hw3

Q1

$$Let\ g_k = \nabla f(x_k),\ \ since\ \ p_k\ \ solves\ \ min\ \frac{1}{2}p^TB_kp + g_k^Tp + f(x_k)$$
$$\Rightarrow p_k = -B_k^{-1}g_k$$
$$so\ \ \langle p_k, -g_k\rangle = \langle -B_k^{-1}g_k, -g_k\rangle = g_k^TB_k^{-T}g_k.$$
$$and\ B_k\ is\ symmetric\ \ positive\ \ definite, so\langle p_k, -g_k\rangle > 0.$$

所以 $p_k$is a decent direction of f at x_k.

Q2

$$F(\alpha) = q_k(\alpha g_k) = \frac{1}{2}\left(g_k^TB_kg_k\right)\alpha^2 + \left(g_k^Tg_k\right)\alpha + f_k$$
$$s.t.\quad \|\alpha g_k\| \leqslant \Delta_k,\quad |\alpha| \leqslant \frac{\Delta_k}{\|g_k\|}$$
$$F(\alpha)' = \left(g_k^TB_kg_k\right)\alpha + \left(g_k^Tg_k\right)$$
$$F(\alpha_0) = 0 \Rightarrow \alpha_0 = -\frac{\|g_k\|^2}{g_k^TB_kg_k}$$

我们分情况讨论：

1. $g_k^TB_kg_k \leq 0$时，$F_{\min}(\alpha) = F\left(-\frac{\Delta_k}{\|gk\|}\right)$
   $-\frac{\Delta_k}{\|g_k\|}\cdot g_k = \alpha g_k = -\tau_k\frac{\Delta_k}{\|g_k\|}g_k \Rightarrow \tau_k = 1$
2. $g_k^TB_kg_k > 0$时, 不知道α_0与$\frac{-\Delta k}{\|g_k\|}$的关系，讨论：
   - $\alpha_0 \leqslant \frac{-\Delta k}{\|g_k\|}$, 此时$\arg\min F(\alpha) = -\frac{\Delta_k}{\|g_k\|} \Rightarrow \tau_k = 1.$
   - $\alpha_0 > \frac{-\Delta k}{\|g_k\|}$, 此时$\text{argmin}F(\alpha) = \alpha_0 = -\frac{\|g_k\|^2}{g_k^TB_kg_k}$
     $\alpha_0g_k = -\frac{\|g_k\|^2}{g_k^TB_kg_k}\cdot g_k = -\tau_k\frac{\Delta_kg_k}{\|g_k\|} \Rightarrow \tau_k = \frac{\|g_k\|^3}{(\Delta_k)\cdot(g_k^TB_kg_k)}$

综上，有：

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

Q3

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

Q4

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

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

$$\tilde{g}=\begin{bmatrix}\|g\|^2\\\beta\left(B^{-1}g,g\right)\end{bmatrix},\quad\tilde{B}=\begin{bmatrix}(Bg,g)&\|g\|^2\\\|g\|^2&\left(B^{-1}g,g\right)\end{bmatrix}$$
$$\text{于是我们发现}m(p)\text{可以写成}u\text{的函数，不妨令}h(u):=m(p)$$

$$H(u):=\|p\|^2=\alpha^2\|g\|^2+\beta^2\|B^{-1}g\|^2+2\alpha\beta\left(g,B^{-1}g\right)=\frac{1}{2}u^T\bar{B}u$$
$$\text{其中}\bar{B}=2\begin{bmatrix}\|g\|^2&\left(B^{-1}g,g\right)\\\left(B^{-1}g,g\right)&\left\|B^{-1}g\right\|^2\end{bmatrix}$$
$$\text{于是}\|p\|\leq\Delta\Rightarrow H(u)\leq\Delta^2$$
$$\text{于是这个问题可以转化成：}\min_{u\in R^2}h(u),s.t.H(u)\leq\Delta^2$$

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

当$H(u)>\Delta^2$时，等价于解$\nabla h(u)+\lambda\nabla H(u)=0,\lambda\geq0,$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

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

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

$$
\nabla^2 r_1(x)=\begin{pmatrix}-2 & 0 \\ 0 & 0\end{pmatrix}, \nabla^2 r_2(x)=\begin{pmatrix}0 & 0 \\ 0 & 0\end{pmatrix}
$$
$$
\text{代入计算} r(x) \nabla^2 r_i(x)=\left(x_2-x_1^2\right) \cdot\begin{pmatrix}-2 & 0 \\ 0 & 0\end{pmatrix}+\left(1-x_2\right)\begin{pmatrix}0 & 0 \\ 0 & 0\end{pmatrix}
$$
$$
=\begin{pmatrix}-2\left(x_2-x_1^2\right) & 0 \\ 0 & 0\end{pmatrix} \rightarrow 0, \text{当} x \rightarrow x^* \text{时。} QED
$$

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$x(\mu)=-\left(A^TA+\mu I\right)^{-1}A^Tr$。于是这其实是关于$\mu$的函数$\delta(\mu)$。

下面我们尝试证明$\delta(\mu)$是递增的。

$$
\delta(x)=(Ax+r)^T(Ax+r)=x^TA^TAx+2r^TAx+r^Tr
$$
$$
\nabla\delta(x)=2\left(A^TAx+A^Tr\right)=2\left(-A^Tr-\mu x+A^Tr\right)=-2\mu x
$$
$$
x'(\mu)=\left(A^TA+\mu I\right)^{-2}A^Tr
$$

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

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