\|A^{-1}\|_2} = \frac{1}{\kappa_2(A)},$$

即在谱范数下，一个矩阵的条件数的倒数恰好等于该矩阵与全体奇异矩阵所组成集合的相对距离.

证明： 只需证明

$$\min\{\|\delta A\|_2 : A+\delta A, \quad \text{奇异} \quad \} = \frac{1}{\|A^{-1}\|_2}$$

即可. 当$\|A^{-1}\|_2\|\delta A\|_2 < 1$时, $A+\delta A$必是非奇异的, 从而有

$$\min\{\|\delta A\|_2 : A+\delta A, \quad \text{奇异} \quad \} \geq \frac{1}{\|A^{-1}\|_2}.$$

由于谱范数是由向量2范数诱导出的算子范数, 因此必存在$x \in \mathbb{R}^n$且$\|x\|_2 = 1$, 使得

$$\|A^{-1}x\|_2 = \|A^{-1}\|_2.$$

令

$$y = \frac{A^{-1}x}{\|A^{-1}x\|_2}, \delta A = -\frac{xy^T}{\|A^{-1}\|_2},$$

则有

$$(A + \delta A)y = Ay + \delta Ay = \frac{x}{\|A^{-1}x\|_2} - \frac{x}{\|A^{-1}\|_2} = 0,$$

$$\|\delta A\|_2 = \max_{\|x\|_2=1} \left\| \frac{xy^T}{\|A^{-1}\|_2} z \right\| = \frac{\|x\|_2}{\|A^{-1}\|_2} \max_{\|x\|_2=1} = \frac{1}{\|A^{-1}\|_2}.$$

这就是说存在δA使得$A + \delta A$是奇异的, 且$\|\delta A\|_2 = \frac{1}{\|A^{-1}\|_2}$. 因此结论成立.

谢谢！