

优化实用算法: 第四次作业

2022 年 5 月 28 日

指导老师: 徐翔

徐圣泽 3190102721

Problem 1

1. Let $B_k \in R^{n \times n}$ symmetric positive definite. p_k solves $\min q_k(p) = \frac{1}{2}p^T B_k p + \nabla f(x_k)^T p + f(x_k)$. Try to prove p_k is a decent direction of f at x_k .

解：根据条件我们有

$$B_k^T p_k + \nabla f(x_k) = 0$$

其中 B_k 可逆，故可以解得 $p_k = -B_k^{-1} \nabla f(x_k)$

代入下式得到

$$\begin{aligned} f(x_k + \alpha p_k) &= f(x_k) + \alpha \nabla f(x_k)^T p_k + \alpha o(\|p_k\|) \\ &= f(x_k) + \alpha \nabla f(x_k)^T (-B_k^{-1} \nabla f(x_k)) + \alpha o(\|p_k\|) \\ &= f(x_k) - \alpha \nabla f(x_k)^T B_k^{-1} \nabla f(x_k) + \alpha o(\|p_k\|) \end{aligned}$$

因此有

$$f(x_k + \alpha p_k) - f(x_k) = -\alpha \nabla f(x_k)^T B_k^{-1} \nabla f(x_k) + \alpha o(\|p_k\|)$$

上式中 $\nabla f(x_k)^T B_k^{-1} \nabla f(x_k) > 0$, B_k^{-1} 是对称正定矩阵，则当 $\alpha \rightarrow 0$ 时， $f(x_k + \alpha p_k) - f(x_k) < 0$ ，因此 p_k 是 f 在 x_k 的下降方向。

Problem 2

2. Let $q_k(p) = \frac{1}{2}p^T B_k p + \nabla f(x_k)^T p + f(x_k)$. Try to prove the Cauchy point is the minimizer of $q_k(p)$ along $\nabla f(x_k)$

$$p_k^C = -\tau_k \frac{\Delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k)$$

where

$$\tau_k = \begin{cases} 1, & \text{if } \nabla f(x_k)^T B_k \nabla f(x_k) \leq 0 \\ \min\{\|\nabla f(x_k)\|^3 / (\nabla f(x_k)^T B_k \nabla f(x_k)), 1\}, & \text{otherwise.} \end{cases}$$

解：取 $p_k^s = -\frac{\Delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k)$ ，则有 $q_k(\tau p_k^s) = \frac{1}{2} \tau^2 p_k^{sT} B_k p_k^s + \tau \nabla f(x_k)^T p_k^s + f(x_k)$ 。

下面分情况进行讨论。

(1) 如果 $\nabla f(x_k)^T B_k \nabla f(x_k) \leq 0$ ，由函数图像可知，在 $\tau > 0$ 时随 τ_k 增大， $q_k(\tau p_k^s)$ 的值减小，因此 τ_k 取不超过信赖域范围内最大值 1 时， $q(p_k^c)$ 取最小值。

(2) 如果 $\nabla f(x_k)^T B_k \nabla f(x_k) > 0$ ，则 $\tau_0 = -\frac{\nabla f(x_k)^T p_k^s}{p_k^{sT} B_k p_k^s}$ 是 $q_k(\tau p_k^s)$ 关于 τ 的最小值点。

(i) 当 p_k^c 在信赖域范围内时， $\tau_0 \leq 1$ ，此时 $\tau_k = \tau_0$ 就是我们需要的值。化简 τ_0 可知，此时有

$$\tau_k = \frac{\|\nabla f(x_k)\|^3}{\Delta_k \nabla f(x_k)^T B_k \nabla f(x_k)}$$

(ii) 当 p_k^c 在信赖域范围外时， $\tau_0 > 1$ ，此时根据函数图像可知 $q_k(\tau p_k^s)$ 在 $\tau \in [0, \tau_0]$ 范围内下降，又 $\tau \in [0, 1]$ ，故 τ_k 取 1 时是我们的目标点。

综上所述， $\tau_k = \begin{cases} 1, & \text{if } \nabla f(x_k)^T B_k \nabla f(x_k) \leq 0 \\ \min\{\frac{\|\nabla f(x_k)\|^3}{\Delta_k \nabla f(x_k)^T B_k \nabla f(x_k)}, 1\}, & \text{otherwise.} \end{cases}$ ，原命题得证。

Problem 3

3. If symmetric $B \in R^{n \times n}$ has factorization $B = Q\Lambda Q^T$ where $Q = (q_1, q_2, \dots, q_n)$ is orthogonal, $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. Try to prove

$$\begin{cases} (B + \lambda I)p = -g \\ \|p\| = \Delta \end{cases}$$

solution can be expressed by

$$p(\lambda) = -\sum_{i=1}^n \frac{q_i^T g}{\lambda_i + \lambda} q_i.$$

Furthermore, try to prove

$$\frac{d}{d\lambda} (\|p(\lambda)\|^2) = -2 \sum_{i=1}^n \frac{(q_i^T g)^2}{(\lambda_i + \lambda)^3}$$

解：我们将 $p(\lambda) = -\sum_{i=1}^n \frac{q_i^T g}{\lambda_i + \lambda} q_i$ 代入 $(B + \lambda I)p = -g$ 后得到

$$\begin{aligned} (B + \lambda I)p &= \sum_{i=1}^n (\lambda_i q_i q_i^T + \lambda e_i e_i^T) \left(-\sum_{i=1}^n \frac{q_i^T g}{\lambda_i + \lambda} q_i \right) \\ &= -\sum_{i=1}^n (\lambda_i q_i q_i^T + \lambda e_i e_i^T) \sum_{i=1}^n \frac{q_i^T g}{\lambda_i + \lambda} q_i \\ &= -\left(\sum_{i=1}^n \frac{\lambda_i q_i q_i^T q_i q_i^T g}{\lambda_i + \lambda} + \lambda E \sum_{i=1}^n \frac{q_i q_i^T g}{\lambda_i + \lambda} \right) \\ &= -\left(\sum_{i=1}^n \frac{\lambda_i q_i q_i^T g}{\lambda_i + \lambda} + \lambda \sum_{i=1}^n \frac{q_i q_i^T g}{\lambda_i + \lambda} \right) \\ &= -\sum_{i=1}^n q_i q_i^T g \\ &= -QE Q^T g = -g \end{aligned}$$

因为上述过程是可逆的，则说明 $p(\lambda) = -\sum_{i=1}^n \frac{q_i^T g}{\lambda_i + \lambda} q_i$ 是方程组的解，这里我们并未用到 $\|p\| = \Delta$ 的条件。

进一步地，有

$$\begin{aligned} \|p(\lambda)\|^2 &= p(\lambda)^T \cdot p(\lambda) \\ &= \sum_{i=1}^n \frac{q_i^T g}{\lambda_i + \lambda} q_i^T \cdot \sum_{j=1}^n \frac{q_j^T g}{\lambda_j + \lambda} q_j^T \\ &= \sum_{i,j} \frac{(q_i^T g)(q_j^T g)}{(\lambda_i + \lambda)(\lambda_j + \lambda)} \delta_{i,j} \end{aligned}$$

其中

$$\delta_{i,j} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

因此有

$$\|p(\lambda)\|^2 = \sum_{i=1}^n \frac{(q_i^T g)^2}{(\lambda_i + \lambda)^2}$$

求导得到

$$\frac{d}{d\lambda}(\|p(\lambda)\|^2) = -2 \sum_{i=1}^n \frac{(q_i^T g)^2}{(\lambda_i + \lambda)^3}$$

Problem 4

4. Let $p_k = \arg \min\{m_k(p) : \|p\| \leq \Delta_k, s \in \text{span}[g_k, B_k^{-1}g_k]\}$, where $m_k(p) = f(x_k) + g_k^T p + \frac{1}{2}p^T B_k p$, B_k is symmetric positive. Try to find the explicit solution p_k .

解:

如果 $p_B = -B_k^{-1}g_k$ 并且 $\|p_B\| \leq \Delta_k$, 则 $p_k = p_B = -B_k^{-1}g_k$ 。

如果 $\|p_B\| > \Delta_k$, 令 $p = p^U + (1 - \tau)(p^U - p^B)$, $\tau \in [1, 2]$, 其中

$$p^U = -\frac{g_k^T g_k}{g_k^T B_k g_k} g_k$$

因此 $p(\tau) = -\frac{g_k^T g_k}{g_k^T B_k g_k} g_k + (1 - \tau)(B_k^{-1}g_k - \frac{g_k^T g_k}{g_k^T B_k g_k} g_k)$, 显然 $\|p^U + (1 - \tau)(p^U - p^B)\|$ 随 $\|(1 - \tau)(p^U - p^B)\|$ 增大而增大。

当 $\tau \in [1, 2]$ 增大时, $\|p\| = \|p^U + (1 - \tau)(p^U - p^B)\|$ 增大, 下面证明 $m_k(p(\tau))$ 随 τ 增大而减小。

$$\begin{aligned} m_k(p(\tau)) &= m_k(p^U + (1 - \tau)(p^U - p^B)) \\ &= f + g^T p + \frac{1}{2}p^T B p \\ &= f + g^T p^U + (\tau - 1)[g^T(p^B - p^U) + \frac{1}{2}p^{UT} B(p^B - p^U) \\ &\quad + \frac{1}{2}(p^B - p^U)^T B p^U] + (\tau - 1)^2 \frac{1}{2}(p^B - p^U)^T B(p^B - p^U) \end{aligned}$$

令

$$\tau_0 = \frac{g^T(p^B - p^U) + \frac{1}{2}p^{UT} B(p^B - p^U) + \frac{1}{2}(p^B - p^U)^T B p^U}{(p^B - p^U)^T B(p^B - p^U)}$$

则 $m_k(p(\tau))$ 会在 $\tau < \tau_0$ 时下降, 在 $\tau > \tau_0$ 时增大。

在无约束条件下, $\tilde{\tau} = 2$ (此时 $p(\tilde{\tau}) = p^B$) 是全局最小值点, 这说明 $\tau_0 = 2 = \tilde{\tau}$, 故 $\tau \in [1, 2]$ 时 $m_k(p(\tau))$ 随 τ 增大而减小。

在 $\|p\| \leq \Delta_k, s \in \text{span}[g_k, B_k^{-1}g_k]$ 的约束条件下, p_k 的显式解由 $\|p^U + (1 - \tau)(p^U - p^B)\| = \Delta_k^2$ 决定, 这是因为 $\tau \in [1, 2]$ 时 $\|p(\tau)\|$ 增大而 $m(p(\tau))$ 减小。

综上, 有

$$p_k = \begin{cases} -B_k^{-1}g_k, & \| -B_k^{-1}g_k \| \leq \Delta_k \\ -\frac{g_k^T g_k}{g_k^T B_k g_k} g_k, & \| -B_k^{-1}g_k \| > \Delta_k \end{cases}$$