1 (a) The second order Taylor expansion of $f(x)$ at x_0 is given below

$$f(x) = f(x_0) + (\nabla f(x_0))^T (x - x_0) + \frac{1}{2}(x-x_0)^T H(x_0)(x-x_0)$$

where $H(x_0)$ is Hessian matrix at x_0.

b (i) $\nabla^2 f(x) \preceq 25I$ for all x. From the above Taylor expansion we can write

$$f(x) \leq f(x_0) + (\nabla f(x_0))^T (x - x_0) + \frac{25}{2}\|x - x_0\|^2$$

Choosing $g(x) = f(x_0) + (\nabla f(x_0))^T (x - x_0) + $
$$+ \frac{25}{2}\|x - x_0\|^2$$

Note $g(x_0) = f(x_0)$ and $g(x) \geq f(x)$ for all x

Substituting $f(x_0) = 6$, $\nabla f(x_0) = [8, 4, 4, 2]^T$ and $x_0 = [4, 0, -2, 1]^T$

we get $g(x) = \frac{25}{2}\|x\|^2 + b^T x + c$ where

$b = [-92, 4, 54, -23]^T$ and $c = \frac{495}{2}$

(i) $Q = 25$ (ii) $b = [-92, 4, 54, -23]^T$ (iii) $c = \frac{495}{2}$

c) $\nabla g(x^*) = 0$ $x^* = -Q^{-1} b$

$$g(x^*) = -\frac{b^T Q^{-1} b}{2} + c$$
$$= 4$$

2(a) $f(x^{k+1}) - f(x^k) = -\alpha g_k^T g_k + \frac{\alpha^2}{2} g_k^T (2Q) g_k$

$$f(x^{k+1}) - f(x^k) = g_k^T Q g_k \left(\alpha^2 - 2\alpha \frac{g_k^T g_k}{2 g_k^T Q g_k} + \left(\frac{g_k^T g_k}{2 g_k^T Q g_k} \right)^2 - \frac{(g_k^T g_k)^2}{4 g_k^T Q g_k} \right)$$

$$= g_k^T Q g_k \left[\left(\alpha - \frac{g_k^T g_k}{2 g_k^T Q g_k} \right)^2 - \left(\frac{g_k^T g_k}{2 g_k^T Q g_k} \right)^2 \right]$$

$$= g_k^T Q g_k \times \alpha \times \left(\alpha - \frac{g_k^T g_k}{g_k^T Q g_k} \right)$$

for this to be < 0 α should lie between

$$0 < \alpha < \frac{g_k^T g_k}{g_k^T Q g_k}$$

$\lambda_{max} g_k^T g_k \geq g_k^T Q g_k \geq \lambda_{min} g_k^T g_k$ (proved in class) so

$$\therefore \frac{1}{\lambda_{max}} \leq \frac{g_k^T g_k}{g_k^T Q g_k} \leq \frac{1}{\lambda_{min}}$$

So if $\alpha \in \left(0, \frac{1}{\lambda_{max}}\right)$ then $f(x^{k+1}) - f(x^k) < 0$

so $a = 0$, $b = 1/\lambda_{max}$

(b) $f(x) = [x_1, x_2] \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [5 \quad 6] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 7$

Eigen values of $\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$ are $\lambda = 1, 5$

$\therefore$ if $\alpha \in \left(0, \frac{1}{5}\right)$ then gradient descent with constant step size α will converge

$\therefore a = 0$ and $b = 1/5$

4(a) $f(x) = \frac{1}{2}\log(x^2+1)$, $f'(x) = \frac{x}{x^2+1}$

$f''(x) = \frac{1-x^2}{(1+x^2)^2}$. $f''(0)>0$ and $f''(2)<0$. Therefore this function is neither convex nor concave

(b) For the function $f(x)$ to be L-smooth, $\exists\, L \geq 0$ such that $|f'(x) - f'(y)| \leq L|x-y|$ for all $x, y \in \mathbb{R}$

or $\left|\frac{f'(x) - f'(y)}{x-y}\right| \leq L$ for all $x, y \in \mathbb{R}$ and $x \neq y$

Consider interval $[x,y]$. If $f'(x)$ is differentiable then by mean value there exists $\alpha \in [x,y]$ such that $f''(\alpha) = \frac{f'(x) - f'(y)}{x-y}$

$\therefore |f''(\alpha)| \leq L$ for all $\alpha \in \mathbb{R}$. $f''(x) = \frac{1-x^2}{(1+x^2)^2}$ is continuous

As $\left|\frac{1-x^2}{(1+x^2)^2}\right| \leq 1$, so $L = 1$

(c) For the general case when $f''(x)$ changes sign we can't use gradient Descent to find minimum as the second order term in taylor expansion can be negative when gradient is 0.

But ~~it~~ for $f(x) = \frac{1}{2}\log(x^2+1)$, $f'(0) = 0$ and $f''(0) > 0$

The function $f(x)$ is convex in the interval $[-1, 1]$ ('around 0'), therefore when gradient Descent stops it will output a local minima $(=0)$

5 (a) $f'(x) = \dfrac{x}{x^2+1}$, $f''(x) = \dfrac{1-x^2}{(1+x^2)^2}$

$f'(0) = 0$ and $f''(0) > 0$

$\therefore$ $x^* = 0$ is a local minimum of f

(b) let $r_k = |x^k - x^*|$. The newton region $(x^* - r_0, x^* + r_0)$ is given by $r_0 < \dfrac{2\mu}{3M}$ where μ is the smallest eigen value of $\nabla^2 f(x^*)$ $[\nabla^2 f(x^*) \succeq \mu I]$ and M is the Lipschitz constant of the Hessian

$$\|H(x) - H(y)\| \leq M\|x-y\| \quad \forall x, y \in \mathbb{R}^d$$

For $f(x) = \frac{1}{2}\log(x^2+1)$, $x^* = 0$, $\nabla^2 f(x^*) = \dfrac{1-0^2}{(1+0^2)^2} = 1$

$\therefore$ $\mu = 1$. We need to find M.

$|f''(x) - f''(y)| \leq M|x-y|$. If $f''(x)$ is differentiable then by mean value theorem we can say

$$|f'''(\alpha)| \leq M \quad \text{for } \alpha \in \mathbb{R}$$

$$f'''(x) = \frac{2x(x^2-3)}{(x^2+1)^2}$$

$\therefore$ $M = \max |f'''(x)| \quad \forall x \in \mathbb{R}$

$= \max \left| \dfrac{2x(x^2-3)}{(x^2+1)^2} \right|$. Finding the maximum of this function is not easy in closed form

I found the max value numerically by varying x from 0 to 1 in steps of 0.01 (See the code for this Question)

$M = 1.457$ at $x = 0.36$

$\therefore$ $r_0 = \dfrac{2\mu}{3M} = \dfrac{2 \times 1}{3 \times 1.457} = 0.4570$

$\therefore$ $a = 0.4570$

$\therefore$ Newton method is effective for $x^{(0)} \in (-0.4570, 0.4570)$

6a) $x = Sy$, $g(y) = f(Sy)$

By chain rule $\nabla g(y) = S^T \nabla f(Sy) = S^T \nabla f(x)$

minimize $g(y) \Rightarrow \nabla g(y^*) = 0 \Rightarrow S^T \nabla f(x) = 0$

As S^T is non singular, so $\nabla f(x) = 0$ or $x = x^*$

$\therefore$ ~~$y^* = Sx^*$~~ $x^* = Sy^*$

(b) $x^{(k+1)} = x^{(k)} - \alpha_k SS^T \nabla f(x^k)$

$$(\nabla f(x^k))^T (-SS^T \nabla f(x^k)) = -(S^T \nabla f(x^k))^T (S^T \nabla f(x^k))$$

$$= -\|S^T \nabla f(x^k)\|^2 \leq 0$$

$\therefore$ $-SS^T \nabla f(x^{(k)})$ is a descent direction when $\nabla f(x^k) \neq 0$ as $-\|S^T \nabla f(x^k)\| < 0$ (this follows from S being non-singular and norm being positive)

(c) By chain rule

$$\nabla^2 g(y) = S^T \nabla^2 f(Sy)\, S$$
$$= S^T \nabla^2 f(x) S$$

(d) if $SS^T = (\nabla^2 f(x^k))^{-1}$ then $x^{k+1} = x^k - \alpha_k (\nabla^2 f(x^k))^{-1} \nabla f(x^k)$

This is similar to Newton Update Method (in newton method $\alpha_k = 1$)

(e) $\nabla^2 g(y) = S^T \nabla^2 f(x) S$

If $SS^T = (\nabla^2 f(x))^{-1} \Rightarrow S^T = S^{-1}(\nabla^2 f(x))^{-1}$

$$\nabla^2 g(y) = S^{-1}(\nabla^2 f(x))^{-1}(\nabla^2 f(x)) S = I$$

$\therefore$ $\nabla^2 g(y)$ is identity matrix with $\lambda_{max} = \lambda_{min} = 1$

$\therefore \rho = 1$

(i) $\alpha_k = \arg\min f(x^k + \alpha u)$

$$f(x^k + \alpha u) - f(x^k) = \alpha g_k^T u + \frac{\alpha^2}{2} u^T Q u \quad \text{(Taylor Expansion)}$$

$\min \left(\frac{\alpha^2}{2} u^T Q u + \alpha g_k^T u\right)$ is obtained at

$$\alpha = -\frac{g_k^T u}{u^T Q u} \qquad \therefore \alpha_k = \frac{-g_k^T u}{u^T Q u}$$

(ii) $x^{k+1} = x^k - \alpha_k S S^T \nabla f(x^k)$ $\quad \bar{\alpha} = \arg\min \left(f(x^{k+1})\right)$

$$f(x^{k+1}) = f(x^k) + g_k^T(-SS^T\alpha_k \nabla f(x^k)) + \frac{\alpha_k^2}{2} g_k^T SS^T Q\, SS^T g_k$$

$$\phi(\alpha) = \frac{\alpha^2}{2}(SS^T g_k)^T Q (SS^T g_k) - \alpha (S^T g_k)^T (S^T g_k) + f(x^k)$$

minimum of $\phi(\alpha)$ is attained at

$$\alpha = \frac{(S^T g_k)^T (S^T g_k)}{(SS^T g_k) Q (SS^T g_k)} \quad \left(\text{minimum of } ax^2 + bx + c \text{ is attained at } x = -\frac{b}{2a}\right)$$

$$\therefore \alpha_k = \frac{(S^T g_k)^T (S^T g_k)}{(SS^T g_k) Q (SS^T g_k)}$$

(iii) Gradient Descent converges in 55 iterations with the given starting point and function value $= -59517.0$

(iv) $\beta = \frac{1}{200} \rightarrow$ Gradient Descent converges in 31 iterations and function value $= -59517.0$

$\beta = \frac{1}{700} \rightarrow$ Gradient Descent converges in 10 iterations and function value $= -59517$

$\beta = \frac{1}{2000} \rightarrow$ Gradient Descent converges in 20 iterations and function value $= -59517$

$\therefore$ The best β is $1/700$ and algorithm converges to $x = [7, 1, 1, 1, 1, 1, 1, 1, 1, 1]$