

数据科学与工程数学基础

作业提交规范及第 1 次作业

教师：黄定江

助教：刘文辉、徐艺玮

2023 年 10 月 3 日

作业提交规范

1. 作业提交形式：建议使用 Word 或 \LaTeX 编写所得到的电子文档。若使用 Word 编写，将其另存为 PDF 形式，然后提交 PDF 文档。若使用 \LaTeX 编写，将其编译成 PDF 形式，然后提交 Tex 和 PDF 两个文档。若使用手写拍照的方式，请务必规整好各题解答的图片，并整合在一个 PDF 文档中，只发图片格式的作业概不批改！
2. 作业命名规范：提交的电子文档必须命名为：“学号_姓名”。命名示例：52200000000_刘某某。
3. 作业提交途径：点击打开每次作业的传送门地址：**第 1 次作业提交传送门**，无需注册和登录，按要求输入个人学号和姓名，然后上传作业文档即可。注意：传送门将会在截至时间点到达后自动关闭。
4. 作业更改说明：如果需要修改已经提交的作业，只要在截至日期前，再次上传更改后的作业（切记保持同名），即可覆盖已有作业。
5. 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；若发现存在抄袭的作业时，相似的两份作业均会被记为 0 分。

第 1 次作业



提交截至时间：**2023/09/26 下周二 12:00（中午）**

理论部分 (范数与二次型)

习题 1. 请证明: 对任意 $A \in \mathbb{R}^{m \times n}$, 由

$$\|A\|_{m_\infty} := \max_{1 \leq i \leq m, 1 \leq j \leq n} |a_{ij}|$$

定义的 $\|\cdot\|_{m_\infty}$ 是 $\mathbb{R}^{m \times n}$ 上的 (广义) 矩阵范数。

解. 非负性: 显然

$$\|A\|_{m_\infty} := \max_{1 \leq i \leq m, 1 \leq j \leq n} |a_{ij}| \geq 0,$$

而且仅当 $A = 0$ 时, $\|A\|_{m_\infty} = 0$.

齐次性:

$$\|c \cdot A\|_{m_\infty} := \max_{1 \leq i \leq m, 1 \leq j \leq n} |c \cdot a_{ij}| = c \max_{1 \leq i \leq m, 1 \leq j \leq n} |a_{ij}| = c \|A\|_{m_\infty}$$

三角不等式: 考虑 $\|A + B\|_{m_\infty}$, 因为

$$|a_{ij} + b_{ij}| \leq |a_{ij}| + |b_{ij}| \leq \max_{1 \leq i \leq m, 1 \leq j \leq n} |a_{ij}| + \max_{1 \leq i \leq m, 1 \leq j \leq n} |b_{ij}|$$

所以

$$\max_{1 \leq i \leq m, 1 \leq j \leq n} |a_{ij} + b_{ij}| \leq \max_{1 \leq i \leq m, 1 \leq j \leq n} |a_{ij}| + \max_{1 \leq i \leq m, 1 \leq j \leq n} |b_{ij}|$$

即 $\|A + B\|_{m_\infty} \leq \|A\|_{m_\infty} + \|B\|_{m_\infty}$

习题 2. 设 a_1, a_2, \dots, a_n 是 n 个正数, 证明: 由

$$\Omega(\mathbf{x}) = \left(\sum_{i=1}^n a_i x_i^2 \right)^{\frac{1}{2}}$$

定义的函数 $\Omega: \mathbb{R}^n \rightarrow \mathbb{R}$ 是一个范数。

解. 同习题 1 根据定义证明: 非负性、齐次性和三角不等式。

习题 3. 求下面矩阵的 1-范数、2-范数和无穷范数:

$$A_1 = \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}, A_2 = \begin{pmatrix} -1 & 0 \\ 1 & 2 \end{pmatrix}.$$

解. $A_1^T A_1 = \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}$

$$\lambda_{\max}(A_1^T A_1) = 3 + \sqrt{5}$$

$$\|A_1\|_1 = 2, \|A_1\|_2 = \sqrt{3 + \sqrt{5}}, \|A_1\|_\infty = 3$$

$$A_2^T A_2 = \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}$$

$$\|A_2\|_1 = 2, \|A_2\|_2 = \sqrt{3 + \sqrt{5}}, \|A_2\|_\infty = 3$$

习题 4. 证明: $\|AB\|_F \leq \|A\|_F \|B\|_F$ 和 $\|AB\|_F \leq \|A\|_F \|B\|_2$.

解. 这道题可以不改已经在习题课给大家讲解了。提示: 可根据矩阵分块的方式证明。

习题 5. 阅读完以下补充材料即可:

将一个带有交叉项的二次型转变成没有交叉项的二次型等同于将对称矩阵合同变换为一个对角矩阵。它的几何意义, 可以看作是将二次方程所对应的几何图形标准化的过程。例如: 在平面上, 即椭圆、抛物线和双曲线等二次曲线的标准化过程; 在空间上, 即二次曲面的标准化过程。下面通过主轴定理及其几何意义窥见一斑。

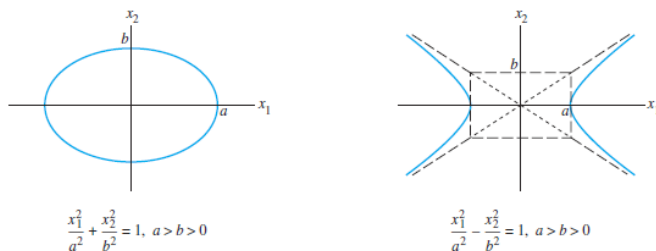


图 1: 对角矩阵时对应标准二次曲线

• 主轴定理

定理 0.0.1. 令 A 是任意一个 $n \times n$ 的对称矩阵, 那么存在一个正交变换 $x = Py$ 使得二次型 $x^T Ax$ 转变为 $y^T Dy$, 其中 $D = P^T AP$ 是对角矩阵。

定理中 P 的列称为二次型 $x^T Ax$ 的主轴。向量 y 是向量 x 在这些主轴下的坐标。

• 主轴定理的几何意义

假设 $Q(x) = x^T Ax$, 其中 $x \in \mathbb{R}^2$, A 是 2×2 可逆对称矩阵, c 是常数。可以证明 \mathbb{R}^2 中所有满足 $x^T Ax = c$ 的 x 的集合可能为椭圆 (圆)、双曲线、两相交直线、点或者空集。如果 A 是对角矩阵, 那么二次曲线将是标准的二次曲线, 如图 1 所示。如果 A 不是对角矩阵, 那么二次曲线将是非标准的二次曲线。图 2 是二次方程为 $5x_1^2 - 4x_1x_2 + 5x_2^2 = 48$ 的椭圆曲线。显然, 它不是一个标准的二次曲线。下面, 我们通过寻找该椭圆的主轴将其转化为标准形式, 从几何的角度验证主轴定理。

易求得该椭圆方程对应二次型的矩阵表示为

$$A = \begin{pmatrix} 5 & -2 \\ -2 & 5 \end{pmatrix}.$$

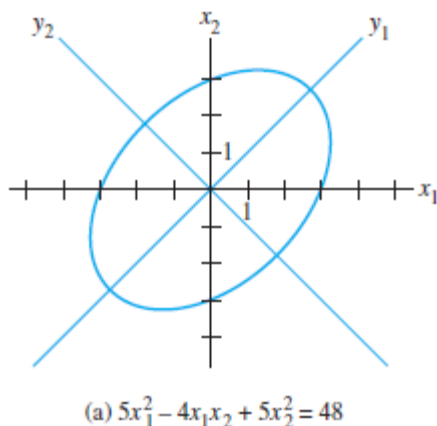


图 2: 非对角矩阵对应非标准二次曲线

可计算出其特征值为 3 和 7, 对应的特征向量分别为

$$v_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}, v_2 = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}.$$

令 $P = [v_1 \ v_2]$, 则从图形中可以看出 P 的列向量, 正好对应该椭圆主轴所在的方向。现将 $x = Py$ 代入二次型, 可得

$$y^T D y = 3y_1^2 + 7y_2^2.$$

即可得到椭圆的标准方程形式。

习题 6. 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 l 范数为列和范数, 无穷范数为行和范数

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

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 l_1 范数

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

与诱导范数的关系

解. (I)

$$\begin{aligned}
 A &= (a, \dots, a_n) \\
 \|Ax\|_1 &= \left\| \sum_i a_i x_i \right\|_1 \\
 &\leq \sum_i \|a_i x_i\|_1 \\
 &= \sum_i \|x_i\| \|a_i\|_1 \\
 &\leq (\max_i \|a_i\|_1) \left(\sum_i |x_i| \right) \\
 &= \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}| \|x\|_1 \\
 \|Ax\|_\infty &= \max_i \left| \sum_j a_{ij} x_j \right| \\
 \left| \sum_j a_{ij} x_j \right| &\leq \sum_j |a_{ij} x_j| \\
 &\leq \sum_j |a_{ij}| \max_j |x_j| \\
 \|Ax\|_\infty &\leq \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}| \|x\|_\infty
 \end{aligned}$$

(2) 以 l_1 范数与 l 范数, 无穷范数为例, 有

$$\begin{aligned}
 \|X\|_1 &\leq \|X\|_1(l_1) \leq n \|X\|_1 \\
 \|X\|_\infty &\leq \|X\|_1(l_1) \leq m \|X\|_\infty
 \end{aligned}$$

习题 7. 阅读完以下补充材料即可:

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

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”

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

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

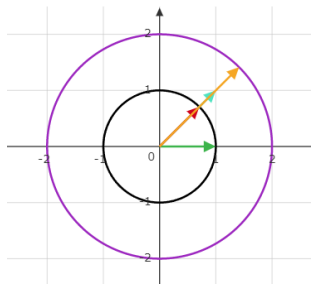


图 3: 诱导范数例 ($p=2$)

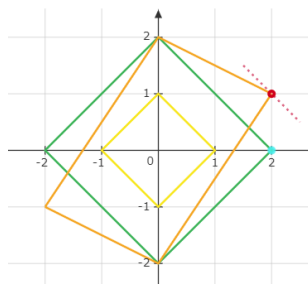


图 4: 诱导范数例 ($p=1$)

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

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

(I) 试证明

$$\|A\|_2 = \sigma_{max} = \sqrt{\lambda(A^T A)}$$

, 其中 σ_{max} 为谱范数, 即矩阵 A 的最大奇异值, $\lambda(A^T A)$ 表示 $A^T A$ 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 $\|Ax\|_2 \leq \sigma_{max}\|x\|_2$ 。

$A^T A$ 是实对称阵, 不同特征值对应的特征向量两两正交。不妨令 $B = A^T A$, 特征向量矩阵为

$$P = \text{diag}(\lambda_1, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}, p_1, \dots, p_n \text{ 为 } B \text{ 的一组标准正交特征向量, 则 } P =$$

(p_1, \dots, p_n) 为正交矩阵, 故

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

(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

$$\|y\|_2^2 = y^T y = (Ax)^T (Ax) = x^T A^T A x = x^T B x = x^T P \lambda P^T x = (P^T x)^T \lambda P^T x$$

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 (矫正) 了图形的位置。记 $z = P^T x$,

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 $\lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ 则进行了掰正位置后的放缩。因而

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

当且仅当除 $z_{\text{opt max}_j(\lambda_j)}$ 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: $\|A\|_2 \leq \|A\|_F$