数据科学与工程数学基础

作业提交规范及第 4 次作业

教师：黄定江
助教：陈诺、刘文辉

2022 年 3 月 9 日

作业提交规范

1. 作业提交形式：**练习本或笔记本**（建议统一使用一般的**练习本**即可，不接收以纸张的方式书写的作业）。
2. 作业书写说明：
   (a) 可以讨论，禁止抄袭！
   (b) 练习本封面至少包含两方面信息：**姓名**和**学号**
   (c) 每一次的作业**请另起一页**，并在**第一行标明第几次作业**。例如“第 4 次作业”；
   (d) 每一题请**标注题号**，无需抄题，直接解答；
   (e) 题与题之间**请空一行**；
   (f) 不要求字好，但要求书写整体清晰易读。
3. 作业提交途径：纸质作业交给**学习委员**，由学习委员**按学号顺序**收齐后统一在截止日期前交到**助教实验室**。**单数周**布置的作业交到助教刘文辉处**数学馆西 109**；**双数周**布置的作业交到助教陈诺处**地理馆 353**。
4. 作业评分说明：正常提交作业的按照实际评分记录；逾期补交作业的根据逾期情况在实际评分基础上酌情扣分；未交作业的当次作业记为 0 分。

第 4 次作业

提交截至时间：**2022/03/11 下周五 20:00（晚上）**

理论部分 (范数与二次型)

习题 1. 对偶范数常在共轭函数及一些不等式中出现。向量的对偶函数定义为

$$\|z\|_* = \sup\left\{z^\top x \mid \|x\| \leq 1\right\}$$

若向量范数 l_p 与 l_q 互为对偶范数，则 $p, q \in R^n$ 需满足 $\frac{1}{p} + \frac{1}{q} = 1$ *(称 p, q 为 Hölder 共轭)*

(1) 试用对偶范数定义及上述性质证明 Hölder 不等式: 对 $p > 1, \frac{1}{p} + \frac{1}{q} = 1$, 以及 $x, y \in R^n$

$$\sum_{i=1}^{n}|x_i y_i| \leq \left(\sum_{i=1}^{n}|x_i|^p\right)^{1/p}\left(\sum_{i=1}^{n}|y_i|^q\right)^{1/q}$$

习题 2. 矩阵的范数主要包括三种主要类型：诱导范数，元素形式范数和 *Schatten* 范数。诱导范数又称矩阵空间上的算子范数 *(operator norm)*，常用的诱导范数为 *p* 范数, 定义如下

$$\|A\|_p = \sup_{\|x\|_p \neq 0}\frac{\|Ax\|_p}{\|x\|_p} = \sup_{\|x\|_p = 1}\|Ax\|_p$$

(1) 设 $A = (a_{ij}) \in C^{m\times n}$, 证明 1 范数为列和范数，无穷范数为行和范数

$$\|A\|_\infty = \max_{1\leqslant i\leqslant m}\sum_{j=1}^{n}|a_{ij}|\,, \ \|A\|_1 = \max_{1\leqslant j\leqslant n}\sum_{i=1}^{m}|a_{ij}|$$

元素形式范数即矩阵按列排成向量，然后采用向量范数的定义得到的矩阵范数，一般称 l_p 范数。

$$l_p : \|A\|_p = \sqrt[P]{\sum_{i,j}|a_{ij}|^P}$$

(2) 试比较 l1 范数

$$l_1 : \|A\|_1 = \sum_{i,j}|a_{ij}|^1$$

与诱导范数的关系

习题 3. *(作为阅读材料, 不计入分数)*

由 $\|A\|_p = \sup_{\|x\|_p \neq 0}\frac{\|Ax\|_p}{\|x\|_p}$,

矩阵 *A* 的诱导范数可理解为线性变换 *Ax* 对向量 *x* 的最大“拉长倍数”

由 $\|A\|_p = \sup_{\|x\|_p = 1}\|Ax\|_p$,

矩阵 *A* 的诱导范数也可理解为 *x* 在单位范数球上运动时 $\|Ax\|_p$ 的最大值

以下情形可便于理解诱导范数

例: 左图为 $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $p(A) = 1, p = 2$ 时的情形，在 $x = (\sqrt{2}/2, \sqrt{2}/2)$ 时 $\|Ax\|_2$ 取到最大值 2

例: 右图绿线为 $A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, $p = 1$ 时的情形，在 $x = (2, 0)$ 时 $\|Ax\|_1$ 取到最大值 2

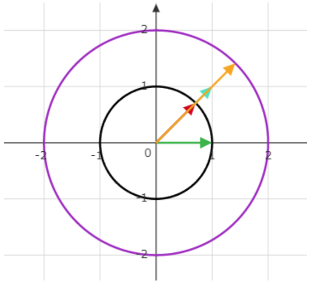

图 1: 诱导范数例 (p=2)

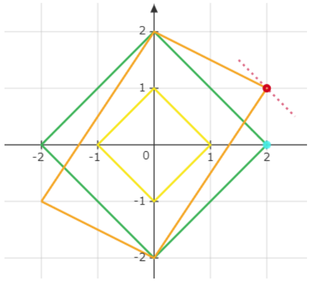

图 2: 诱导范数例 (p=1)

例: 右图橙线为 $A=\begin{pmatrix}2 & 0\\ 1 & 2\end{pmatrix}$, $p=1$ 时的情形，在 $x=(2,1)$ 时 $\|Ax\|_1$ 取到最大值 3

(1) 试证明

$$||A||_2=\sigma_{max}=\sqrt{\lambda\left(A^{\top}A\right)}$$

, 其中 σ_{max} 为谱范数，即矩阵 A 的最大奇异值，$\lambda\left(A^{\top}A\right)$ 表示 $A^{\top}A$ 的最大特征值。

证: 即证明对于矩阵 $A_{m\times n}$, 对任意向量 x, 在矩阵 A 的变换 (即 Ax) 后, 其长度不大于 $\sigma_{max}\|x\|_2$, 即 $\|Ax\|_2\le\sigma_{max}\|x\|_2$ 。

$A^{\top}A$ 是实对称阵,其特征向量两两正交。不妨令 $B=A^{\top}A$, 特征向量矩阵为 $\lambda=diag(\lambda_1,...,\lambda_n)=\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix}$, $p_1,...,p_n$ 为 B 的一组标准正交特征向量，则 $P=(p_1,...,p_n)$ 为正交矩阵, 故

$$BP=P\lambda\Leftrightarrow B=P\lambda P^{-1}\Leftrightarrow B=P\lambda P^{\top}$$

(称 λ 合同于 B)

假设对一个向量 x, 在矩阵 A 的变换（即 Ax ）后得到 y, 即满足 $y=Ax$。则

$$\|y\|_2^2=y^{\top}y=(Ax)^{\top}(Ax)=x^{\top}A^{\top}Ax=x^{\top}Bx=x^{\top}P\lambda P^{\top}x=(P^{\top}x)^{\top}\lambda P^{\top}x$$

对于二次型，可以看作是一个二次齐次多项式的图形，而正交矩阵 P 的变换，可以保证变换图形的形状和大小不变，仅仅做了位移、旋转或翻转的变换，类似把物体从一个地方移到另一个地方。(可以想象一个三维坐标系，在坐标系上的点构成的图形通过一个非正交的基表示，现坐标系换了一组标准正交基 P，用这组基变换图形不过是移动 (掰正) 了图形的位置。记 $z=P^{\top}x$, 所以 z 不过是一个与 x 一样的（同范数的）向量，只是换了位置，而 $\lambda=diag(\lambda_1,...,\lambda_n)$ 则进行了掰正位置后的放缩。因而

$$
\begin{aligned}
|y|^2 = z^T \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} z &= \sum_i \lambda_i z_i^2 \\
&= \sum_i \left(\sqrt{\lambda_i} z_i\right)^2 \leq \sum_i \left(\max_j \left(\sqrt{\lambda_j}\right) z_i\right)^2 \\
&= \max\left(\sqrt{\lambda_j}\right)^2 \sum_i z_i^2 = \max\left(\lambda_j\right)|z|^2 == \max\left(\lambda_j\right)|x|^2
\end{aligned}
$$

当且仅当除 $z_{opt\,\max_j(\lambda_j)}$ 以外的其他元素均等于 *0* 时，该不等式的等号成立。

即 $\|Ax\|_2 \leq \sigma_{max}\|x\|_2$ ，证毕

(2) 元素形式下矩阵的 l_2 范数称为 *Frobenius* 范数，即

$$l_2 : \|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2}$$

试比较 $\|A\|_2$ 与 $\|A\|_F$ 的大小

答:$\|A\|_2 \leq \|A\|_F$

习题 4. 上题提到矩阵 A 的诱导范数也可理解为 x 在单位范数球上运动时 $\|Ax\|_p$ 的最大值，而对于某些矩阵，x 的巨大变化只能引起 Ax 很小的变化（旋转而非放缩）。反之，对于线性方程组 $Ax = b$，b（或 A）的微小变化会带来解 x 的巨大变化。这样的矩阵 A（及线性方程组）称作病态的。如 $A = \begin{pmatrix} 1 & 1 \\ 1 & 1.0001 \end{pmatrix}$

$$\begin{pmatrix} 1 & 1 \\ 1 & 1.0001 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 \\ 1 & 1.0001 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2.0001 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$\|\delta b\|_\infty = \left\|(2,2)^{\mathrm{T}} - (2,2.0001)^{\mathrm{T}}\right\|_\infty = 0.0001, \|\delta x\|_\infty = \left\|(2,0)^{\mathrm{T}} - (1,1)^{\mathrm{T}}\right\|_\infty = 1$，方程组解 x 的变化程度是 b 变化程度的 *10000* 倍，因此称矩阵 A 是病态的。

(1) 试推测什么样的矩阵是病态矩阵，用矩阵范数表示。

提示：对线性方程组 $Ax = b$，设 A 是精确的，b 有微小的扰动 δb，新方程组的解为 $x + \delta x$，即 $A(x + \delta x) = b + \delta b$，用

$$Ax = b \Rightarrow \|A\|\,\|x\| \geq \|b\|$$

$$A\delta x = A\delta b, \delta x = A^{-1}\delta b \Rightarrow \|\delta x\| = \|A^{-1}\delta b\| \leq \|A^{-1}\|\|\delta b\|$$

证明

$$\frac{\|\delta x\|}{\|x\|} \leq \left\|A^{-1}\right\|\|A\|\frac{\|\delta b\|}{\|b\|}$$