1 第一题

考虑函数:

$$q_k(p) = \frac{1}{2}p^T B_k p + \nabla f(x_k)^T p + f(x_k)$$

对 p 求导，由于 p 使 $q_k(p)$ 最小，

$$B_k p_k + \nabla f(x_k) = 0, \text{ 因为} B_k \text{ 对称正定，所以可逆.}$$

解得:

$$p_k = -B_k^{-1}\nabla f(x_k)$$

于是:

$$\nabla f(x_k)^T p_k = -\nabla f(x_k)^T B_k^{-1}\nabla f(x_k) < 0$$

因此 p_k 是下降方向。

2 第二题

沿负梯度方向搜索:

$$p = -s \cdot \frac{\nabla f(x_k)}{\|\nabla f(x_k)\|}, \quad \text{其中} s \geq 0 \text{ 且} \|p\| = s \leq \Delta_k.$$

其中，s 是步长，Δ_k 是信赖域半径。

二次模型为:

$$q_k(p) = \frac{1}{2}\left(\frac{\nabla f(x_k)^T B_k \nabla f(x_k)}{\|\nabla f(x_k)\|^2}\right)s^2 - \|\nabla f(x_k)\| \cdot s + f(x_k).$$

记为:

$$q_k(p) = \frac{1}{2}cs^2 - \|d\|s + f(x_k),$$

其中 $d = \nabla f(x_k)$, $c = \dfrac{d^T B_k d}{\|d\|^2}$.

求导得:

$$\frac{dq_k}{ds} = cs - \|d\| = 0 \quad \Rightarrow \quad s^* = \frac{\|d\|^3}{d^T B_k d}.$$

柯西点定义为:

$$p_k^C = -\tau_k \frac{\Delta_k}{\|d\|} d,$$

其中 τ_k 的取值如下:

当 $d^T B_k d > 0$ 时:

- 若 $\dfrac{\|d\|^3}{d^T B_k d} \leq 1$，则 $\tau_k = \dfrac{\|d\|^3}{d^T B_k d}$，对应 $s = \tau_k \Delta_k = s^*$（内部最小值）。
- 若 $\dfrac{\|d\|^3}{d^T B_k d} > 1$，则 $\tau_k = 1$，对应 $s = \Delta_k$（边界最小值）。

当 $d^T B_k d \leq 0$ 时:

$$\tau_k = 1, \quad \text{对应} s = \Delta_k \text{(边界最小值）.}$$

综上，柯西点是 $q_k(p)$ 在梯度方向上的最小化点。

3 第三题

首先，

$$B + \lambda I = Q\Lambda Q^T + \lambda QQ^T = Q(\Lambda + \lambda I)Q^T.$$

代入方程：

$$(B + \lambda I)p = Q(\Lambda + \lambda I)Q^T p = -g.$$

左乘 Q^T 得：

$$(\Lambda + \lambda I)Q^T p = -Q^T g.$$

令 $y = Q^T p$，则：

$$(\Lambda + \lambda I)y = -Q^T g.$$

由于 $\Lambda + \lambda I$ 是对角阵，其逆为：

$$(\Lambda + \lambda I)^{-1} = \mathrm{diag}\left(\frac{1}{\lambda_1 + \lambda}, \frac{1}{\lambda_2 + \lambda}, \ldots, \frac{1}{\lambda_n + \lambda}\right).$$

因此，第 i 个分量为：

$$y_i = -\frac{q_i^T g}{\lambda_i + \lambda}, \quad i = 1, \ldots, n,$$

其中 $q_i^T g$ 是向量 $Q^T g$ 的第 i 个分量。

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

即第一部分得证：

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

接下来，

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

对 λ 求导：

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

即第二部分得证：

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

4 第四题

忽略约束条件，先求函数 $m_k(p)$ 的无约束最小值点。由于 B_k 对称正定，$m_k(p)$ 是严格凸二次函数。
令梯度为零：

$$\nabla_p m_k(p) = g_k + B_k p = 0 \quad \Rightarrow \quad p^* = -B_k^{-1} g_k.$$

即牛顿方向，显然属于子空间 $\mathrm{span}\{g_k, B_k^{-1} g_k\}$。
考虑约束 $\|p\| \le \Delta_k$。
若

$$\|B_k^{-1} g_k\| \le \Delta_k,$$

则最优解未超出信赖域范围，可直接采用：

$$p_k = -B_k^{-1} g_k.$$

若

$$\|B_k^{-1} g_k\| > \Delta_k,$$

则 p^* 超出允许步长，此时最小值出现在边界 $\|p\| = \Delta_k$ 上。

$$p_k = -\frac{\Delta_k}{\|B_k^{-1} g_k\|} B_k^{-1} g_k.$$

综上所述，

$$\boxed{p_k = \begin{cases} -B_k^{-1} g_k, & \text{if } \|B_k^{-1} g_k\| \le \Delta_k, \\ -\dfrac{\Delta_k}{\|B_k^{-1} g_k\|} B_k^{-1} g_k, & \text{otherwise.} \end{cases}}$$

5 第五题

a 小题已由题目给出，下面解答 b 小题：

$$\begin{aligned} \nabla f(x) &= \nabla r(x)^\top \nabla r(x) \\ \nabla^2 f(x) &= \nabla r(x)^\top \nabla r(x) + r_1(x) \nabla^2 r_1(x) + r_2(x) \nabla^2 r_2(x) \\ \nabla^2 r_1(x) &= \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix}, \quad \nabla^2 r_2(x) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ \text{代入得: } r_1(x)\nabla^2 r_1(x) + r_2(x)\nabla^2 r_2(x) &= (x_2 - x_1^2) \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} + (1 - x_2) \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} -2(x_2 - x_1^2) & 0 \\ 0 & 0 \end{pmatrix} \to 0 \\ \text{当} x \to x^* = (1,1) \text{ 时} \end{aligned}$$

6 第六题

令 $\delta(x) = \|Ax + r\|_2^2$，将 x 看成是 u 的函数，即 $\delta(u)$，

$$x(u) = -(A^\top A + uI)^{-1} A^\top r.$$

尝试证明 $\delta(u)$ 是递增的。由 $A^\top Ax = -ux - A^\top r$，

$$\begin{aligned} \nabla \delta(x) &= 2A^\top (Ax + r) \\ &= 2(A^\top Ax + A^\top r) \\ &= 2(-ux - A^\top r + A^\top r) \\ &= -2ux. \end{aligned}$$

$$x'(u) = (A^\top A + uI)^{-2} A^\top r,$$

$$\begin{aligned}\delta'(u) = (\nabla\delta(x))^\top x'(u) &= -2ux^\top(A^\top A + uI)^{-2}A^\top r \\ &= 2ux^\top(A^\top A + uI)^{-1}\left[-(A^\top A + uI)^{-1}A^\top r\right] \\ &= 2ux^\top(A^\top A + uI)^{-1}x > 0.\end{aligned}$$

因此，$\delta(u)$ 关于 u 单调递增。对于 $u_1 > u_2 > 0$，有 $\|Ax_2 + r\|_2^2 < \|Ax_1 + r\|_2^2$。