1

（1）记$f(x)=(x_1-2)^2+2(x_2-1)^2$，写为如下形式

$$\min_x f(x)=\frac{1}{2}x^TQx+q^Tx+C$$

其中$Q=\begin{bmatrix}2 & 0\\ 0 & 4\end{bmatrix}$，$q=(-4,-4)^T$，$C=5$。

求解过程中，β的值由PRP算法给出

$$\beta_k^{\mathrm{PRP}}=\frac{\nabla f(x^k)^T(\nabla f(x^k)-\nabla f(x^{k-1}))}{\|\nabla f(x^{k-1})\|^2}$$

下降步长为

$$\alpha_k=-\frac{\nabla f(x^{(k)})^Td^{(k)}}{{d^{(k)}}^TQd^{(k)}}$$

对f求导，$g(x)=\nabla f(x)=Qx+q=(2x_1-4,4x_2+4)^T$。起始点为$x^{(1)}=(1,3)^T$。

第一次迭代：$g^{(1)}=(-2,8)^T$，$d^{(1)}=-g^{(1)}=(2,-8)^T$，$\alpha^{(1)}=\frac{17}{66}$，$x^{(2)}=x^{(1)}+\alpha^{(1)}d^{(1)}=\left(\frac{50}{33},\frac{31}{33}\right)$。

第二次迭代：$g^{(2)}=(-\frac{32}{33},-\frac{8}{33})^T$，$\beta^{(2)}=\frac{16}{1089}$，$d^{(2)}=-g^{(2)}+\beta^{(2)}d^{(1)}=(\frac{1088}{1089},\frac{136}{1089})^T$，$\alpha^{(2)}=\frac{33}{68}$，$x^{(3)}=x^{(2)}+\alpha^{(2)}d^{(2)}=(2,1)$。

$g\left(x^{(3)}\right)=0$，由于f是$\mathbb{R}$上的凸函数，$x^{(3)}=(2,1)^T$是全局极小点，$\min(x_1-2)^2+2(x_2-1)^2=0$。

（2）记$f(x)=2x_1^2+2x_1x_2+x_2^2+3x_1-4x_2$，写为如下形式

$$\min_x f(x)=\frac{1}{2}x^TQx+q^Tx$$

其中$Q=\begin{bmatrix}4 & 2\\ 2 & 2\end{bmatrix}$，$q=(3,-4)^T$。

求解过程中，β的值由FR算法给出

$$\beta_k^{\mathrm{FR}}=\frac{\|\nabla f(x^k)\|^2}{\|\nabla f(x^{k-1})\|^2}$$

下降步长为

$$\alpha_k=-\frac{\nabla f(x^{(k)})^Td^{(k)}}{{d^{(k)}}^TQd^{(k)}}$$

对f求导，$g(x)=\nabla f(x)=Qx+q=(4x_1+2x_2+3,2x_2+2x_1-4)^T$。起始点为$x^{(1)}=(3,4)^T$。

第一次迭代：$g^{(1)}=(23,10)^T$，$d^{(1)}=-g^{(1)}=(-23,10)^T$，$\alpha^{(1)}=\frac{629}{3236}$，$x^{(2)}=x^{(1)}+\alpha^{(1)}d^{(1)}=\left(-\frac{1578}{1073},\frac{1499}{729}\right)$。

第二次迭代：$g^{(2)}=(\frac{995}{809},-\frac{2809}{993})^T$，$\beta^{(2)}=\frac{93}{6148}$，$d^{(2)}=-g^{(2)}+\beta^{(2)}d^{(1)}=(\frac{-669}{424},\frac{2300}{859})^T$，$\alpha^{(2)}=\frac{809}{629}$，$x^{(3)}=x^{(2)}+\alpha^{(2)}d^{(2)}=(-3.5,5.5)$。

$g\left(x^{(3)}\right)=0$，由于f是$\mathbb{R}$上的凸函数，$x^{(3)}=(-3.5,5.5)^T$是全局极小点，$\min 2x_1^2+2x_1x_2+x_2^2+3x_1-4x_2=0$。

2

记

$$f(x)=\frac{1}{2}x^TAx$$

有

$$\nabla f=Ax$$

令$x'=\bar{x}-\alpha(\bar{x}-b)$，在$x=x'$处应用泰勒展开

$$\begin{aligned}\frac{1}{2}(x')^TAx'&=\frac{1}{2}\bar{x}^TA\bar{x}-\alpha(A\bar{x})^T(\bar{x}-b)+\frac{1}{2}\alpha^2(\bar{x}-b)^TA(\bar{x}-b)\\&=\frac{1}{2}\bar{x}^TA\bar{x}-\alpha\bar{x}^TA(\bar{x}-b)+\frac{1}{2}\alpha^2(\bar{x}-b)^TA(\bar{x}-b)\end{aligned}$$

假设$\bar{x}^TA(\bar{x}-b)\neq 0$，则$\bar{x}^TA(\bar{x}-b)>0$或$\bar{x}^TA(\bar{x}-b)<0$。

如果$\bar{x}^TA(\bar{x}-b)>0$，使$0<\alpha\le 1$，$x'=\bar{x}-\alpha(\bar{x}-b)=(1-\alpha)\bar{x}+\alpha b$。对于凸集$S=\{x|x\le b\}$，由于$\bar{x}\in S,b\in S$，所以$x'\in S$。令$\epsilon=\frac{2\bar{x}^TA(\bar{x}-b)}{(\bar{x}-b)^TA(\bar{x}-b)}$，有$\epsilon>0$，取$0<x'\le\min(1,\epsilon)$。有$f(x')<f(\bar{x})$，与$\bar{x}$是最优解的前提矛盾。

如果$\bar{x}^TA(\bar{x}-b)<0$，使$\alpha<0$，有$x'=\bar{x}-\alpha(\bar{x}-b)>\bar{x}\ge b$。令$\epsilon=\frac{2\bar{x}^TA(\bar{x}-b)}{(\bar{x}-b)^TA(\bar{x}-b)}$，有$\epsilon<0$，取$x'<\epsilon$。有$f(x')<f(\bar{x})$，与$\bar{x}$是最优解的前提矛盾。

所以，当$\bar{x}$是最优解时，一定有$\bar{x}^TA(\bar{x}-b)=0$。即$\bar{x}^T$与$\bar{x}-b$关于A共轭。

3

令

$$f(\mathbf{x})=x_1^2+3x_2^2$$

f在(x_1,x_2)处的梯度为

$$\nabla f(\mathbf{x})=(2x_1,6x_2)^T$$

从$x^{(1)}=(1,-1)^T$处开始迭代，

第一次迭代：$g^{(1)}=(2,-6)^T,d^{(1)}=-H_1g^{(1)}=(2,4)^T,\alpha^{(1)}=\frac{5}{26},x^{(2)}=x^{(1)}+\alpha^{(1)}d^{(1)}=(\frac{18}{13},-\frac{3}{13})^T$

第二次迭代：$g^{(2)}=(\frac{36}{13},-\frac{18}{13})^T$，其中$H_2$由BFGS公式给出：

$$H_{k+1}=\left(I-\frac{s^k\left(y^k\right)^T}{\left(y^k\right)^Ts^k}\right)H_k\left(I-\frac{s^k\left(y^k\right)^T}{\left(y^k\right)^Ts^k}\right)^T+\frac{s^k\left(s^k\right)^T}{\left(y^k\right)^Ts^k}$$

其中$y^{(k)}=\nabla f\left(x^{(k+1)}\right)-\nabla f\left(x^{(k)}\right)=g^{(k+1)}-g^{(k)}$，$s^{(k)}=x^{(k+1)}-x^{(k)}$，于是

$$H_2=\begin{bmatrix}\frac{373}{338}&-\frac{17}{169}\\-\frac{17}{169}&\frac{31}{169}\end{bmatrix}$$

$d^{(2)}=-H_2g^{(2)}=(\frac{540}{169},-\frac{90}{169})^T,\alpha^{(2)}=-\frac{13}{30}$，$x^{(3)}=(0,0)^T$

$$\nabla f\left(x^{(3)}\right)=\mathbf{0}$$

由于f是$\mathbb{R}$上的凸函数，所以我们找到了全局极小值点$(0,0)$，对应的最小值为0。

4

记$H(x)=\nabla^2 f(x)$，由牛顿法可知

$$
\begin{aligned}
x^{k+1} &= x^k - H(x^k)^{-1}\nabla f(x^k) \\
x^{k+1} - x^* &= x^k - x^* - H(x^k)^{-1}\nabla f(x^k) \\
&= x^k - x^* + H(x^k)^{-1}(\nabla f(x^*) - \nabla f(x^k))
\end{aligned}
$$

令$g(t)=\nabla f(x_t + t(x^* - x_t))$，$g'(t)=H(x_t + t(x^* - x_t))(x^* - x_t)$

$$
\begin{aligned}
\nabla f(x^*) - \nabla f(x^k) &= g(1) - g(0) \\
&= \int_0^1 g'(t)dt \\
&= \int_0^1 H(x^k + t(x^* - x^k))(x^* - x^k)dt
\end{aligned}
$$

将上式带入$x_{t+1} - x^*$，

$$
\begin{aligned}
x^{k+1} - x^* &= x^k - x^* + H(x^k)^{-1}\int_0^1 H(x^k + t(x^* - x^k))(x^* - x^k)dt \\
&= H(x^k)^{-1}\left(H(x^k)(x^k - x^*) + \int_0^1 H(x^k + t(x^* - x^k))(x^* - x^k)dt\right) \\
&= H(x^k)^{-1}\int_0^1 (H(x^k + t(x^* - x^k)) - H(x^k))(x^* - x^k)dt
\end{aligned}
$$

取模长

$$
\begin{aligned}
||x^{k+1} - x^*|| &= \left|\left|H(x^k)^{-1}\int_0^1 (H(x^k + t(x^* - x^k)) - H(x^k))(x^* - x^k)dt\right|\right| \\
&\le ||H(x^k)^{-1}|| \cdot \left|\left|\int_0^1 (H(x^k + t(x^* - x^k)) - H(x^k))(x^* - x^k)dt\right|\right| \\
&\le ||H(x^k)^{-1}||\int_0^1 ||(H(x^k + t(x^* - x^k)) - H(x^k))(x^* - x^k)||dt \\
&\le ||H(x^k)^{-1}||\int_0^1 ||(H(x^k + t(x^* - x^k)) - H(x^k))|| \cdot ||(x^* - x^k)||dt \\
&= ||H(x^k)^{-1}|| \cdot ||x^* - x^k||\int_0^1 ||(H(x^k + t(x^* - x^k)) - H(x^k))||dt
\end{aligned}
$$

由于$H(x)$在x^*的邻域Lipschitz连续，$\exists L > 0$满足

$$
\begin{aligned}
\int_0^1 ||H(x^k + t(x^* - x^k)) - H(x^k)||dt &\le \int_0^1 L||t(x^* - x^k)||dt \\
&= L||x^* - x^k||\int_0^1 tdt \\
&= \frac{L}{2}||x^* - x^k||
\end{aligned}
$$

同样由Lipschitz连续，$\exists \mu > 0$，使得

$$
||H(x^k)^{-1}|| \le \frac{1}{\mu}
$$

于是

$$
||x^{k+1} - x^*|| \le \frac{1}{\mu}||x^* - x^k||\frac{L}{2}||x^* - x^k|| = \frac{L}{2\mu}||x^* - x^k||^2
$$

$$
\frac{||x^{k+1} - x^*||}{||x^k - x^*||^2} \le \frac{L}{2\mu}
$$

所以$\{x^k\}$收敛到x^*，且收敛速度是Q-二次的。我们证明了结论（1）与结论（2）。

对$\nabla f(x^k)$与$\nabla f(x^{k+1})$泰勒展开

$$\begin{aligned}\nabla f(x^k) &= \nabla f(x^*) + H(x^*)(x^k - x^*) + o(x^k - x^*)\\ \nabla f(x^{k+1}) &= \nabla f(x^*) + H(x^*)(x^{k+1} - x^*) + o(x^{k+1} - x^*)\end{aligned}$$

取模

$$\begin{aligned}\|\nabla f(x^k)\|^2 &= \|\nabla f(x^*) + H(x^*)(x^k - x^*) + o(x^k - x^*)\|^2\\ &= \|H(x^*)\|^2\|(x^k - x^*)\|^2 + o(\|x^k - x^*\|)\\ \|\nabla f(x^{k+1})\| &= \|H(x^*)\|\|(x^{k+1} - x^*)\| + o(\|x^{k+1} - x^*\|)\end{aligned}$$

所以，存在一个有界常数$\epsilon > 0$，在x^*邻域内

$$\begin{aligned}\frac{\|\nabla f(x^{k+1})\|}{\|\nabla f(x^k)\|^2} &= \frac{\|H(x^*)\|\|x^{k+1} - x^*\| + o(\|x^{k+1} - x^*\|)}{\|H(x^*)\|^2\|x^k - x^*\|^2 + o(\|x^k - x^*\|)}\\ &\le \frac{\|H(x^*)\|\|x^{k+1} - x^*\| + o(\|x^{k+1} - x^*\|)}{\|H(x^*)\|^2\|x^k - x^*\|^2} + \epsilon\\ &= \frac{1}{\|H(x^*)\|}\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^2} + \frac{o(\|x^{k+1} - x^*\|)}{\|H(x^*)\|^2\|x^k - x^*\|^2} + \epsilon\end{aligned}$$

由于$H(x)$在x^*的邻域Lipschitz连续，$\exists \mu' > 0$满足

$$||H(x^*)^{-1}|| \le \frac{1}{\mu'}$$

有

$$\frac{1}{\|H(x^*)\|}\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^2} \le \frac{L}{2\mu\mu'}$$

x^{k+1}距离x^*足够近，所以

$$\frac{o(\|x^{k+1} - x^*\|)}{\|H(x^*)\|^2\|x^k - x^*\|^2} \le \frac{\|x^{k+1} - x^*\|}{\|H(x^*)\|^2\|x^k - x^*\|^2} \le \frac{L}{2\mu\mu'^2}$$

所以

$$\frac{\|\nabla f(x^{k+1}) - \mathbf{0}\|}{\|\nabla f(x^k) - \mathbf{0}\|^2} = \frac{\|\nabla f(x^{k+1})\|}{\|\nabla f(x^k)\|^2} \le \frac{L(\mu' + 1)}{2\mu\mu'^2} + \epsilon$$

于是$\{\nabla f(x^k)\}$Q-二次收敛到0，结论（3）证毕。