最优化方法：基础知识习题解答

162350107 冉茂印

March 14, 2025

Code Repository: `https://github.com/myRan-cyber/OMW/tree/main/first_hw`
More Detals `https://zhuanlan.zhihu.com/p/30307011306`

Problem 1. 设 $f(x) = \|A\boldsymbol{x} - \boldsymbol{y}\|^2$，其中 A 为列满秩矩阵。

a. 使用逐分量计算梯度方式计算 $\frac{\partial f}{\partial x_k}$；

b. 使用整体的方法计算梯度 $\frac{\partial f}{\partial \boldsymbol{x}}$。

Ans.

Part (a)：逐分量梯度计算 展开平方范数：

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij} x_j - y_i \right)^2 .$$

对 x_k 求偏导：

$$\frac{\partial f}{\partial x_k} = 2 \sum_{i=1}^{m} A_{ik} \left(\sum_{j=1}^{n} A_{ij} x_j - y_i \right).$$

$$\frac{\partial f}{\partial x_k} = 2 \left[A^\top (A\boldsymbol{x} - \boldsymbol{y}) \right]_k .$$

$$\boxed{\nabla f(\boldsymbol{x}) = 2A^\top (A\boldsymbol{x} - \boldsymbol{y})}.$$

Part (b)：整体梯度计算 将 $f(\boldsymbol{x})$ 展开平方范数：

$$f(\boldsymbol{x}) = (A\boldsymbol{x} - \boldsymbol{y})^\top (A\boldsymbol{x} - \boldsymbol{y}) = \boldsymbol{x}^\top A^\top A\boldsymbol{x} - 2\boldsymbol{y}^\top A\boldsymbol{x} + \boldsymbol{y}^\top \boldsymbol{y}.$$

对 $\boldsymbol{x}$ 求梯度：

$$\frac{\partial}{\partial \boldsymbol{x}} (\boldsymbol{x}^\top A^\top A\boldsymbol{x}) = 2A^\top A\boldsymbol{x},$$
$$\frac{\partial}{\partial \boldsymbol{x}} (-2\boldsymbol{y}^\top A\boldsymbol{x}) = -2A^\top \boldsymbol{y},$$
$$\frac{\partial}{\partial \boldsymbol{x}} (\boldsymbol{y}^\top \boldsymbol{y}) = \boldsymbol{0}.$$

得：

$$\boxed{\nabla f(\boldsymbol{x}) = 2A^\top (A\boldsymbol{x} - \boldsymbol{y})}.$$

Problem 2. 设 $f(\boldsymbol{x}) = tr(\sigma^2 I + \boldsymbol{x}\boldsymbol{x}^\top)$，其中 $\sigma > 0$，I 为 $n \times n$ 单位矩阵，$\boldsymbol{x} \in \mathbb{R}^n$。

a. 使用逐分量计算梯度方式计算 $\frac{\partial f}{\partial x_k}$；

b. 使用整体的方法计算梯度 $\frac{\partial f}{\partial \boldsymbol{x}}$。

Ans.

Part (a)：逐分量梯度计算

$$f(\boldsymbol{x}) = \operatorname{tr}(\sigma^2 I) + \operatorname{tr}(\boldsymbol{x}\boldsymbol{x}^\top).$$

$$f(\boldsymbol{x}) = n\sigma^2 + \sum_{i=1}^{n} x_i^2.$$

对 x_k 求偏导：

$$\frac{\partial f}{\partial x_k} = \frac{\partial}{\partial x_k}\left(n\sigma^2 + \sum_{i=1}^{n} x_i^2\right) = 2x_k.$$

得：

$$\boxed{\nabla f(\boldsymbol{x}) = (2x_1, 2x_2, \ldots, 2x_n)^\top}.$$

Part (b)：整体梯度计算 对 $f(\boldsymbol{x}) = \operatorname{tr}(\sigma^2 I + \boldsymbol{x}\boldsymbol{x}^\top)$ 求微分：

$$df = \operatorname{tr}(d(\boldsymbol{x}\boldsymbol{x}^\top)) = \operatorname{tr}(d\boldsymbol{x} \cdot \boldsymbol{x}^\top + \boldsymbol{x} \cdot d\boldsymbol{x}^\top).$$

$$df = 2\operatorname{tr}(\boldsymbol{x}^\top d\boldsymbol{x}) = 2\boldsymbol{x}^\top d\boldsymbol{x}.$$

$$\boxed{\nabla f(\boldsymbol{x}) = 2\boldsymbol{x}}.$$

Problem 3. 考虑二次型 $f(x) = x^\top Q x$，其中 $x \in \mathbb{R}^n$，$Q \in \mathbb{R}^{n\times n}$，且 Q 的特征值分解为

$$Q = U\Lambda U^\top,$$

其中 U 为正交矩阵，特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$。

a. 证明：

$$\lambda_1 = \max_{\|x\|=1} x^\top Q x = \max_{x\neq 0} \frac{x^\top Q x}{x^\top x}.$$

b. 证明：对任意 $x \in \mathbb{R}^n$，$x \neq 0$，有

$$\lambda_n \|x\|^2 \leq x^\top Q x \leq \lambda_1 \|x\|^2.$$

Ans.

Part (a) 令 $y = U^\top x$，则 $x = Uy$，

$$x^\top Q x = y^\top U^\top Q U y = y^\top \Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2.$$

当 $\|x\| = 1$ 时，$\|y\| = 1$，即 $\sum_{i=1}^{n} y_i^2 = 1$

$$x^\top Q x = \sum_{i=1}^{n} \lambda_i y_i^2 \leq \lambda_1 \sum_{i=1}^{n} y_i^2 = \lambda_1.$$

当 $y_1 = 1$，其余 $y_i = 0$ 时，等号成立，故：

$$\max_{\|x\|=1} x^\top Q x = \lambda_1.$$

对任意 $x \neq 0$，

$$\frac{x^\top Qx}{x^\top x} = \frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2} \leq \lambda_1.$$

当 $y_1 \neq 0$，其余 $y_i = 0$ 时，等号成立，故：

$$\max_{x\neq 0} \frac{x^\top Qx}{x^\top x} = \lambda_1.$$

Part (b) 由 $\lambda_n \leq \lambda_i \leq \lambda_1$，对任意 $y \in \mathbb{R}^n$，有：

$$\lambda_n \sum_{i=1}^n y_i^2 \leq \sum_{i=1}^n \lambda_i y_i^2 \leq \lambda_1 \sum_{i=1}^n y_i^2.$$

由于 $\sum_{i=1}^n y_i^2 = \|x\|^2$，代入得：

$$\lambda_n \|x\|^2 \leq x^\top Qx \leq \lambda_1 \|x\|^2.$$

当 x 分别为 Q 的最小和最大特征值对应的特征向量时，等号成立。

Problem 4. 给定二次型 $f(x) = 3x_1^2 + 4x_2^2 + 3x_3^2 + 2x_1x_3$，完成以下问题：

a. 将此二次型写成矩阵-向量形式，并求其标准型；

b. 证明：$\min_{x\neq 0} \frac{f(x)}{\|x\|^2} = 2$；

c. 证明集合 $\{x \in \mathbb{R}^3 \mid f(x) = 1\}$ 是一个椭球，并计算其长轴与短轴之比；

d. 计算 $\nabla^2 f(x)$ 的条件数。

Ans.

Part (a): 将二次型写成 $f(x) = x^\top Qx$，其中 Q 为：

$$Q = \begin{pmatrix} 3 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 3 \end{pmatrix}.$$

对 Q 特征分解，$\det(Q - \lambda I) = 0$，得：

$$\lambda_1 = 4 \text{ (二重根)}, \quad \lambda_2 = 2.$$

通过 $x = Uy$，二次型化为标准型：

$$\boxed{f(x) = 4y_1^2 + 4y_2^2 + 2y_3^2}.$$

Part (b): 由瑞利商性质，最小值等于 Q 的最小特征值（上一题已证）：

$$\boxed{\min_{x\neq 0} \frac{x^\top Qx}{x^\top x} = \lambda_{\min}(Q) = 2}.$$

当 x 为 Q 对应特征值 2 的特征向量时，等号成立。

Part (c): $f(x) = 1$ 对应标准型：

$$4y_1^2 + 4y_2^2 + 2y_3^2 = 1 \implies \frac{y_1^2}{(1/2)^2} + \frac{y_2^2}{(1/2)^2} + \frac{y_3^2}{(1/\sqrt{2})^2} = 1.$$

椭球的半轴长度为 $a = 1/2, b = 1/2, c = 1/\sqrt{2}$。长轴为 c，短轴为 a，其比为：

$$\boxed{\frac{c}{a} = \frac{1/\sqrt{2}}{1/2} = \sqrt{2}}.$$

Part (d): Hessian 矩阵为 $\nabla^2 f(x) = 2Q$，其特征值为 $8, 8, 4$。条件数为最大特征值与最小特征值之比：

$$\boxed{\kappa(\nabla^2 f(x)) = \frac{8}{4} = 2}.$$

Problem 5. 计算函数

$$f(x) = \log \sum_{i=1}^{n} \exp x_i$$

的梯度和 *Hessian* 矩阵，并证明 *Hessian* 矩阵正定。

Ans.

对 x_j 求偏导数：

$$\frac{\partial f}{\partial x_j} = \frac{\exp x_j}{\sum_{i=1}^{n} \exp x_i}.$$

令 $S = \sum_{i=1}^{n} \exp x_i$，则梯度向量为：

$$\nabla f(x) = \left(\frac{\exp x_1}{S}, \frac{\exp x_2}{S}, \ldots, \frac{\exp x_n}{S} \right)^T.$$

对角元素（$i = j$）

$$\frac{\partial^2 f}{\partial x_j^2} = \frac{\exp x_j (S - \exp x_j)}{S^2}.$$

非对角元素（$i \neq j$）

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{-\exp x_i \exp x_j}{S^2}.$$

因此，Hessian 矩阵为：

$$H = \begin{pmatrix} \frac{\exp x_1 (S - \exp x_1)}{S^2} & \frac{-\exp x_1 \exp x_2}{S^2} & \cdots & \frac{-\exp x_1 \exp x_n}{S^2} \\ \frac{-\exp x_2 \exp x_1}{S^2} & \frac{\exp x_2 (S - \exp x_2)}{S^2} & \cdots & \frac{-\exp x_2 \exp x_n}{S^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-\exp x_n \exp x_1}{S^2} & \frac{-\exp x_n \exp x_2}{S^2} & \cdots & \frac{\exp x_n (S - \exp x_n)}{S^2} \end{pmatrix}.$$

对任意非零向量 $z = (z_1, z_2, \ldots, z_n)^T$，计算二次型：

$$z^T H z = \sum_{i=1}^{n} \frac{\exp x_i (S - \exp x_i)}{S^2} z_i^2 + \sum_{i \neq j} \frac{-\exp x_i \exp x_j}{S^2} z_i z_j.$$

$$= \frac{1}{S^2} \left(S \sum_{i=1}^{n} \exp x_i z_i^2 - \sum_{i=1}^{n} (\exp x_i z_i)^2 - \sum_{i \neq j} \exp x_i \exp x_j z_i z_j \right).$$

$$= \frac{1}{S^2} \left(S \sum_{i=1}^{n} \exp x_i z_i^2 - \left(\sum_{i=1}^{n} \exp x_i z_i \right)^2 \right).$$

令 $p_i = \frac{\exp x_i}{S}$，则：

$$z^T H z = \frac{1}{S} \left(\sum_{i=1}^{n} p_i z_i^2 - \left(\sum_{i=1}^{n} p_i z_i \right)^2 \right).$$

上式等价于概率分布 p_i 下 z 的方差，故：

$$z^T H z \geq 0.$$

当且仅当 z 为常数向量时，方差为零，因此 $\boxed{H \text{ 是半正定矩阵}}$。

Problem 6. 设 A 为对称矩阵，计算下式的梯度：

$$\ell(\mathbf{x}) = \|A - \mathbf{x}\mathbf{x}^T\|^2$$

Ans.

$$\ell(\mathbf{x}) = \mathrm{tr}\left((A - \mathbf{x}\mathbf{x}^T)^T(A - \mathbf{x}\mathbf{x}^T)\right).$$

展开后得到：

$$\ell(\mathbf{x}) = \mathrm{tr}(A^TA) - 2\mathrm{tr}(A\mathbf{x}\mathbf{x}^T) + \mathrm{tr}(\mathbf{x}\mathbf{x}^T\mathbf{x}\mathbf{x}^T).$$

第一项 tr(A^TA)

$$d\,\mathrm{tr}(A^TA) = 0.$$

第二项 -2**tr**$(A\mathbf{x}\mathbf{x}^T)$

$$d\,\mathrm{tr}(A\mathbf{x}\mathbf{x}^T) = \mathrm{tr}(A(d\mathbf{x})\mathbf{x}^T + A\mathbf{x}(d\mathbf{x}^T)).$$

$$d\,\mathrm{tr}(A\mathbf{x}\mathbf{x}^T) = 2\mathrm{tr}(A\mathbf{x}d\mathbf{x}^T).$$

$$d\left(-2\mathrm{tr}(A\mathbf{x}\mathbf{x}^T)\right) = -4\mathrm{tr}(A\mathbf{x}d\mathbf{x}^T).$$

第三项 tr$(\mathbf{x}\mathbf{x}^T\mathbf{x}\mathbf{x}^T)$

$$\mathrm{tr}(\mathbf{x}\mathbf{x}^T\mathbf{x}\mathbf{x}^T) = (\mathbf{x}^T\mathbf{x})^2 = \|\mathbf{x}\|^4.$$

对 $\|\mathbf{x}\|^4$ 求微分：

$$d\left(\|\mathbf{x}\|^4\right) = 4\|\mathbf{x}\|^2\mathbf{x}^Td\mathbf{x}.$$

$$d\ell = -4\mathrm{tr}(A\mathbf{x}d\mathbf{x}^T) + 4\|\mathbf{x}\|^2\mathbf{x}^Td\mathbf{x}.$$

$$d\ell = \left(-4A\mathbf{x} + 4\|\mathbf{x}\|^2\mathbf{x}\right)^T d\mathbf{x}.$$

$$\nabla\ell(\mathbf{x}) = -4A\mathbf{x} + 4\|\mathbf{x}\|^2\mathbf{x}.$$

$$\nabla\ell(\mathbf{x}) = 4\left(\|\mathbf{x}\|^2\mathbf{x} - A\mathbf{x}\right).$$

梯度向量为：

$$\boxed{\nabla\ell(\mathbf{x}) = 4\left(\|\mathbf{x}\|^2\mathbf{x} - A\mathbf{x}\right)}.$$

Problem 7. 设 $f:\mathbb{R}^n \to \mathbb{R}$ 为正定二次型 $f(x) = x^\top Qx$，多变量正态分布为：

$$p(x|\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)\right),$$

计算积分：

$$J(\mu,\Sigma) = \int_{\mathbb{R}^n} f(x)p(x|\mu,\Sigma)\,dx.$$

Ans.

$J(\mu, \Sigma)$ 等价于二次型 $f(x)$ 在正态分布下的期望：

$$J(\mu, \Sigma) = \mathbb{E}_{x\sim\mathcal{N}(\mu,\Sigma)}\left[x^\top Q x\right].$$

$$x^\top Q x = \mathrm{tr}(x^\top Q x) = \mathrm{tr}(Q x x^\top).$$

$$\mathbb{E}[x^\top Q x] = \mathrm{tr}\left(Q \cdot \mathbb{E}[x x^\top]\right).$$

对于 $x \sim \mathcal{N}(\mu, \Sigma)$，其二阶矩为：

$$\mathbb{E}[x x^\top] = \Sigma + \mu\mu^\top.$$

$$\mathbb{E}[x^\top Q x] = \mathrm{tr}\left(Q(\Sigma + \mu\mu^\top)\right).$$

$$\mathrm{tr}\left(Q(\Sigma + \mu\mu^\top)\right) = \mathrm{tr}(Q\Sigma) + \mathrm{tr}(Q\mu\mu^\top).$$

$$\mathrm{tr}(Q\mu\mu^\top) = \mathrm{tr}(\mu^\top Q \mu) = \mu^\top Q \mu.$$

$$\boxed{J(\mu, \Sigma) = \mu^\top Q \mu + \mathrm{tr}(Q\Sigma)}.$$