2021 级华东师范大学本科生期末考试

《数据科学与工程数学基础》笔试试卷 2023.2

学院：数据科学与工程学院　　考试形式：闭卷　　所需时间：120 分钟

考生姓名：______　　学号：______　　专业：______　　任课教师：黄定江

题目	一	二	三	四	五	六	七	八	总分
满分	13	12	12	13	13	12	13	12	100
得分									
阅卷人									

题 1 (13') 假设 $Q \in \mathbb{R}^{n\times n}\backslash\{0\}$ 是一个投影矩阵.

(1) 证明：$Qz = z,\ \forall z \in \mathcal{R}(Q)$ 以及 $Qy - y \in \mathcal{N}(Q),\ \forall y \in \mathbb{R}^n$.

(2) 证明：Q 的特征值 $\lambda \in \Lambda(Q) \subseteq \{0,1\}$。

(3) 假设 $\mathcal{R}(Q) = \operatorname{span}(u_1,\ldots,u_r)$，$\mathcal{N}(Q) = \operatorname{span}(v_{r+1},\ldots,v_n)$ 这里 r 表示投影矩阵的秩，请据此写出 Q 的特征分解 $Q = XDX^{-1}$ 的具体形式。

(4) 证明：当 $Q \neq I_n$，$\det(Q) = 0$。

解 (1) $\forall\, y \in \mathcal{R}(Q)$ 必存在 $x \in \mathbb{R}^n$，使得 $y = Qx$ 成立。因此，$Qy = Q^2x = Qx = y$。另外，$\forall x \in \mathbb{R}^n$，$Q(Qx - x) = Q^2x - Qx = Qx - Qx = 0$。　**（3 分）**

(2) 对 $\lambda \in \Lambda(Q)$，$x \in \mathbb{R}^n \backslash \{0\}$，有 $Qx = \lambda x$. 由于 $Q = Q^2$, $\lambda x = Qx = Q(Qx) = Q(\lambda x) = \lambda Qx = \lambda^2 x$。因为 $x \neq 0 \in \mathbb{R}^n$，故 $\lambda = \lambda^2$，$\lambda \in \{0,1\}$。因此 $\Lambda(Q) \subseteq \{0,1\}$. **（3 分）**

(3) 由 (1) 可知，$\forall i = 1, ..., r, u_i \in \mathcal{R}(Q), Qu_i = u_i$，以及 $\forall j = r+1, ..., n, v_j \in \mathcal{N}(Q), Qv_j = 0$。故令

$$X := (u_1|\cdots|u_r\,|v_{r+1}|\cdots|\ v_n) \in \mathbb{R}^{n\times n},$$

$$D := \mathrm{diag}_{n\times n}(\underbrace{1,\ldots,1}_{r \text{ times}},0\ldots,0) \in \mathbb{R}^{n\times n},$$

此时 $Q = XDX^{-1}$。 **（4 分）**

(3) 反证 $\det(Q) \neq 0 \Longrightarrow Q = I_n$。由于 $\det(Q) \neq 0$，故 Q 可逆。故由 $Q^2 = Q$，得 $Q^{-1}Q^2 = Q^{-1}Q$，即 $Q = I_n$。 **（3 分）**

题 2　(12’) 已知矩阵

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

(1) 先求矩阵 $\boldsymbol{M}$ 的 QR 分解，然后根据 QR 分解求解方程组 $\boldsymbol{Mx} = \boldsymbol{b}$，其中 $\boldsymbol{b} = [6,4,1]^\top$；

(2) 请根据你对 LU 分解、Cholesky 分解和 QR 分解的理解，谈谈它们之间的联系与差异。

解（1）因为 $\boldsymbol{\alpha}_1 = (0,0,2)^T$, 记 $a_1 = \|\alpha_1\|_2 = 2$, 令　$\boldsymbol{w}_1 = \frac{\alpha_1 - a_1 e_1}{\|\alpha_1 - a_1 e_1\|_2} = \frac{1}{\sqrt{2}}(-1,0,1)^T$, 则

$$\boldsymbol{H}_1 = I - 2\boldsymbol{w}_1\boldsymbol{w}_1^H = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

从而

$$\boldsymbol{H}_1\boldsymbol{M} = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 0 & 4 \\ 0 & -3 & 3 \end{pmatrix}$$

记 $\boldsymbol{\beta} = (0, -3)^T$, 则 $b_2 = \|\boldsymbol{\beta}_2\|_2 = 3$, 令 $\boldsymbol{w}_2 = \frac{\boldsymbol{\beta}_2 - b_2\boldsymbol{e}_1}{\|\boldsymbol{\beta}_2 - b_2\boldsymbol{e}_1\|_2} = \frac{1}{\sqrt{2}}(-1, -1)^T$

$$\widetilde{\boldsymbol{H}}_2 = \boldsymbol{I} - 2\boldsymbol{w}_2\boldsymbol{w}_2^{\boldsymbol{H}} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$$

记

$$\boldsymbol{H}_2 = \begin{pmatrix} 1 & 0^T \\ 0 & \widetilde{\boldsymbol{H}}_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$

则

$$\boldsymbol{H}_2\left(\boldsymbol{H}_1\boldsymbol{M}\right) = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 3 & -3 \\ 0 & 0 & -4 \end{pmatrix} = \boldsymbol{R}$$

易知 Rx $= H_2H_1b = (1, -6, -4)^T$, 故可求得 x $= (1/2, \quad -1, \quad 1)^T$ **（7 分）**

（2）（总体上，只要理解正确即可）这三种分解方式的联系在于将一个非三角矩阵转化为三角矩阵，这样便可以有效提高实际应用中一系列方程组求解的效率。它们区别在于 LU 分解和 Cholesky 分解是基于高斯消去法，分别适用于一般地可逆矩阵和对称矩阵，QR 分解则是利用正交投影的方法，将向量分别投影到左边平面上的方式，从而达到对角化。 **（5 分）**

题 3 (12’) 完成下列矩阵函数求梯度：

(1) 求行列式函数 $f(\boldsymbol{X}) = |\boldsymbol{X}^3|$ 的梯度矩阵，其中变量 $\boldsymbol{X} \in \mathbb{R}^{n\times n}$ 非奇异；

(2) 求迹函数 $f(\boldsymbol{X}) = Tr(\boldsymbol{X}^\top\boldsymbol{X}^{-1}\boldsymbol{A})$ 的梯度矩阵，其中变量 $\boldsymbol{X} \in \mathbb{R}^{n\times n}$ 非奇异。

解 （1）因为 $d|\boldsymbol{X}|=Tr(|\boldsymbol{X}|\boldsymbol{X}^{-1}d\boldsymbol{X})$，因此

$$
\begin{aligned}
d(|\boldsymbol{X}^3|) &= Tr(\boldsymbol{X}\boldsymbol{X}^{-3}d\boldsymbol{X}^3)\\
&= Tr(|\boldsymbol{X}^3|X^{-3}(d\boldsymbol{X}\cdot\boldsymbol{X}^2+\boldsymbol{X}d\boldsymbol{X}\cdot\boldsymbol{X}+\boldsymbol{X}^2d\boldsymbol{X}))\\
&= Tr(3|\boldsymbol{X}|^3\boldsymbol{X}^{-1}d\boldsymbol{X})
\end{aligned}
$$

所以梯度矩阵 $3|\boldsymbol{X}|^3(\boldsymbol{X}^{-1})^\top$。 **（6 分）**

（2）

$$
\begin{aligned}
dTr(\boldsymbol{X}^\top\boldsymbol{X}^{-1}\boldsymbol{A}) &= Tr(d(\boldsymbol{X}^\top\boldsymbol{X}^{-1}\boldsymbol{A}))\\
&= Tr(d\boldsymbol{X}^\top\boldsymbol{X}^{-1}\boldsymbol{A}-\boldsymbol{X}^\top\boldsymbol{X}^{-1}d\boldsymbol{X}\boldsymbol{X}^{-1}\boldsymbol{A})\\
&= Tr((\boldsymbol{A}^\top(\boldsymbol{X}^{-1})^\top-\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}^\top\boldsymbol{X}^{-1})d\boldsymbol{X})
\end{aligned}
$$

即

$$
\frac{\partial Tr(\boldsymbol{X}^\top\boldsymbol{X}^{-1}\boldsymbol{A})}{\partial\boldsymbol{X}}=(\boldsymbol{A}^\top(\boldsymbol{X}^{-1})^\top-\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}^\top\boldsymbol{X}^{-1})^\top
$$

（6 分）

题 4 (13’) 矩阵

$$
\boldsymbol{A}=\begin{pmatrix}7/5 & -1/5\\ 1/5 & 7/5\\ 0 & 0\end{pmatrix}
$$

(1) 计算矩阵 $\boldsymbol{A}$ 的 SVD 分解；

(2) 假设 $\boldsymbol{M}$ 是任意一个非奇异 $n\times d$ 的矩阵，已知其奇异值分解为 $\boldsymbol{M}=\boldsymbol{U}\Sigma\boldsymbol{V}^\top$，其中 $\boldsymbol{U}=[u_1,u_2,\cdots,u_n]$，$\Sigma=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$，$\boldsymbol{V}=[v_1,v_2,\cdots,v_n]$。请分别写出矩阵 $\boldsymbol{M}$ 的逆矩阵的 SVD 分解；

(3) 请写出 (2) 中 $\boldsymbol{M}$ 的转置矩阵的 SVD 分解。

解（1）易计算

$$A^{\top}A=\begin{pmatrix}2&0\\0&2\end{pmatrix}$$

求得其特征值为 $2,9$。对应的特征向量分别为 $(1,0)^{\top},(0,1)^{\top}$，因此，

$$V=\begin{pmatrix}1&0\\0&1\end{pmatrix}$$

（3 分）

由 $\boldsymbol{AV}$ 与 $\boldsymbol{U}$ 的关系可直接计算出

$$U=\begin{pmatrix}7/5\sqrt{2}&-1/5\sqrt{2}&0\\1/5\sqrt{2}&7/5\sqrt{2}&0\\0&0&1\end{pmatrix}$$

（3 分）

（2）因为 $\boldsymbol{M}$ 的 SVD 分解为

$$\boldsymbol{M}=U\Sigma V^{\mathrm{T}}=(u_1|\cdots|u_n)\left[\operatorname{diag}_{n\times n}(\lambda_1,\ldots,\lambda_n)\right](v_1|\cdots|v_n)^{\mathrm{T}}$$

其中 $U,V\in\mathbb{R}^{n\times n}$ 正交, $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_n\geq 0$。因为 $\boldsymbol{M}$ 可逆，有 $\operatorname{rank}(\boldsymbol{M})=n$，故 $\lambda_i>0$, $\forall i\in\{1,\ldots,n\}$. 又由 $\boldsymbol{M}=U\Sigma V^{\mathrm{T}}$ 有 $\boldsymbol{M}v_i=\lambda_i u_i\ \forall i\in\{1,\ldots,n\}$ ，因此

$$\boldsymbol{M}^{-1}u_i=\boldsymbol{M}^{-1}\left(\frac{1}{\lambda_i}\boldsymbol{M}v_i\right)=\frac{1}{\lambda_i}v_i$$

其中 $\frac{1}{\lambda_n}\geq\cdots\geq\frac{1}{\lambda_2}\geq\frac{1}{\lambda_1}>0$。综上

$$\boldsymbol{M}^{-1}=(v_n|\cdots|v_2\mid v_1)\left[\operatorname{diag}_{n\times n}\left(\frac{1}{\lambda_n},\ldots,\frac{1}{\lambda_2},\frac{1}{\lambda_1}\right)\right](u_n|\cdots|u_2\mid u_1)^{\mathrm{T}},$$

整理可得

$$\boldsymbol{M}^{-1}=(VP)\left(P\Sigma^{-1}P\right)(UP)^{\mathrm{T}}.$$

记 $P:=(e_n|\cdots|e_2\mid e_1)\in\mathbb{R}^{n\times n}$. (由于 $PP^{\mathrm{T}}=P^2=I_n$，P 是正交阵，且 VP, UP 是正交阵)。 **（5 分）**

（3）因为 $\boldsymbol{M}$ 的 SVD 分解为

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

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

$$\boldsymbol{M}^{\top} = V\Sigma U^{T}.$$

显然 V, U 是正交阵。 **（2 分）**

题 5 (13’) 已知矩阵

$$\boldsymbol{M} = \begin{pmatrix} 2 & 0 \\ -1 & 2 \\ 0 & 3 \end{pmatrix}$$

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

(2) 计算矩阵 $\boldsymbol{M}$ 的 F 范数和 2 范数。

(3) 试比较任意矩阵 $\boldsymbol{A}$ 的 F 范数和 2 范数的大小，并给出理由。

解 （1）

$$\|\boldsymbol{M}\|_1 = \max\{2 + \|-1\|, 0 + 2 + 3\} = 5$$

$$\|\boldsymbol{M}\|_{\infty} = \max\{2 + \|0\|, \|-1\| + 2, 0 + 3\} = 3$$

（4 分）

（2）$\boldsymbol{M}^{\top}\boldsymbol{M} = \begin{pmatrix} 5 & -2 \\ -2 & 13 \end{pmatrix}$

$\lambda_{max}(\boldsymbol{M}^{\top}\boldsymbol{M}) = 9 + 2\sqrt{2}$

$\|\boldsymbol{M}\|_2 = \sqrt{9 + 2\sqrt{2}}$

易求得 $\|\boldsymbol{M}\|_F = \sqrt{18} = 3\sqrt{2}$ **（5 分）**

（3）显然 F 范数大于 2 范数，因为 F 范数的平方是所有奇异值的平方之和，而 2 范数的平方是最大奇异值的平方，因此可知这一点成立。 **（4 分）**

题 6 (12') 求解以下问题:

(1) 证明: Gauss 概率密度函数的累积分布函数

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-u^2/2}du$$

是对数-凹函数. 即 $\log(\Phi(x))$ 是凹函数。

(2) 设总体 $X\sim E(\theta), X_1, X_2, \ldots, X_n$ 为来自总体 X 的样本, θ 的先验分布为指数分布 $E(\lambda)$ (λ 已知), 求 θ 的最大后验估计。

解 (1) 由题意得,

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-u^2/2}du$$

$$\Phi^{'}(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

$$\Phi^{''}(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}(-x)$$

$$(\Phi^{'}(x))^2=\frac{1}{2\pi}e^{-x^2}$$

$$\Phi(x)\log\Phi^{''}(x)=\frac{1}{2\pi}\int_{-\infty}^{x}e^{-u^2/2}du\cdot e^{-x^2/2}(-x)$$

(3 分)

当 $x\geq 0$ 时, $(\Phi^{'}(x))^2\geq 0\geq \Phi(x)\Phi^{''}(x)$.
当 $x<0$ 时, 由于 $\frac{u^2}{2}$ 是凸函数, 则

$$\frac{u^2}{2}\geq\frac{x^2}{2}+(u-x)x\geq xu-\frac{x^2}{2}$$

所以,

$$\int_{-\infty}^{x}e^{-u^2/2}du\leq\int_{-\infty}^{x}e^{\frac{x^2}{2}-xu}du$$

$$=e^{\frac{x^2}{2}}\cdot\frac{e^{-xu}}{-x}\Big|_{u=-\infty}^{x}$$

$$= e^{\frac{x^2}{2}} \cdot \frac{e^{-x^2}}{-x}$$

因此 $\Phi(x)\Phi''(x) \leq \frac{1}{2\pi}e^{-x^2} = (\Phi'(x))^2$, $\Phi(x)$ 是对数凹函数. **（3 分）**

（2）因为先验概率密度函数为:

$$\pi(\theta) = \begin{cases} \lambda e^{-\lambda\theta} & \theta > 0 \\ 0 & \theta \leq 0 \end{cases}$$

样本 $(X_1, X_2, \ldots, X_n)$ 的联合概率密度为:

$$q(\mathbf{x} \mid \theta) = \prod_{n=2}^{n} f(x_i \mid \theta) = \begin{cases} \theta^n e^{-\theta \sum_{i=1}^n x_i} & x_1, x_2, \ldots, x_n > 0 \\ 0 & \text{其它} \end{cases}$$

所以 θ 的后验分布密度

$$h(\theta \mid \mathbf{x}) \propto \theta^n e^{-\left(\lambda + \sum_{i=1}^n x_i\right)\theta}$$

$$\ln h(\theta \mid \mathbf{x}) = n \ln \theta - \left(\lambda + \sum_{i=1}^n x_i\right)\theta + \ln c(\mathbf{x})$$

$$\frac{\partial \ln h(\theta \mid \mathbf{x})}{\partial \theta} = \frac{n}{\theta} - \left(\lambda + \sum_{i=1}^n x_i\right) = 0$$

求得 θ 的最大后验估计为 $\hat{\theta} = \frac{1}{\bar{x} + \lambda/n}$ 。当 $n \to \infty$ 时, $\hat{\theta} \to \frac{1}{\bar{x}}$, 与传统意义下的极大似然估计是一致的。 **（6 分）**

题 7 (13’) 求解下述问题:

(1) 写出下述非线性规划的 KKT 条件，并求解

$$\begin{aligned} &\texttt{maximize} \quad f(x) = (x-4)^2 \\ &\texttt{suject to} \quad 1 \leq x \leq 5 \end{aligned}$$

(2) 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 （1）原问题等价于

$$\begin{cases}\texttt{minimize} \quad -f(x)=(x-4)^2\\ g_1(x)=-x+1\le 5\\ g_2(x)=x-5\le 0\end{cases}$$

求目标函数和约束函数的梯度得,

$$\nabla_x f(x)=-2(x-4),\nabla_x g_1(x)=-1,\nabla_x g_2(x)=1$$

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

$$\begin{cases}-2(x^*-4)-v_1^*+v_2^*=0\\ v_1^*(-x^*+1)=0\\ v_2^*(x^*-5)=0\\ v_1^*\ge 0,v_2^*\ge 0\end{cases}$$

（4 分）

若 $v_1^*\ne 0,v_2^*\ne 0$, 无解.
若 $v_1^*=0,v_2^*\ne 0$, 得 $x^*=5,v_2^*=2,-f(x^*)=-1$.
若 $v_1^*\ne 0,v_2^*=0$, 得 $x^*=1,v_1^*=6,-f(x^*)=-9$.
若 $v_1^*=0,v_2^*=0$, 得 $x^*=4,f(x^*)=0$.
因此最优点 $x^*=1,\texttt{maximize} f(x)=9$。 **（3 分）**

（2）优化问题为

$$\texttt{maximize} \quad f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$$

$$\texttt{suject to} \quad \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=1$$

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:

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

（3 分）

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

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

（3 分）

题 8　(12') 求解如下优化问题

(1) 考虑问题

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

从初始点 $\boldsymbol{x}^{(0)} = (0,1)^{\mathrm{T}}$ 出发, 写出用最速下降法迭代两步的求解过程，并说明迭代是否可以终止。

(2) 尽管牛顿法在求解无约束优化问题时，具有更高阶的收敛速度，但在处理大规模问题时，该方法涉及的 Hessian 矩阵及其逆矩阵的计算是非常耗时的。因此，研究者们探究了对 Hessian 矩阵逆近似的迭代方法，也就是拟牛顿法或变尺度法。通常这样近似 Hessian 矩阵或逆矩阵需要保持原有的性质，即满足割线方程:

$$\Delta x^{(k)} = \bar{H}^{(k+1)}\Delta g^{(k)},$$

其中 $\bar{H}^{(k)}$ 表示矩阵的 Hessian 矩阵逆的近似。如果已经得到上一轮迭代时的近似 $\bar{H}^{(k)}$，则一般地做法是通过秩一修正或秩二修正得到下一轮的近似矩阵 $\bar{H}^{(k+1)}$。当采用秩二修正时，便得到著名的 DFP 法。现已知第 k 次迭代时的 $\Delta x^{(k)}, \bar{H}^{(k)}, \Delta g^{(k)}$，请给出下一轮迭代的 Hessian 矩阵逆的近似的推导过程。

解 （1）由题意得, $\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = (6x_1 - 4x_1x_2, 6x_2 - 2x_1^2)^T$，故 $g^{(0)} = \nabla_{\boldsymbol{x}} f(\boldsymbol{x}^{(0)}) = (0, 6)$。现设最速下降法的步长为 λ, 那么

$$f(\boldsymbol{x}^{(0)} - \lambda g^{(0)}) = 3(1 - 6\lambda^2)$$

在 $\nabla_{\boldsymbol{x}} f(\boldsymbol{x}^{(0)})$ 方向上, 使 $f(\boldsymbol{x})$ 最小的 λ 为

$$\lambda = \frac{1}{6}$$

所以,

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

易知此时 $\boldsymbol{x}^{(1)}$ 处的梯度已经为 0，可见已经收敛。显然

$$\boldsymbol{x}^{(2)} = (0, 0)^T$$

（7 分）

（2）可设

$$\overline{\boldsymbol{H}}^{(k+1)} = \overline{\boldsymbol{H}}^{(k)} + a\boldsymbol{u}\boldsymbol{u}^T + b\boldsymbol{v}\boldsymbol{v}^T$$

其中 $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$，$a, b \in \mathbb{R}$ 待定。

同上所述，仍然利用割线方程，有

$$\begin{aligned}\Delta\boldsymbol{x}^{(k)} &= \overline{\boldsymbol{H}}^{(k+1)}\Delta\boldsymbol{g}^{(k)} \\ &= (\overline{\boldsymbol{H}}^{(k)} + a\boldsymbol{u}\boldsymbol{u}^T + b\boldsymbol{v}\boldsymbol{v}^T)\Delta\boldsymbol{g}^{(k)}\end{aligned}$$

整理得

$$a(\boldsymbol{u}^T\Delta\boldsymbol{g}^{(k)})\boldsymbol{u} + b(\boldsymbol{v}^T\Delta\boldsymbol{g}^{(k)})\boldsymbol{v} = \Delta\boldsymbol{x}^{(k)} - \overline{\boldsymbol{H}}^{(k)}\Delta\boldsymbol{g}^{(k)}.$$

因此，$\boldsymbol{u}, \boldsymbol{v}$ 的线性组合等于 $\Delta\boldsymbol{x}^{(k)} - \overline{\boldsymbol{H}}^{(k)}\Delta\boldsymbol{g}^{(k)}$。同样地，不妨令 $\boldsymbol{u} = \Delta\boldsymbol{x}^{(k)}$，$\boldsymbol{v} = \overline{\boldsymbol{H}}^{(k)}\Delta\boldsymbol{g}^{(k)}$，则代入可得

$$a = \frac{1}{(\Delta\boldsymbol{x}^{(k)})^T\Delta\boldsymbol{g}^{(k)}},$$

$$b = -\frac{1}{(\Delta\boldsymbol{g}^{(k)})^T\overline{\boldsymbol{H}}^{(k)}\Delta\boldsymbol{g}^{(k)}}.$$

从而，得到更新公式：

$$\overline{\boldsymbol{H}}^{(k+1)} = \overline{\boldsymbol{H}}^{(k)} + \frac{\Delta \boldsymbol{x}^{(k)})\Delta \boldsymbol{x}^{(k)})^T}{(\Delta \boldsymbol{x}^{(k)})^T \Delta \boldsymbol{g}^{(k)}} - \frac{\overline{\boldsymbol{H}}^{(k)} \Delta \boldsymbol{g}^{(k)} (\overline{\boldsymbol{H}}^{(k)} \Delta \boldsymbol{g}^{(k)})^T}{(\Delta \boldsymbol{g}^{(k)})^T \overline{\boldsymbol{H}}^{(k)} \Delta \boldsymbol{g}^{(k)}}.$$

（5 分）