

hw3

顾格非 3210103528

Q1

Let $g_k = \nabla f(x_k)$, since p_k solves $\min \frac{1}{2} p^T B_k p + g_k^T p + f(x_k)$

$$\Rightarrow p_k = -B_k^{-1} g_k$$

$$\text{so } \langle p_k, -g_k \rangle = \langle -B_k^{-1} g_k, -g_k \rangle = g_k^T B_k^{-1} g_k.$$

and B_k is symmetric positive definite, so $\langle p_k, -g_k \rangle > 0$.

所以 p_k is a decent direction of f at x_k .

Q2

$$F(\alpha) = q_k(\alpha g_k) = \frac{1}{2} (g_k^T B_k g_k) \alpha^2 + (g_k^T g_k) \alpha + f_k$$

$$\text{s.t. } \|\alpha g_k\| \leq \Delta_k, \quad |\alpha| \leq \frac{\Delta_k}{\|g_k\|}$$

$$F(\alpha)' = (g_k^T B_k g_k) \alpha + (g_k^T g_k)$$

$$F(\alpha_0) = 0 \Rightarrow \alpha_0 = -\frac{\|g_k\|^2}{g_k^T B_k g_k}$$

我们分情况讨论：

1. $g_k^T B_k g_k \leq 0$ 时, $F_{\min}(\alpha) = F\left(-\frac{\Delta_k}{\|g_k\|}\right)$

$$-\frac{\Delta_k}{\|g_k\|} \cdot g_k = \alpha g_k = -\tau_k \frac{\Delta_k}{\|g_k\|} g_k \Rightarrow \tau_k = 1$$

2. $g_k^T B_k g_k > 0$ 时, 不知道 α_0 与 $\frac{-\Delta_k}{\|g_k\|}$ 的关系, 讨论:

○ $\alpha_0 \leq \frac{-\Delta_k}{\|g_k\|}$, 此时 $\arg \min F(\alpha) = -\frac{\Delta_k}{\|g_k\|} \Rightarrow \tau_k = 1$.

○ $\alpha_0 > \frac{-\Delta_k}{\|g_k\|}$, 此时 $\arg \min F(\alpha) = \alpha_0 = -\frac{\|g_k\|^2}{g_k^T B_k g_k}$

$$\alpha_0 g_k = -\frac{\|g_k\|^2}{g_k^T B_k g_k} \cdot g_k = -\tau_k \frac{\Delta_k g_k}{\|g_k\|} \Rightarrow \tau_k = \frac{\|g_k\|^3}{(\Delta_k) \cdot (g_k^T B_k g_k)}$$

综上, 有:

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

Q3

$$\begin{aligned}
P &= -(B + \lambda I)^{-1}g = -(Q\Lambda Q^T + \lambda I)^{-1}g \\
&= -(Q\Lambda Q^T + \lambda Q Q^{-1}I)^{-1}g = -(Q\Lambda Q^T + \lambda Q Q^T I)^{-1}g - Q(\Lambda + \lambda I)^{-1}Q^T g \\
&= -(q_1, q_2, \dots, q_n)(\Lambda + \lambda I)^{-1} \begin{pmatrix} q_1^T g \\ q_2^T g \\ \vdots \\ q_n^T g \end{pmatrix} \\
&= - \left(\frac{q_1}{\lambda_1 + \lambda}, \frac{q_2}{\lambda_2 + \lambda}, \dots, \frac{q_n}{\lambda_n + \lambda} \right) \cdot \begin{pmatrix} q_1^T g \\ q_2^T g \\ \vdots \\ q_n^T g \end{pmatrix} \\
&= - \sum_{i=1}^n \frac{q_i^T g}{\lambda_i + \lambda} q_i \\
\|P(\lambda)\|^2 &= p(\lambda)^T \cdot p(\lambda) = \left(\sum_{i=1}^n \frac{q_i^T g}{\lambda_i + \lambda} q_i^T \right) \cdot \left(\sum_{j=1}^n \frac{q_j^T g}{\lambda_j + \lambda} q_j \right) = \sum_{i=1}^n \frac{(q_i^T g)^2}{(\lambda_i + \lambda)^2} \\
\text{求导得: } \frac{d(\|p(\lambda)\|^2)}{d\lambda} &= -2 \sum_{i=1}^n \frac{(q_i^T g)^2}{(\lambda_i + \lambda)^3}
\end{aligned}$$

Q4

题目应该是 $p \in \text{span}[g_k, B_k^{-1}g_k]$ 吧。令 $p = \alpha g + \beta B^{-1}g$, $u = [\alpha, \beta]^T$

$$\begin{aligned}
m(p) &= f + (\alpha g + \beta B^{-1}g, g) + \frac{1}{2}(\alpha g + \beta B^{-1}g, \alpha Bg + \beta g) \\
&= f + \alpha \|g\|^2 + \beta (B^{-1}g, g) + \frac{\alpha^2}{2}(g, Bg) + \frac{\beta^2}{2}(B^{-1}g, g) + \alpha\beta \|g\|^2 \\
&= f + \tilde{g}^T u + \frac{1}{2}u^T \tilde{B}u
\end{aligned}$$

$$\tilde{g} = \begin{bmatrix} \|g\|^2 \\ \beta (B^{-1}g, g) \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} (Bg, g) & \|g\|^2 \\ \|g\|^2 & (B^{-1}g, g) \end{bmatrix}$$

于是我们发现 $m(p)$ 可以写成 u 的函数，不妨令 $h(u) := m(p)$

$$H(u) := \|p\|^2 = \alpha^2 \|g\|^2 + \beta^2 \|B^{-1}g\|^2 + 2\alpha\beta (g, B^{-1}g) = \frac{1}{2}u^T \bar{B}u$$

$$\text{其中 } \bar{B} = 2 \begin{bmatrix} \|g\|^2 & (B^{-1}g, g) \\ (B^{-1}g, g) & \|B^{-1}g\|^2 \end{bmatrix}$$

$$\text{于是 } \|p\| \leq \Delta \Rightarrow H(u) \leq \Delta^2$$

于是这个问题可以转化成: $\min_{u \in R^2} h(u), s.t. H(u) \leq \Delta^2$

显然当这是一个无约束优化问题时, $\text{argmin} h(u) = -\tilde{B}^{-1}\tilde{g}$ 。

当 $H(u) > \Delta^2$ 时, 等价于解 $\nabla h(u) + \lambda \nabla H(u) = 0, \lambda \geq 0$, 由于 $(\tilde{B} + \lambda \bar{B})$ 是对称正定的, 故 $u = -(\tilde{B} + \lambda \bar{B})^{-1}\tilde{g}$. 其中 λ 满足 $H\left(-(\tilde{B} + \lambda \bar{B})^{-1}\tilde{g}\right) = \Delta^2$

Q5

a题应该已经给出了吧

$$\nabla f(x) = \nabla r(x)^T r(x)$$

$$\nabla^2 f(x) = \nabla r(x)^T \nabla r(x) + r(x) \nabla^2 r_i(x)$$

要证明 $\nabla^2 f(x) \rightarrow \nabla r(x) \nabla r(x)^T$, 即证明 $x \rightarrow x^*$ 时, $r(x) \nabla^2 r_i(x) \rightarrow 0$ 。

$$\nabla^2 r_1(x) = \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix}, \nabla^2 r_2(x) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\begin{aligned} \text{代入计算 } r(x) \nabla^2 r_i(x) &= (x_2 - x_1^2) \cdot \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} + (1 - x_2) \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} -2(x_2 - x_1^2) & 0 \\ 0 & 0 \end{pmatrix} \rightarrow 0, \text{ 当 } x \rightarrow x^* \text{ 时。} \quad QED \end{aligned}$$

Q6

令 $\delta(x) = \|Ax + r\|_2^2$, 将这里的 x 看成是关于 μ 的函数 (因为 μ 确定了, x 也能从方程中解出来). 我们有 $x(\mu) = -(A^T A + \mu I)^{-1} A^T r$. 于是这其实是关于 μ 的函数 $\delta(\mu)$ 。

下面我们尝试证明 $\delta(\mu)$ 是递增的。

$$\begin{aligned} \delta(x) &= (Ax + r)^T (Ax + r) = x^T A^T A x + 2r^T A x + r^T r \\ \nabla \delta(x) &= 2(A^T A x + A^T r) = 2(-A^T r - \mu x + A^T r) = -2\mu x \\ x'(\mu) &= (A^T A + \mu I)^{-2} A^T r \\ \delta'(\mu) &= (\nabla \delta(x))^T x'(\mu) = -2\mu x^T (A^T A + \mu I)^{-2} A^T r \\ &= 2\mu x^T (A^T A + \mu I)^{-1} \left[-(A^T A + \mu I)^{-1} A^T r \right] \\ &= 2\mu x^T (A^T A + \mu I)^{-1} x > 0 \end{aligned}$$

于是 $\delta(\mu)$ 关于 μ 递增, 于是对于 $\mu_1 > \mu_2 > 0$, 有 $\|Ax_2 + r\|_2^2 < \|Ax_1 + r\|_2^2$ 。