第四章 向量与矩阵的范数

本章数域 $\mathbb{F}$ 指实数域 $\mathbb{R}$ 或复数域 $\mathbb{C}$.

4.1 向量范数

定义

定义（向量范数）

设 V 是 $\mathbb{F}$ 上的线性空间. V 上的实值函数

$$\|\cdot\| : V \to \mathbb{R}^+$$

称为 V 上的一个范数，如果它满足

（1）正定性：对任意非零向量 x，$\|x\| > 0$.

（2）正齐性：对任意向量 x 及任意数 k，$\|xk\| = \|x\| \, |k|$.

（3）三角不等式：对任意两个向量 x 和 y，$\|x+y\| \le \|x\| + \|y\|$.

定理（向量范数的性质）

（1） $\|0\| = 0$.

（2） $\|-x\| = \|x\|$.

（3） $\|x-y\| \geq |\,\|x\| - \|y\|\,|$.

（4） $\|x+y\| \geq |\,\|x\| - \|y\|\,|$.

例（内积空间中向量的长度是范数）

以下约定，凡是说到长度，都是专指内积空间中由内积定义的长度.

问题：范数何时是长度?

定理（范数何时是长度?）

范数是长度的充要条件是它满足平行四边形公式，即：

设 $\|\cdot\|$ 是线性空间 V 上的一个范数．如果它满足平行四边形公式，即对任意两个向量 x 和 y，

$$\|x+y\|^2+\|x-y\|^2=2\left(\|x\|^2+\|y\|^2\right),$$

则存在唯一的 V 上的一个内积 $\langle\cdot\rangle$，使得对任意向量 x

$$\|x\|=\sqrt{\langle x,x\rangle}.$$

例（$\mathbb{C}^n$ 上的 p –范数）

对任意 $x \in \mathbb{C}^n$， $1 \le p \le \infty$， 定义

$$\|x\|_p = \begin{cases} \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{\frac{1}{p}}, & 1 \le p < \infty, \\ \max\{|x_1|, |x_2|, \ldots, |x_n|\}, & p = \infty. \end{cases}$$

可以证明， $\|\cdot\|_p$ 是 $\mathbb{C}^n$ 上的范数， 称为 p –范数.

以下事实

$$\lim_{p\to\infty}\left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{\frac{1}{p}} = \max\{|x_1|, |x_2|, \ldots, |x_n|\}$$

解释了记号 $\|\cdot\|_\infty$ 的合理性.

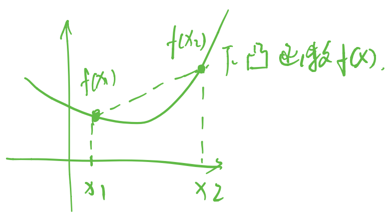
证: p-范数是范数

(1) (2) 易证

(3). $\|x+y\|_p \le \|x\|_p + \|y\|_p$

即 $\left(\sum_{i=1}^n |x_i+y_i|^p\right)^{\frac{1}{p}} \le \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |y_i|^p\right)^{\frac{1}{p}}$

引理:(凸组合)



x_1和x_2的凸组合为线段 x_1x_2

$f(x_1)$ 和 $f(x_2)$ --- . . $f(x_1)f(x_2)$

用泰勒公式可证:

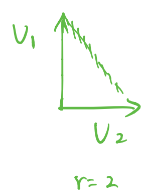
$f\left(\sum_{i=1}^n x_i\lambda_i\right) \le \sum_{i=1}^n f(x_i)\lambda_i$

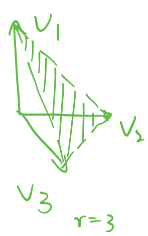
$f''(x) > 0 \Rightarrow f(x)$ 是下凸的.

凸组合: r个向量 $\{v_1, v_2, \dots, v_r\}$

其凸组合 $\sum_{i=1}^r v_i\lambda_i$, 其中

$0 \le \lambda_i \le 1$, $\sum_{i=1}^r \lambda_i = 1$ (证明, 省)





阴影部分为凸组合可以落在的空间.

回到(3). 两边同除左边, 得 $1 \le \left[\frac{\sum a_i^p}{\sum_k (a_k+b_k)^p}\right]^{\frac{1}{p}} + \left[\frac{\sum b_i^p}{\sum_k (a_k+b_k)^p}\right]^{\frac{1}{p}}$

$$\frac{\sum a_i^p}{\sum_k (a_k+b_k)^p} = \sum_i \frac{a_i^p}{\sum_k (a_k+b_k)^p}$$

$$= \sum \frac{(a_i+b_i)^p}{\sum (a_k+b_k)^p} \cdot \frac{a_i^p}{(a_i+b_i)^p}$$

构造出一个λ来

$$= \sum_i \underbrace{\frac{(a_i+b_i)^p}{\sum_k (a_k+b_k)^p}}_{\lambda_i} \cdot \left(\frac{a_i}{a_i+b_i}\right)^p \ge \left(\sum_i \lambda_i \frac{a_i}{a_i+b_i}\right)^p$$

(引理)

设 $x_i = \frac{a_i}{a_i+b_i}$, $f(x) = x^p$

$f''(x) \ge 0$, 当 $p \ge 2$ 时, 下凸

$$
\text{原式} = \left[\sum_i \frac{(a_i+b_i)^p}{\sum_k (a_k+b_k)^p} \cdot \left(\frac{a_i}{a_i+b_i}\right)^p\right]^{\frac{1}{p}} + \left[\sum_i \frac{(a_i+b_i)^p}{\sum_k (a_k+b_k)^p} \cdot \left(\frac{b_i}{a_i+b_i}\right)^p\right]^{\frac{1}{p}}
$$

$$
\geq \left[\left(\sum_i \lambda_i \frac{a_i}{a_i+b_i}\right)^p\right]^{\frac{1}{p}} + \left[\left(\sum_i \lambda_i \frac{b_i}{a_i+b_i}\right)^p\right]^{\frac{1}{p}}
$$

$$
= \sum_i \lambda_i \frac{a_i}{a_i+b_i} + \sum_i \lambda_i \frac{a_i}{a_i+b_i}
$$

$$
= \sum_i \lambda_i \left(\frac{a_i}{a_i+b_i} + \frac{b_i}{a_i+b_i}\right)
$$

$$
= \sum_i \lambda_i = 1
$$

例（p-范数中只有 2-范数是长度）

只需证明，当 $p \neq 2$ 时，p-范数不满足平行四边形公式．例如，在 $\mathbb{C}^2$ 中，取

$$x=\begin{bmatrix}1\\0\end{bmatrix},\quad y=\begin{bmatrix}0\\1\end{bmatrix},$$

分别计算四个范数

$$\|x\|_p=\|y\|_p=1,\quad \|x+y\|_p=\|x-y\|_p=2^{\frac{1}{p}}.$$

$$\|x\|_\infty=\|y\|_\infty=\|x+y\|_\infty=\|x-y\|_\infty=1.$$

易见，

$$\left(\|x+y\|_p\right)^2+\left(\|x-y\|_p\right)^2\neq 2\left(\|x\|_p{}^2+\|y\|_p{}^2\right).$$

问题：既然已经有了长度，为什么还要考虑不是长度的范数?

引入范数使得我们可以考虑两个向量的距离．而距离好处有两个，分别是定性的和定量的．定性的方面是可以考虑极限，进而引入连续，导数与积分等数学分析的方法．定量的方面是可以为最优逼近问题提供逼近性能指标．定性的角度来说，各种范数的作用是等价的(范数的等价性)；但就提供符合实际需求的逼近性能指标来说，不同的范数起着不同的作用，给出的“最优解”也不同．

定义（范数定义距离）

设 $\|\cdot\|$ 是向量空间 V 上的范数. 对任意 $x, y \in V$，记 $\mathrm{d}(x, y) = \|x - y\|$. 如此决定了 V 上的二元函数

$$\mathrm{d}(\cdot,\cdot) : V \times V \to \mathbb{R}^+,$$

称为由范数 $\|\cdot\|$ 定义的距离.

例（中位数作为最小一乘估计）

范数的等价性

中位数.

数组$\{x_1, x_2 \cdots x_n\}$的有序排列 $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$

$$中位数 = \begin{cases} x_{(\frac{n+1}{2})}, & n为奇数 \\ \left[x_{(\frac{n}{2})},\ x_{(\frac{n}{2}+1)}\right], & n为偶数. \end{cases}$$

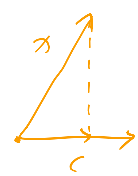
(补) 算术平均作为最小二乘估计:

$\sum_{i=1}^{n} |x_i - c|^2 \to \min$, 当$c$是算术平均

2范数意义下的最优解

$$f(c) = \sum_{i=1}^{n} (x_i - c)^2 = \left\| \begin{bmatrix} x_1 - c \\ x_2 - c \\ \vdots \\ x_n - c \end{bmatrix} \right\|_2$$

$$= \| x - \mathbb{1} \cdot c \|, \quad \mathbb{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$



x在$\vec{c}$上的正交投影.

(几何观点.)

中位数的极值刻画:

$\sum_{i=1}^{n} |x_i - c| \longrightarrow \min$, 当$c$是中位数,

1范数意义下的最优解

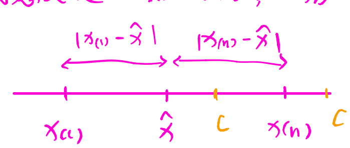
记: $\sum_{i=1}^{n} |x_{(i)} - \hat{x}| \le \sum_{i=1}^{n} |x_{(i)} - c|, \ \forall c$

$\Rightarrow x_{(n)} - x_{(1)} \le |x_{(1)} - \hat{x}| + |x_{(n)} - \hat{x}| \le |x_{(1)} - c| + |x_{(n)} - c|$

$x_{(n-1)} - x_{(2)} \le |x_{(2)} - \hat{x}| + |x_{(n-1)} - \hat{x}| \le |x_{(2)} - c| + |x_{(n-1)} - c|$

$\vdots$

几何意义是两点间距; 若c在两点间, 可以取等, 否则必大于两点间距.



杂谈.

比较 $p=1$ 和 $p=2$.

$p=2$ 最大的好处是光滑的，可以利用求导

$\bar{x} = \sum_{i=1}^{n} \frac{1}{n} x_i$，$x_i$ 有相同概率影响 $\bar{x}$. 若有很大噪声会影响很大.

所以说 不稳健.

$p=1$ 是稳健的，可以很大程度的去除噪声.

范数等价性：

有限维空间中的任意两个范数 $\|\cdot\|_2$，$\|\cdot\|_\beta$，存在 m, M，

$m\|x\|_\beta \le \|x\|_2 \le M\|x\|_\beta$.（$\|\cdot\|_2$，$\|\cdot\|_\beta$ 可相互控制）

等价性是说范数对研究极限的效果是一样的，

但作为评估指标性能是不同的.

例：$\|\cdot\|$，$\|\cdot\|_2$.

① $x = I\cdot x = [e_1\, e_2 \cdots e_n]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_i e_i x_i$

$\|x\| = \|\sum_i e_i x_i\| \overset{(3)}{\le} \sum_i \|e_i x_i\| \overset{(2)}{=} \sum_i \|e_i\|\cdot|x_i| \le \underbrace{\sqrt{\sum_i \|e_i\|}}_{\text{(柯西不等式)}}\cdot\|x\|_2$

② $\|x\|: \mathbb{C}^n \to \mathbb{R}$，$\|x\|$ 是 $\mathbb{C}$ 上的连续函数，则 $\lim_{x\to p}\|x\| = \|p\|$

$S = \{x \mid \|x\|_2 = 1\}$ 是有界闭集，定理：有界闭集上的连续函数一定有最大小值 (M/m).

$0 \le m \le \|x\| \le M$. 则 $m \le \|\underbrace{\frac{x}{\|x\|_2}}_{单位化}\| \le M$

4. 2 矩阵范数

定义 （矩阵范数）

设对任意的正整数 m 和 n， 及任意的 $m\times n$ 矩阵 A， 都有一个对应的实数 $\|A\|$． 若该对应关系满足

（1） 正定性： 对任意 $m\times n$ 的任意非零矩阵 A， 都有 $\|A\|>0$．

（2） 正齐性： 对任意矩阵 A 及数 k， $\|kA\|=|k|\,\|A\|$．

（3） 三角不等式（加法相容性）： 若矩阵 A 和 B 可以相加， 则 $\|A+B\|\le\|A\|+\|B\|$．

（4） 乘法相容性： 若矩阵 A 和 B 可以相乘， 则 $\|AB\|\le\|A\|\|B\|$

则称对应关系 $\|\cdot\|$ 是一个矩阵范数.

其实，严格地说，$\|\cdot\|$应理解为一个映射

$$\|\cdot\|: \bigcup_{m,n\geq 1} \mathbb{F}^{m\times n} \to \mathbb{R}.$$

例（Hilbert-Schmidt 范数）

向量范数导出矩阵范数

观察：由于具体向量 $x\in\mathbb{F}^n$ 其实也就是 $n\times 1$ 的矩阵，因此任一矩阵范数也就同时给出了标准向量空间 $\mathbb{F}^n$，也就是 $\mathbb{F}^{n\times 1}$，$n=1,2,\ldots$，上的向量范数.

反之，也可以从已知的标准向量空间 $\mathbb{F}^n$，$n=1,2,\ldots$，上的向量范数，决定出一个矩阵范数.

定理（向量范数诱导矩阵范数）

设对任意的正整数 $k=1,2,\ldots$，给定了标准向量空间 $\mathbb{F}^k$ 上的向量范数 $\|\cdot\|_{\alpha_k}$．对任意的正整数 m 和 n，及任意的 $m\times n$ 矩阵 A，定义

$$\|A\|:=\max\left\{\frac{\|Ax\|_{\alpha_m}}{\|x\|_{\alpha_n}}:x\in\mathbb{F}^n,x\neq 0\right\}=\max\left\{\|Ax\|_{\alpha_m}:x\in\mathbb{F}^n,\|x\|_{\alpha_n}=1\right\}$$

则 $\|\cdot\|$ 为一矩阵范数，称为由向量范数 $\|\cdot\|_{\alpha_k}$，$k=1,2,\ldots$，导出的矩阵范数.

注意，当 $\|\cdot\|_{\alpha_1}$ 是 $\mathbb{F}^1$（即实数域或复数域）上的绝对值时，如上导出的矩阵范数限制在 $\mathbb{F}^k=\mathbb{F}^{k\times 1}$ 上，才与 $\|\cdot\|_{\alpha_k}$ 重合.

注（向量范数诱导矩阵范数的几何意义）

最大增益(信号与系统的观点)；长度最大放大率(几何的观点).

因此矩阵范数是许多系统设计指标的要利用的工具.

例（向量的 p -范数导出的矩阵范数）

4.3 范数应用举例

矩阵值序列与函数

矩阵指数函数

范数控制特征值

可逆性的范数条件

矩阵的条件数