数据科学与工程数学基础

作业提交规范及第 7 次作业

教师：黄定江
助教：刘文辉、徐艺玮

2024 年 1 月 8 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 7 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 7 次作业

提交截至时间：**2024/01/08 下周一 12:00（中午）**

理论部分

习题 1. 写出下述非线性规划的 *KKT* 条件并求解

$$
\begin{aligned}
(1)\quad & \mathtt{maximize} \quad f(x)=(x-3)^2\\
& \mathtt{suject\ to} \quad 1\le x\le 5\\
(2)\quad & \mathtt{minimize} \quad f(x)=(x-3)^2\\
& \mathtt{suject\ to} \quad 1\le x\le 5
\end{aligned}
$$

解. *(1)* 原问题等价于

$$
\begin{cases}
\mathtt{minimize} \quad -f(x)=(x-3)^2\\
g_1(x)=-x+1\le 5\\
g_2(x)=x-5\le 0
\end{cases}
$$

求目标函数和约束函数的梯度得,

$$\nabla_x f(x)=-2(x-3),\nabla_x g_1(x)=-1,\nabla_x g_2(x)=1$$

将约束引入广义 *Lagrange* 乘子 v_1,v_2, 在 *KKT* 条件上有

$$
\begin{cases}
-2(x^*-3)-v_1^*+v_2^*=0\\
v_1^*(-x^*+1)=0\\
v_2^*(x^*-5)=0\\
v_1^*\ge 0,v_2^*\ge 0
\end{cases}
$$

若 $v_1^*\ne 0,v_2^*\ne 0$, 无解.

若 $v_1^*=0,v_2^*\ne 0$, 得 $x^*=5,v_2^*=4,-f(x^*)=-4$.

若 $v_1^*\ne 0,v_2^*=0$, 得 $x^*=1,v_1^*=4,-f(x^*)=-4$.

若 $v_1^*=0,v_2^*=0$, 得 $x^*=3,f(x^*)=0$.

因此最优点 $x^*=1$ 或 $x^*=5$, $\mathtt{maximize} f(x)=4$.

(2) 原问题等价于

$$
\begin{cases}
\mathtt{minimize} \quad f(x)=(x-3)^2\\
g_1(x)=-x+1\le 5\\
g_2(x)=x-5\le 0
\end{cases}
$$

求目标函数和约束函数的梯度得,

$$\nabla_x f(x)=-2(x-3),\nabla_x g_1(x)=-1,\nabla_x g_2(x)=1$$

将约束引入广义 *Lagrange* 乘子 v_1,v_2, 在 *KKT* 条件上有

$$
\begin{cases}
2(x^*-3)-v_1^*+v_2^*=0\\
v_1^*(-x^*+1)=0\\
v_2^*(x^*-5)=0\\
v_1^*\ge 0,v_2^*\ge 0
\end{cases}
$$

若 $v_1^* \neq 0, v_2^* \neq 0$, 无解.
若 $v_1^* = 0, v_2^* \neq 0$, 得 $x^* = 5, v_2^* = -4 < 0$, 不是 *KKT* 点.
若 $v_1^* \neq 0, v_2^* = 0$, 得 $x^* = 1, v_1^* = -4 < 0$, 不是 *KKT* 点.
若 $v_1^* = 0, v_2^* = 0$, 得 $x^* = 3, f(x^*) = 0$.
因此最优点 $x^* = 3$, $minimize f(x) = 0$.

习题 2. 考虑等式约束的最小二乘问题

$$\begin{aligned} &minimize \quad \|\mathbf{Ax} - \mathbf{b}\|_2^2 \\ &suject\ \ to \quad \mathbf{Gx} = \mathbf{h} \end{aligned}$$

其中 $\mathbf{A} \in \mathbb{R}^{m\times n}, \text{rank}(\mathbf{A}) = n, \mathbf{G} \in \mathbb{R}^{p\times n}, \text{rank}(\mathbf{G}) = p$. 给出 *KKT* 条件, 推导原问题最优解 x^* 以及对偶问题最优解 v^* 的表达式.

解. 求得 *Lagrangian* 函数为

$$\begin{aligned} L(x, v) =& \|Ax - b\|_2^2 + v^T(Gx - h) \\ =& x^T A^T A x + (G^T v - 2A^T b)^T x - v^T h \end{aligned}$$

可通过如下最优性条件得到函数最小值. 令梯度为 *0* 得,

$$\nabla_x L(x, v) = 2A^T A x + G^T v - 2A^T b = 0$$

因此当 $x = \frac{1}{2}(A^T A)^{-1}(G^T v - 2A^T b)$ 时, *Lagrangian* 函数取得最小值.
对偶函数为 $g(x) = -\frac{1}{4}(G^T v - 2A^T b)^T (A^T A)^{-1} (G^T v - 2A^T b) - v^T h$.
最优性条件为

$$\begin{cases} 2A^T(Ax^* - b) + G^T v^* = 0 \\ Gx^* = h \end{cases}$$

解方程得,

$$\begin{cases} v^* = 2(G(A^T A)^{-1} G^T)^{-1}(G(A^T A)^{-1} A^T b - h) \\ x^* = (A^T A)^{-1}(A^T b - G^T(G(A^T A)^{-1} G^T)^{-1}(G(A^T A)^{-1} A^T b - h)) \end{cases}$$

习题 3. 用 *Lagrange* 乘子法证明：矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 的 *2* 范数

$$\|\boldsymbol{A}\|_2 = \max_{\|\boldsymbol{x}\|_2 = 1, \boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{A}\boldsymbol{x}\|_2$$

的平方是 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的最大特征值。

解. 优化问题为

$$\begin{aligned} &maximize \quad f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \\ &suject\ \ to \quad \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = 1 \end{aligned}$$

Lagrange 函数为：

$$L(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \lambda(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - 1)$$

$$\frac{\partial L}{\partial \boldsymbol{x}} = 2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - 2\lambda\boldsymbol{x}$$

令 $\frac{\partial L}{\partial \boldsymbol{x}} = 0$，有：

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \lambda$$

这表示在 $f(\boldsymbol{x})$ 的极大值点，$\boldsymbol{x}$ 是 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的特征向量，λ 是对应的特征值。此时，

$$f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\lambda\boldsymbol{x} = \lambda\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \lambda$$

因此说明，为使 $f(\boldsymbol{x})$ 最大，$f(\boldsymbol{x}) = \lambda_{max}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$，其中 λ_{max} 表示最大特征值。即

$$\|\boldsymbol{A}\|_2^2 = \lambda_{max}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$$

习题 4. 用 *Lagrange* 乘子法求欠定方程 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 的最小二范数解，其中 $\boldsymbol{A} \in \mathbb{R}^{m\times n}, m \leq n, \mathrm{rank}(\boldsymbol{A}) = m$

解. 优化问题为

$$\mathtt{minimize} \quad f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x}\|_2^2$$

$$\mathtt{suject\ to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

Lagrange 函数为：

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$$

$$\frac{\partial L}{\partial \boldsymbol{x}} = \boldsymbol{x} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}$$

令 $\frac{\partial L}{\partial \boldsymbol{x}} = 0$，有：

$$\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}$$

$$g(\boldsymbol{\lambda}) = -\frac{1}{2}\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\lambda + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{b}$$

令 $\frac{\partial g}{\partial \boldsymbol{\lambda}} = 0$：

$$-\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\lambda + \boldsymbol{b} = 0$$

由 $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \mathrm{rank}(\boldsymbol{A}) = m$ 得 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 可逆，因此

$$\boldsymbol{\lambda} = (\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{b}$$

因此，$\boldsymbol{x}$ 满足 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 的最小二范数解：

$$\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{b}$$

习题 5. 用最速下降法和精确线搜索计算 $\min f(x) = x_1^2 + x_2^2 + x_3^2$ ，初始点 $x^{(0)} = (2, 2, 1)^T$. 当 $(f(x^{(n+1)}) - f(x^{(n)})) < 0.001$ 时迭代终止.

解. 由题意得, $f(x)=x^Tx, \nabla_x f(x)=2x$, 设最速下降法的步长为 λ, 那么

$$\begin{aligned} f(x-\lambda\nabla_x f(x)) =&(x-\lambda\nabla_x f(x))^T(x-\lambda\nabla_x f(x)) \\ =&x^Tx-2\lambda\nabla_x f(x)^Tx+\lambda^2\nabla_x f(x)^T\nabla_x f(x) \end{aligned}$$

在 $x-\lambda\nabla_x f(x)$ 方向上, 使 $f(x)$ 最小的 λ 满足

$$\frac{\partial f(x-\lambda\nabla_x f(x))}{\partial\lambda}=-2\nabla_x f(x)^Tx+2\lambda\nabla_x f(x)^T\nabla_x f(x)$$

得

$$\lambda=\frac{\nabla_x f(x)^Tx}{\nabla_x f(x)^T\nabla_x f(x)}=\frac{1}{2}$$

所以,

$$x^{(1)}=x^{(0)}-\frac{1}{2}\nabla_x f(x^{(0)})=(0,0,0)^T$$
$$f(x^{(1)})=0$$
$$x^{(2)}=x^{(1)}-\frac{1}{2}\nabla_x f(x^{(1)})=(0,0,0)^T$$
$$f(x^{(2)})=0$$

同理可得, $f(x^{(n)})=0(n>0)$, 因此当 $|(f(x^{(n+1)})-f(x^{(n)})|=0<0.001$ 时, 迭代终止.

习题 6. 使用梯度下降法和固定步长 $\lambda=0.01$ 计算 $\min f(x)=(x_1-1)^2+16(x_2-2)^2$ ，初始点 $x^{(0)}=(2,3)^T$，迭代两步后终止.

解. 具体迭代结果：

k	${x^{(k)}}^{\mathrm{T}}$	g_k^{T}	f_k	$\|g_k\|_\infty$
0	$(2,3)$	$(2,32)$	17	32
1	$(1.98,2.68)$	$[1.96,21.76]$	8.3588	21.76
2	$(1.96,2.46)$	$[1.9208,14.7968]$	4.3434	14.7968

习题 7. 考虑问题

$$\min f(x)=3x_1^2+3x_2^2-x_1^2x_2.$$

从初始点 $x^{(0)}=(1.5,1.5)^{\mathrm{T}}$ 出发, 用 *Newton* 方法求迭代两步后该问题的解（可用编写程序辅助计算）.

解. $f(x)$ 的一、二阶导数分别为

$$g(x)=\left(6x_1-2x_1x_2,6x_2-x_1^2\right)^{\mathrm{T}},\quad G(x)=\begin{bmatrix} 6-2x_2 & -2x_1 \\ -2x_1 & 6 \end{bmatrix}$$

$f(x)$ 有三个稳定点: 极小点 $x_{(1)} = (0,0)^{\mathrm{T}}$, 鞍点 $x_{(2)} = (3\sqrt{2},3)^{\mathrm{T}}$ 和 $x_{(3)} = (-3\sqrt{2},3)^{\mathrm{T}}$. 在这三个点的 *Hesse* 矩阵分别为

$$G\left(x_{(1)}\right) = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}, \quad G\left(x_{(2)}\right) = \begin{bmatrix} 0 & -6\sqrt{2} \\ -6\sqrt{2} & 6 \end{bmatrix}$$

$$G\left(x_{(3)}\right) = \begin{bmatrix} 0 & 6\sqrt{2} \\ 6\sqrt{2} & 6 \end{bmatrix}.$$

下面我们从 $x^{(0)} = (1.5, 1.5)^{\mathrm{T}}$, 这时 *Newton* 方法在每一迭代步的信息见下表

k	${x^{(k)}}^{\mathrm{T}}$	f_k	$\|g_k\|_\infty$
0	$(1.5000, 1.5000)$	10.1250	8.1125
1	$(-3.7500, -2.2500)$	89.0156	48.0633
2	$(0.6250, -3.1250)$	31.6895	20.6151
3	$(0.3190, 0.0014)$	0.3052	1.9155
4	$(-0.0020, -0.0172)$	0.0009	0.1037
5	$(-0.0000, -0.0000)$	0.0000	0.0000
6	$(-0.0000, -0.0000)$	0.0000	0.0000