向量范数的定义

定义

令记号$\|\cdot\| : \mathbb{R}^n \to \mathbb{R}^+$是一种非负函数，如果它满足：

- 正定性：对于一切定义域内的$v \in \mathbb{R}^n$，有$\|v\| \geqslant 0$，且$\|v\| = 0 \Leftrightarrow v = 0_{n\times 1}$；
- 齐次性：对于一切$v \in \mathbb{R}^n$和$\alpha \in \mathbb{R}$，有$\|\alpha v\| = |\alpha|\,\|v\|$；
- 三角不等式：对于一切$v, w \in \mathbb{R}^n$，均成立$\|v+w\| \leqslant \|v\| + \|w\|$.

则称$\|\cdot\|$是定义在向量空间$\mathbb{R}^n$上的**向量范数**.

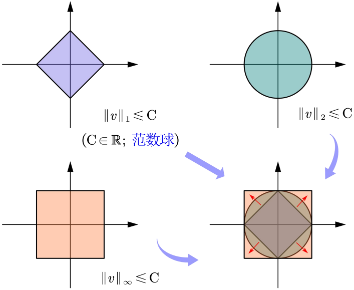
最常用的向量范数即我们熟知的ℓ_p范数(其中$p \geqslant 1$):

$$\|v\|_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{\frac{1}{p}};$$

我们知道，这也是ℓ^p空间中最常用的距离范数.
当$p = \infty$时，距离范数定义为向量所有分量的最大值，即

$$\|v\|_\infty = \max_{1\leqslant j\leqslant n} |v_{(j)}|.$$

向量范数的定义



容易看出, $p=\infty$时, 有关"最大值"的定义要求向量的分量是有限的. 在一般化的空间中, 这一要求很可能不成立, 此时我们只需将"最大值"更换成"上确界"即可.

向量范数度量的是v与零点之间的距离. 在实际应用时, 我们通常使用$p=1,2,\infty$的情形, 即分别使用$\|v\|_1, \|v\|_2, \|v\|_\infty$度量$v$在不同意义下的距离, 这是因为它们具有鲜明的度量特征.

左图是它们各自的范数球实例, 请想一想不同范数所度量的距离分别具有怎样的特征? 这些特征分别适用于度量什么情形?

加权的向量范数

定义

给定任意的向量范数$\|v\|$, 可以定义基于矩阵$W \in \mathbb{R}^{n\times n}$的加权范数

$$\|v\|_W = \|Wv\|,$$

特别, 若W仍是对角矩阵, 则W的对角元分别对应向量各分量的权系数.

A-范数

A-范数是一类由正定矩阵诱导的向量范数.

定义

给定任意的向量范数$\|v\|$, 可以基于正定矩阵$A \in \mathbb{R}^{n\times n}$诱导

$$\|v\|_A = \sqrt{v^T A v}.$$

这即是v的*A*-范数, 记为$\|v\|_A$.

根据正定矩阵的定义, 容易验证A-范数是合法的向量范数. 请验证A-范数是关于"三角不等式"合法的.

事实上, 若规定A仍是对称的, 则A-范数是一类以对称正定矩阵加权的向量范数:

$$\|v\|_A = \sqrt{v^{\mathrm{T}} A v} = \sqrt{v^{\mathrm{T}} A^{1/2} A^{1/2} v} = \sqrt{\left(A^{1/2} v\right)^{\mathrm{T}} \left(A^{1/2} v\right)} = \|v\|_{2,A^{1/2}} \ .$$

定理

设$a, b \in \mathbb{R}^n$, 则

$$|a^{\mathrm{T}}b| \leqslant \|a\|_2 \|b\|_2 ,$$

且等号成立的条件是a与b线性相关.

Cauchy不等式在估值和放缩等不同的情形下具有重要的作用, 在本课后续的算法分析中有所体现.

我们下面简证其成立. 若已熟知此结论, 请直接跳过下一页.

Cauchy不等式的简证

Proof: 设$a, b \in \mathbb{R}^n$, 如果a, b线性相关, 那么存在$k \in \mathbb{R}$, 且有$a = kb$. 将此结果代入Cauchy不等式, 即得

$$|\,(kb)^{\mathrm{T}}\, b| = |k| \cdot |b^T b| = |k| \cdot \|b\|_2^2 = \|kb\|_2 \, \|b\|_2 \,.$$

若a, b不是线性无关的, 那么可先定义$v \in \mathbb{R}^n$, 且$v = a - \frac{a^{\mathrm{T}} b}{b^{\mathrm{T}} b} b$, 这即是Gram-Schmidt正交化的过程, 因此v与b正交.
因此, 利用勾股定理, 成立

$$\|a\|_2^2 = \|v\|_2^2 + \left|\frac{a^{\mathrm{T}} b}{b^{\mathrm{T}} b}\right|^2 \|b\|_2^2 = \|v\|_2^2 + \frac{|a^{\mathrm{T}} b|^2}{\|b\|_2},$$

最终由范数的正定性推出Cauchy不等式:

$$\|a\|_2^2 \geqslant \frac{|a^{\mathrm{T}} b|^2}{\|b\|_2} \Rightarrow \|a\|_2 \, \|b\|_2 \geqslant |a^{\mathrm{T}} b|.$$

矩阵范数的定义

矩阵范数的定义是将向量范数的定义推广至$\mathbb{R}^{m\times n}$所直接形成的.

定义

令记号$\|\cdot\|:\mathbb{R}^{m\times n}\to\mathbb{R}^{+}$是一种非负函数, 如果它满足:

- 正定性: 对于一切定义域内的$A\in\mathbb{R}^{m\times n}$, 有$\|A\|\geqslant 0$,
  且$\|A\|=0\Leftrightarrow A=0_{m\times n}$;
- 齐次性: 对于一切$A\in\mathbb{R}^{m\times n}$和$\alpha\in\mathbb{R}$, 有$\|\alpha A\|=|\alpha|\,\|A\|$;
- 三角不等式: 对于一切$A,B\in\mathbb{R}^{m\times n}$, 均成立$\|A+B\|\leqslant\|A\|+\|B\|$.

则称$\|\cdot\|$是定义在向量空间$\mathbb{R}^{m\times n}$上的矩阵范数.

矩阵的ℓ_p范数

和向量的ℓ_p范数类似，矩阵的ℓ_p范数也与矩阵中各分量的信息密切相关．在接下来的定义中，我们认为a_{ij}是矩阵A的第i行j列的分量．

- $p=1$时，$\|A\|_1=\sum_{i=1}^m\sum_{j=1}^n|a_{ij}|$.
- $p=2$时，又称此时定义的"2-范数"为Frobenius范数(F-范数)，记为$\|A\|_F$，具体形式为$\|A\|_F=\sqrt{\mathbf{Tr}\left(AA^\mathrm{T}\right)}=\sqrt{\sum_{i=1}^m\sum_{j=1}^n a_{ij}^2}$.

· F-范数具有正交不变性．

定理

对于任意的正交矩阵$\mathcal{U}\in\mathbb{R}^{m\times m},\mathcal{V}\in\mathbb{R}^{n\times n}$，成立$\|\mathcal{U}A\mathcal{V}\|_F^2=\|A\|_F^2$.

请自行证明上述定理．
提示：$\mathbf{Tr}\left(AB\right)=\mathbf{Tr}\left(BA\right)$.

矩阵的内积

对于空间$\mathbb{R}^{m\times n}$的两个矩阵A, B, 除了定义它们各自的范数以外, 我们还可以定义它们之间的内积.

定义

Frobenius内积 矩阵A, B的内积定义为

$$\langle A, B\rangle = \operatorname{Tr}\left(AB^{\mathrm{T}}\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ij}.$$

Frobenius内积实际上就是矩阵逐分量相乘再求和, 因此满足内积的定义.

实际上, 类似向量的内积可表征向量间的夹角, 矩阵的内积也可表征矩阵间的夹角.

明确指出了矩阵内积的定义后，可以类比向量给出矩阵的F-范数对应的Cauchy不等式.

定理

设$A, B \in \mathbb{R}^{m\times n}$，则

$$|\langle A, B\rangle| \leqslant \|A\|_F \, \|B\|_F \,,$$

等号成立当且仅当A和B线性相关.

矩阵的算子范数

算子范数是一类特殊的矩阵范数，它是由向量范数所诱导的. 不过，我们对"诱导"不应感到陌生，因为前言所述向量范数中的A-范数即是由正定矩阵诱导的范数.

定义

给定矩阵$A \in \mathbb{R}^{m\times n}$，以及$\mathbb{R}^m$中的向量范数$\|\cdot\|_{(m)}$和$\mathbb{R}^n$中的向量范数$\|\cdot\|_{(n)}$，其诱导的矩阵范数形式为

$$\|A\|_{(m,n)} = \max_{x\in\mathbb{R}^n, \|x\|_{(n)}=1} \|Ax\|_{(m)} .$$

如果将$\|\cdot\|_{(m)}$和$\|\cdot\|_{(n)}$均取为相应向量空间的ℓ_p范数，我们可以定义矩阵的p范数. 请注意，矩阵的p范数和上一页中定义的ℓ_p范数是不同的.

特殊算子范数: p范数

根据矩阵p范数的定义, 我们可以得到结论:

- $p=1$时, $\|A\|_{p=1}=\max\limits_{\|x\|_1=1}\|Ax\|_1=\max\limits_{1\leqslant j\leqslant n}\sum_{i=1}^{m}|a_{ij}|$.
- $p=2$时, $\|A\|_{p=2}=\max\limits_{\|x\|_2=1}\|Ax\|_2=\sqrt{\lambda_{A^{\mathrm{T}}A}}$. 其中, $\lambda_{A^{\mathrm{T}}A}$是$A^{\mathrm{T}}A$的最大特征值. 此时, 又称该范数为$A$的谱范数.
- $p=\infty$时, $\|A\|_{p=\infty}=\max\limits_{\|x\|_\infty=1}\|Ax\|_\infty=\max\limits_{1\leqslant i\leqslant m}\sum_{j=1}^{n}|a_{ij}|$.

矩阵谱范数和其他p范数的性质

矩阵的谱范数虽然看似是最不好计算的(要求特征值), 但是用谱范数具有诸多的不变性质和估计性质, 因此经常被应用于理论分析中.

我们以定理的形式分别给出几个重要的情形, 然后请读者作为习题自行证明.

定理

设$A \in \mathbb{R}^{n\times n}$, 则

- $\|A\|_2^2 = \|A^{\mathrm{T}}\|_2^2 = \|A^{\mathrm{T}}A\|_2 = \|AA^{\mathrm{T}}\|_2$.
- 对于任意n阶酉矩阵C, D均有$\|CA\|_2 = \|AD\|_2 = \|CAD\|_2 = \|A\|_2$.

定理

设$A \in \mathbb{R}^{n\times n}$, 则

- $\|A\|_2 = \max\limits_{\|x\|_2=1, \|y\|_2=1} |x^{\mathrm{T}}Ay|$.
- $\|A\|_2^2 \leqslant \|A\|_1 \|A\|_\infty$.

矩阵范数的相容性

回顾矩阵范数的定义，我们即可得到一个不等式

$$\|Ax\|_{(m)} \leqslant \|A\|_{(m,n)} \|x\|_{(n)} .$$

这揭示了矩阵范数具有相容性.

更具体地，当$m = n = 2$时，关于$p = 2$的矩阵范数具有相容性$\|Ax\|_2 \leqslant \|A\|_2 \|x\|_2$.

要注意的是，这里我们仅说明了$p = 2$时的算子范数与$p = 2$时的向量距离范数是相容的，并没有指出所有符合定义的矩阵范数均与任何指定的向量范数都相容. 读者要注意推导的因果关系，并在今后的应用中严格遵循.

矩阵范数的自相容性

在讨论矩阵范数时, 我们更多关注的是其自相容性.

定义

如果对于可乘的有限维矩阵A, B, 成立$\|AB\| \leqslant \|A\|\,\|B\|$, 则称矩阵范数$\|\cdot\|$是自相容的.

自相容性并不是一般的矩阵范数所具有的.

例 定义矩阵范数$\|A\| \triangleq \max\limits_{i,j}|a_{ij}|$, 证明其并不是自相容的矩阵范数.

Proof: 只需取得$A = B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, 则不等式不成立.

请读者自行验证以下的命题是正确的. 提示: 利用Cauchy不等式.

定理

- 矩阵的ℓ_1范数是自相容的.
- 矩阵的ℓ_2范数是自相容的.

矩阵范数的等价性

本页将简单叙述一个在泛函分析中特别重要的理论, 即同一矩阵空间内, 矩阵范数彼此之间是相互等价的. 事实上, 在泛函分析中, 这一结论可以拓展至任意的有限维向量空间中(要存在范数).

定理

对于定义在$\mathbb{R}^{n\times n}$上的任意两个矩阵范数$\|\cdot\|_\alpha, \|\cdot\|_\beta$, 总是存在正的常数$c_1, c_2$, 使得

$$\|\cdot\|_\alpha \leqslant c_1 \|\cdot\|_\beta, \|\cdot\|_\beta \leqslant c_2 \|\cdot\|_\alpha .$$

这一结论我们在这里先并不要求证明(但以目前所学的知识确实可以证明), 因为在泛函分析的课程中, 我们将对其进行更具思想价值的证明. 不过, 我们需要对这一结论有所了解.

核范数

上面介绍的范数是我们在数值分析或泛函分析中最常见的范数形式. 在最优化理论分析中, 我们另外规定核范数以衡量矩阵的秩的大小.

定义

给定矩阵$A \in \mathbb{R}^{m\times n}$, 其核范数定义为

$$\|A\|_* = \sum_{i=1}^{r} \sigma_i,$$

其中$\sigma_i (i = 1, \cdots, r)$为$A$的所有非零奇异值, $r = \mathbf{rank}\,(A)$.

· 请读者验证如上规定的核范数是满足三角不等式的.

- 想一想: 形式$\|A\| = \sum_{i=1}^{r} |\sigma_i|$可以成为$\mathbb{R}^{m\times n}$空间中的合法范数吗? 为什么?