# 华东师范大学期末试卷 (A 卷) 答案

## 2023－2024 学年第一学期

考试科目：数据科学与工程数学基础　　任课教师：黄定江

姓　　名：________　　学　　号：________

专　　业：________　　班　　级：________

| 题号 | 一 | 二 | 三 | 四 | 五 | 六 | 七 | 八 | 总分 | 阅卷人签名 |
|---|---|---|---|---|---|---|---|---|---|---|
| 得分 | | | | | | | | | | |

**题 1**　(13’) 完成以下问题：

(1) 对矩阵 $\boldsymbol{M}=\begin{pmatrix}5&3&-1&3\\0&1&1&-2\\-5&-3&4&-4\\0&1&1&0\end{pmatrix}$ 进行 LU 分解；

(2) 设 $A$ 对称且 $a_{11}\neq 0$, 并假经过一步 Gauss 消去之后, $A$ 具有如下形式

$$\begin{bmatrix}a_{11}&a_1^{\mathrm{T}}\\\mathbf{0}&A_2\end{bmatrix}$$

证明：$A_2$ 仍是对称矩阵。

**解**（1）

$$\boldsymbol{A}=\begin{pmatrix}5&3&-1&3\\0&1&1&-2\\-5&-3&4&-4\\0&1&1&0\end{pmatrix}=\begin{pmatrix}1&0&0&0\\0&1&0&0\\-1&0&1&0\\0&1&0&1\end{pmatrix}\begin{pmatrix}5&3&-1&3\\0&1&1&-2\\0&0&3&-1\\0&0&0&2\end{pmatrix}=\boldsymbol{LU}$$

**（7 分）**

（2）记矩阵 $A$ 和高斯变换矩阵分别为：

$$A=\begin{bmatrix}a_{11}&a_1^{\mathrm{T}}\\a_1&A_1\end{bmatrix},G=\begin{bmatrix}1&\mathbf{0}^{\mathrm{T}}\\-l_1&I\end{bmatrix}$$

则

$$GA=\begin{bmatrix}a_{11}&a_1^{\mathrm{T}}\\-a_{11}l_1+a_1&-l_1a_1^{\mathrm{T}}+A_1\end{bmatrix}$$

易知 $-a_{11}l_1+a_1=\mathbf{0}$ 以及 $A_2=-l_1a_1^{\mathrm{T}}+A_1$。由第一个等式可得 $l_1=1/a_{11}a_1$，代入第二个等式知

$$A_2=-\frac{1}{a_{11}}a_1a_1^{\mathrm{T}}+A_1.$$

故可知 $A_2$ 仍是对称矩阵。 **（6 分）**

题 2 (10’) $\mathbb{R}^5$ 的欧氏空间。子空间 $U\subseteq\mathbb{R}^5$ 和 $\boldsymbol{x}\in\mathbb{R}^5$ 如下：

$$U=\textbf{span}\left\{\begin{bmatrix}0\\-1\\2\\0\\1\end{bmatrix},\begin{bmatrix}1\\-1\\1\\-1\\1\end{bmatrix}\right\},\quad \boldsymbol{x}=\begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix}$$

请确定 $\boldsymbol{x}$ 在子空间 $\mathbb{U}$ 上的正交投影 $\pi_{\mathbb{U}}(\boldsymbol{x})$;

**解** 首先，将 $U$ 中的基向量构成矩阵 $A$：

$$A = \begin{bmatrix} 0 & 1 \\ -1 & -1 \\ 2 & 1 \\ 0 & -1 \\ 1 & 1 \end{bmatrix}$$

易计算出

$$A^{\mathrm{T}}A = \begin{bmatrix} 6 & 4 \\ 4 & 5 \end{bmatrix}, \quad (A^{\mathrm{T}}A)^{-1} = \begin{bmatrix} 5/14 & -2/7 \\ -2/7 & 3/7 \end{bmatrix}$$

**（4 分）**

然后，计算投影矩阵 $P_U$：

$$P_U = \begin{bmatrix} 3/7 & -1/7 & -1/7 & -3/7 & 1/7 \\ -1/7 & 3/14 & -2/7 & 1/7 & -3/14 \\ -1/7 & -2/7 & 5/7 & 1/7 & 2/7 \\ -3/7 & 1/7 & 1/7 & 3/7 & -1/7 \\ 1/7 & -3/14 & 2/7 & -1/7 & 3/14 \end{bmatrix}$$

因此，投影

$$\pi_U(x) = P_U x = (-1/7,\ 3/14,\ -2/7,\ 1/7,\ -3/14)$$

**（6 分）**

**题 3** (13’) 完成下列矩阵函数求梯度：

(1) 求函数 $f(\boldsymbol{x}) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$ 和 $g(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ 的 Hessian 矩阵；

(2) 考虑一个两层的全连接神经网络：

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) = \mathrm{ReLU}(\boldsymbol{A}_2(\mathrm{ReLU}(\boldsymbol{A}_1\boldsymbol{x} + \boldsymbol{b}_1)) + \boldsymbol{b}_2)$$

其中

$$A_1=\begin{bmatrix}2&3\\-2&1\\3&-1\end{bmatrix},A_2=\begin{bmatrix}3&-1&0\\1&-2&2\end{bmatrix},b_1=\begin{bmatrix}-1\\4\\-2\end{bmatrix},b_2=\begin{bmatrix}2\\-3\end{bmatrix}$$

假设输入为 $\boldsymbol{x}=(1,-1)$，并且对应的真实输出为 $\hat{\boldsymbol{y}}=(0,1)$，采用平方损失 $L=\frac{1}{2}\|\boldsymbol{y}-\hat{\boldsymbol{y}}\|_2^2$。试计算函数 $L$ 关于 $\boldsymbol{b}_1$ 的梯度。

**解** （1）因为 $\frac{\partial f}{\partial \boldsymbol{x}}=\boldsymbol{a}$，因此 $H_f=\boldsymbol{O}$。又

$$\begin{aligned}d(g)&=Tr(d\boldsymbol{x}^{\mathrm{T}}\cdot\boldsymbol{A}\boldsymbol{x}+\boldsymbol{A}\boldsymbol{x}d\boldsymbol{x})\\&=Tr(\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A})d\boldsymbol{x})\end{aligned}$$

所以 Hessian 矩阵 $H_g=\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}$。 **（5 分）**

（2）

先计算前项过程：

$$\boldsymbol{A}_1\boldsymbol{x}+\boldsymbol{b}_1=\begin{bmatrix}-2\\1\\2\end{bmatrix},\boldsymbol{A}_2(\mathrm{ReLU}(\boldsymbol{A}_1\boldsymbol{x}+\boldsymbol{b}_1))+\boldsymbol{b}_2=\begin{bmatrix}1\\0\end{bmatrix}$$

故 $\boldsymbol{y}=(1,0)$，从而 $L=1$。记

$$\boldsymbol{k}=\mathrm{ReLU}(\boldsymbol{A}_1\boldsymbol{x}+\boldsymbol{b}_1)$$

然后分别计算

$$\frac{\partial L}{\partial \boldsymbol{y}}=\begin{bmatrix}1\\-1\end{bmatrix},\frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial \boldsymbol{k}}=\begin{bmatrix}3&0\\-1&0\\0&0\end{bmatrix},\frac{\partial \boldsymbol{k}^{\mathrm{T}}}{\partial \boldsymbol{b}_1}=\begin{bmatrix}0&0&0\\0&1&0\\0&0&1\end{bmatrix}$$

所以有

$$\frac{\partial L}{\partial \boldsymbol{b}_1}=\frac{\partial \boldsymbol{k}^{\mathrm{T}}}{\partial \boldsymbol{b}_1}\frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial \boldsymbol{k}}\frac{\partial L}{\partial \boldsymbol{y}}=\begin{bmatrix}0&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}3&0\\-1&0\\0&0\end{bmatrix}\begin{bmatrix}1\\-1\end{bmatrix}=\begin{bmatrix}0\\-1\\0\end{bmatrix}$$

**（8 分）**

**题 4** (13') 完成下列问题：

(1) 设已知矩阵 $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ 的 SVD 分解为 $\boldsymbol{U\Sigma V}^{\mathrm{T}}$，请写出拼接矩阵 $\boldsymbol{A} = [\boldsymbol{M} \quad \boldsymbol{M}]$ 的奇异值分解；

(2) 令 $\boldsymbol{A}$ 为 $m \times n$ 矩阵，且 $\boldsymbol{P}$ 为 $m \times m$ 正交矩阵。证明：$\boldsymbol{PA}$ 与 $\boldsymbol{A}$ 的奇异值相同。并说明矩阵 $\boldsymbol{PA}$ 与 $\boldsymbol{A}$ 的左、右奇异向量有何关系？

**解**（1）因为 $\boldsymbol{M} = \boldsymbol{U\Sigma V}^{\mathrm{T}}$，因此

$$\boldsymbol{AA}^{\mathrm{T}} = 2\boldsymbol{MM}^{\mathrm{T}} = \boldsymbol{U} \cdot 2\boldsymbol{\Sigma}_{m\times m}^2 \cdot \boldsymbol{U}^{\mathrm{T}}.$$

故 $\boldsymbol{A}$ 的奇异值为 $\sqrt{2}\boldsymbol{\Sigma}_{m\times 2n}$，左奇异向量构成的正交矩阵就是 $\boldsymbol{U}$。 **（3 分）**

由 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{A}$ 与右奇异向量（设为 $\boldsymbol{V_A}$）的关系可直接计算出

$$[\boldsymbol{\Sigma}_{m\times n}\boldsymbol{V}^{\mathrm{T}} \quad \boldsymbol{\Sigma}_{m\times n}\boldsymbol{V}^{\mathrm{T}}] = \sqrt{2}\boldsymbol{\Sigma}_{m\times 2n}\boldsymbol{V_A}^{\mathrm{T}}$$

对比等式两边可得 $\boldsymbol{V_A}^{\mathrm{T}}$ 的前 $n$ 行即为 $[\frac{1}{\sqrt{2}}\boldsymbol{V}^{\mathrm{T}} \quad \frac{1}{\sqrt{2}}\boldsymbol{V}^{\mathrm{T}}]$。对此进行正交扩充，可得

$$\boldsymbol{V_A}^{\mathrm{T}} = \begin{bmatrix} \frac{1}{\sqrt{2}}\boldsymbol{V}^{\mathrm{T}} & \frac{1}{\sqrt{2}}\boldsymbol{V}^{\mathrm{T}} \\ \frac{1}{\sqrt{2}}\boldsymbol{V}^{\mathrm{T}} & -\frac{1}{\sqrt{2}}\boldsymbol{V}^{\mathrm{T}} \end{bmatrix}$$

**（4 分）**

（2）因为 $(\boldsymbol{PA})^{\mathrm{T}}(\boldsymbol{PA}) = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{PA} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$，故它们的奇异值相同。

**（2 分）**

假设 $\boldsymbol{A}$ 的 SVD 分解为

$$\boldsymbol{A} = U\Sigma V^{\mathrm{T}}$$

其中 $U, V \in \mathbb{R}^{n\times n}$ 正交。故

$$\boldsymbol{PA} = \boldsymbol{P}U\Sigma V^{\mathrm{T}}$$

因为 $\boldsymbol{P}$ 为正交矩阵，因此 $\boldsymbol{P}U$ 为其左奇异向量，$V$ 就是 $\boldsymbol{PA}$ 的右奇异向量。

**（4 分）**

题 5 (12’) 已知矩阵

$$\boldsymbol{M}=\begin{pmatrix}5&1&2\\-2&-3&4\\1&3&-1\end{pmatrix}$$

(1) 计算矩阵 $\boldsymbol{M}$ 的 1 范数和 $\infty$ 范数；

(2) 证明：如果矩阵 $\boldsymbol{A}=[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n]$ 是按列分块的，那么 $\|\boldsymbol{A}\|_F^2=\|\boldsymbol{a}_1\|_2^2+\|\boldsymbol{a}_2\|_2^2+\cdots+\|\boldsymbol{a}_n\|_2^2$。

(3) 设 $a_1,a_2,\cdots,a_n$ 是 $n$ 个正数，证明：由

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

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

解（1）

$$\|\boldsymbol{M}\|_1=\max\{5+|-2|+1,1+|-3|+3,2+4+|-1|\}=8$$

$$\|\boldsymbol{M}\|_\infty=\max\{5+1+2,|-2|+|-3|+4,1+3+|-1|\}=9$$

**（2 分）**

（2）因为矩阵 $\boldsymbol{A}$ 的 $F$ 范数的平方是矩阵各元素的平方和，而 $\boldsymbol{a}_1$ 的 $\ell_2$ 范数的平方是矩阵第一列元素的平方和，以此类推，$\boldsymbol{a}_n$ 的 $\ell_2$ 范数的平方是矩阵第 $n$ 列元素的平方和。因此，的证 **（5 分）**

（3）可以通过定义证明该函数具有范数的非负性、齐次性和三角不等式，只要证明过程合理即可。 **（5 分）**

题 6 (12’) 求解以下问题：

(1) 假设 $X_1 \to X_2 \to X_3 \to \cdots \to X_n$ 是一个马尔科夫链，即

$$p\left(x_1, x_2, \ldots, x_n\right) = p\left(x_1\right) p\left(x_2 \mid x_1\right) \cdots p\left(x_n \mid x_{n-1}\right)$$

试化简 $I\left(X_1; X_2, \ldots, X_n\right)$。

(2) $X$ 和 $Y$ 是 $\{0,1,2,3\}$ 上的独立、等概率分布的随机变量，求 $H(X+Y)$。

**解**（1）

$$\begin{aligned}
I\left(X_1; X_2, \ldots, X_n\right) &= H\left(X_1\right) - H\left(X_1 \mid X_2, \ldots, X_n\right) \\
&= H\left(X_1\right) - \left[H\left(X_1, X_2, \ldots, X_n\right) - H\left(X_2, \ldots, X_n\right)\right] \\
&= H\left(X_1\right) - \left[\sum_{i=1}^{n} H\left(X_i \mid X_{i-1}, \ldots, X_1\right) - \sum_{i=2}^{n} H\left(X_i \mid X_{i-1}, \ldots, X_2\right)\right] \\
&= H\left(X_1\right) - \Bigg[\Bigg(H\left(X_1\right) + \sum_{i=2}^{n} H\left(X_i \mid X_{i-1}\right)\Bigg) - \Bigg(H\left(X_2\right) + \sum_{i=3}^{n} H\left(X_i \mid X \right. \\
&= H\left(X_2\right) - H\left(X_2 \mid X_1\right) \\
&= I\left(X_2; X_1\right) \\
&= I\left(X_1; X_2\right)
\end{aligned}$$

**（7 分）**

（2）。解：$X+Y$ 的分布律

| $X+Y$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 概率 | 1/16 | 2/16 | 3/16 | 4/16 | 3/16 | 2/16 | 1/16 |

因此

$$H(X+Y) = -(2\frac{1}{16}\log_2\frac{1}{16} + 2\frac{2}{16}\log_2\frac{2}{16} + 2\frac{3}{16}\log_2\frac{3}{16} + \frac{4}{16}\log_2\frac{4}{16})$$

也可以用其他对数，这里可化简也可以不化简。

**（5 分）**

**题 7** (13') 求解下述问题:

(1) 请计算负熵函数 $f(\boldsymbol{x}) = \sum_{i=1}^{n} x_i \log x_i$ 的共轭函数。

(2) 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 $\boldsymbol{A} \in \mathbb{R}^{m\times n}, m \leq n, \text{rank}(\boldsymbol{A}) = m$。

**解** (1)设 $g(x) = \langle x, y\rangle - f(x) = \langle x, y\rangle - \sum_{i=1}^{n} x_i \log x_i$,对 $x$ 求梯度可得:

$$\nabla g = [y_1 - \log x_1 - 1 \ y_2 - \log x_2 - 1 \ \cdots \ y_n - \log x_n - 1]^{\mathrm{T}}$$

因此,可得最优解为 $x_i = e^{y_i - 1}$,代入即得:

$$f^*(\boldsymbol{y}) = \sum_{i=1}^{n} e^{y_i - 1}.$$

**(6 分)**

(2)优化问题为

$$\begin{aligned} \texttt{min} \quad & f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x}\|_2^2 \\ \texttt{suject to} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

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 $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \text{rank}(\boldsymbol{A}) = m$ 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 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最小二范数解：

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

**（7 分）**

题 8 (14’) 求解如下优化问题

(1) 使用梯度下降法和固定步长 $\lambda=0.1$ 计算 $\min f(x)=(x_1-1)^2+16(x_2-2)^2$ ，初始点 $x^{(0)}=(3,2)^T$，迭代至第三步后终止，请计算迭代过程中的解、梯度和函数值（第三步不用计算梯度和函数值，计算结果保留小数点后两位）。

(2) 请谈谈牛顿法与 BFGS 方法之间的区别与联系。

解 （1）易求得梯度为：

$$\nabla f=[2(x_1-1)\quad 32(x_2-2)]^{\mathrm{T}}$$

具体迭代结果：

| $k$ | $x^{(k)^{\mathrm{T}}}$ | $g_k^{\mathrm{T}}$ | $f_k$ |
|---|---|---|---|
| 0 | $(3,2)$ | $(4,0)$ | 4 |
| 1 | $(2.6,2)$ | $(3.2,0)$ | 2.56 |
| 2 | $(2.28,2)$ | $(2.56,0)$ | 1.6384 |
| 3 | $(2.024,2)$ | | |

**（8 分）**

（2）牛顿法（Newton’s method）和 BFGS 算法（Broyden-Fletcher-Goldfarb-Shanno algorithm）都是优化算法，用于求解无约束非线性优化问题。它们都属于迭代优化算法，但有一些关键的区别。

**1. 原理和基本思想：**

- 牛顿法：基于二阶导数（Hessian 矩阵）的信息，通过迭代更新当前点的位置，以找到目标函数的极小值点。牛顿法通常收敛速度很快，但需要计算和存储 Hessian 矩阵，这可能在高维问题中成为一个挑战。

- BFGS 算法：也是基于二阶导数的信息，但它通过估计 Hessian 矩阵的逆矩阵来进行迭代。相较于牛顿法，BFGS 不需要直接计算和存储 Hessian 矩阵，这降低了存储和计算的复杂性。

**2. 更新规则：**

- 牛顿法：通过解线性方程系统来计算下降方向。牛顿法的更新规则是直接基于 Hessian 矩阵和梯度的。

- BFGS 算法：通过迭代更新逆 Hessian 矩阵的估计值，以获得下降方向。BFGS 的更新规则更为复杂，但避免了直接计算 Hessian 矩阵的需求。

**3. 收敛性：**

- 牛顿法：在适当的条件下，牛顿法通常具有二次收敛性，收敛速度较快，但可能受到 Hessian 矩阵的病态性（ill-conditioning）的影响。

- BFGS 算法：通常也具有二次收敛性，但相对于牛顿法更具有稳定性，特别是在高维问题中，因为它不需要直接处理 Hessian 矩阵。

总体上，只要言之有理，且答到关键点即可给分。

**（6 分）**