Reference sol for HW3

1 (i) First note that

$$\nabla f(\vec{x}) = 2A\vec{x} + 2\vec{b}$$

then

$$\|\nabla f(\vec{x}) - \nabla f(\vec{y})\| = \|2A\vec{x} - 2A\vec{y}\|$$
$$= 2\|A(\vec{x}-\vec{y})\|$$
$$\leq 2\|A\|\cdot\|\vec{x}-\vec{y}\|$$

Recall the definition of $\|A\| := \max_{\vec{z}\neq\vec{0}} \frac{\|A\vec{z}\|}{\|\vec{z}\|}$, we conclude that the smallest Lipchiz constant of ∇f is $2\|A\|$.

(ii) let $\vec{z} = A\vec{x} + \vec{b} \in \mathbb{R}^m$, then

$$\nabla_x g(\vec{x}) = A^T \nabla_z f(A\vec{x}+\vec{b}),$$ thus

$$\|\nabla g(\vec{x}) - \nabla g(\vec{y})\| = \|A^T\nabla f(A\vec{x}+\vec{b}) - A^T\nabla f(A\vec{y}+\vec{b})\|$$
$$\leq \|A\|\cdot\|\nabla f(A\vec{x}+\vec{b}) - \nabla f(A\vec{y}+\vec{b})\|$$
$$\leq \|A\|\cdot L\,\|A\vec{x} - A\vec{y}\|$$
$$\leq \|A\|^2 L\,\|\vec{x}-\vec{y}\|$$

Finally, we conclude that the smallest Lipchiz constant of ∇g is $\tilde{L} = \|A\|^2 L$

2. proof one : Denote $Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12} & Q_{22} \end{bmatrix}$, then solve the following characteristic function

$$(\lambda - Q_{11})(\lambda - Q_{22}) - Q_{12}Q_{12} = 0$$

we obtain the eigenvalues are

$$\lambda_1 = \frac{Q_{11}+Q_{22} + \sqrt{(Q_{11}-Q_{22})^2 + 4Q_{12}^2}}{2}, \quad \lambda_2 = \frac{Q_{11}+Q_{22} - \sqrt{(Q_{11}-Q_{22})^2 + 4Q_{12}^2}}{2}$$

On the other hand, the scaled matrix

$$\tilde{Q} = D^{1/2} Q D^{1/2} = \begin{bmatrix} Q_{11}^{-1/2} & 0 \\ 0 & Q_{22}^{-1/2} \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12} & Q_{22} \end{bmatrix} \begin{bmatrix} Q_{11}^{-1/2} & 0 \\ 0 & Q_{22}^{-1/2} \end{bmatrix} = \begin{bmatrix} 1 & \frac{Q_{12}}{\sqrt{Q_{11}Q_{22}}} \\ \frac{Q_{12}}{\sqrt{Q_{11}Q_{22}}} & 1 \end{bmatrix}$$

has the eigenvalues

$$\check{\lambda}_1 = 1 + \frac{Q_{12}}{\sqrt{Q_{11} Q_{22}}}, \quad \check{\lambda}_2 = 1 - \frac{Q_{12}}{\sqrt{Q_{11} Q_{22}}}.$$

Case I: $Q_{12} = 0$, then

$$\kappa(Q) = \frac{Q_{11} + Q_{22} + \sqrt{(Q_{11} - Q_{22})^2}}{Q_{11} + Q_{22} - \sqrt{(Q_{11} - Q_{22})^2}} \geq 1$$

$$\kappa(\check{Q}) = 1$$

it is obvious that $\kappa(Q) \geq \kappa(\check{Q})$

Case II: $Q_{12} > 0$, then

$$\kappa(Q) = \frac{\frac{Q_{11}}{Q_{12}} + \frac{Q_{22}}{Q_{12}} + \sqrt{\left(\frac{Q_{11}}{Q_{12}} - \frac{Q_{22}}{Q_{12}}\right)^2 + 4}}{\frac{Q_{11}}{Q_{12}} + \frac{Q_{22}}{Q_{12}} - \sqrt{\left(\frac{Q_{11}}{Q_{12}} - \frac{Q_{22}}{Q_{12}}\right)^2 + 4}}$$

$$\kappa(\check{Q}) = \frac{1 + \frac{1}{\sqrt{\frac{Q_{11}}{Q_{12}} \frac{Q_{22}}{Q_{12}}}}}{1 - \frac{1}{\sqrt{\frac{Q_{11}}{Q_{12}} \frac{Q_{22}}{Q_{12}}}}}$$

Denote $a := \frac{Q_{11}}{Q_{12}}$, $b := \frac{Q_{22}}{Q_{12}}$, then we have $ab > 1$, $a > 0$, $b > 0$ from the positive definiteness of Q.

Since

$$\frac{a+b+\sqrt{(a-b)^2+4}}{a+b-\sqrt{(a-b)^2+4}} - \frac{1+\frac{1}{\sqrt{ab}}}{1-\frac{1}{\sqrt{ab}}}$$

$$= \frac{\left(a+b+\sqrt{(a-b)^2+4}\right)\left(1-\frac{1}{\sqrt{ab}}\right) - \left(a+b-\sqrt{(a-b)^2+4}\right)\left(1+\frac{1}{\sqrt{ab}}\right)}{\left(a+b-\sqrt{(a-b)^2+4}\right)\left(1-\frac{1}{\sqrt{ab}}\right)}$$

$$= \frac{-\frac{2}{\sqrt{ab}}(a+b) + 2\sqrt{(a-b)^2+4}}{\left(a+b-\sqrt{(a-b)^2+4}\right)\left(1-\frac{1}{\sqrt{ab}}\right)} \geq 0,$$ that is, $\kappa(Q) \geq \kappa(\check{Q})$

Case III : $Q_{12} < 0$, then

$$k(Q) = \frac{\frac{Q_{11}}{Q_{12}} + \frac{Q_{22}}{Q_{12}} - \sqrt{\left(\frac{Q_{11}}{Q_{12}} - \frac{Q_{22}}{Q_{12}}\right)^2 + 4}}{\frac{Q_{11}}{Q_{12}} + \frac{Q_{22}}{Q_{12}} + \sqrt{\left(\frac{Q_{11}}{Q_{12}} - \frac{Q_{22}}{Q_{12}}\right)^2 + 4}}$$

$$k(\check{Q}) = \frac{1 + \frac{1}{\sqrt{\frac{Q_{11}}{Q_{12}} \frac{Q_{22}}{Q_{12}}}}}{1 - \frac{1}{\sqrt{\frac{Q_{11}}{Q_{12}} \frac{Q_{22}}{Q_{12}}}}}$$

Denote $a := \frac{Q_{11}}{Q_{12}}$, $b := \frac{Q_{22}}{Q_{12}}$. then we have $ab > 1$, $a < 0$, $b < 0$ from the positive definiteness of Q.

Since

$$\frac{a+b + \sqrt{(a-b)^2+4}}{a+b + \sqrt{(a-b)^2+4}} - \frac{1 + \frac{1}{\sqrt{ab}}}{1 - \frac{1}{\sqrt{ab}}}$$

$$= \frac{-\frac{2}{\sqrt{ab}}(a+b) - 2\sqrt{(a-b)^2+4}}{\left(a+b + \sqrt{(a-b)^2+4}\right)\left(1 - \frac{1}{\sqrt{ab}}\right)} \geq 0, \text{ that is, } k(Q) \geq k(\check{Q})$$

proof two: Alternatively, we can first simplify the matrix Q by using the positive definiteness of Q. Note that if $Q_{12} \neq 0$, then

$$\frac{1}{Q_{12}} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12} & Q_{22} \end{bmatrix} = \begin{bmatrix} \frac{Q_{11}}{Q_{12}} & 1 \\ 1 & \frac{Q_{22}}{Q_{12}} \end{bmatrix} =: \hat{Q}$$ has the same eigenvalues with Q. Then we can consider different cases.

Case I: $Q_{12} = 0$, see the proof in proof one.

Case II: $Q_{12} > 0$, then denote $a := \frac{Q_{11}}{Q_{12}}$, $b := \frac{Q_{22}}{Q_{12}}$. and further we have $a > 0$, $b > 0$, $ab > 1$ from the positive definiteness of Q. Now instead of considering Q, we can consider $\hat{Q}$ and its scaled version, then compare their condition number. (the steps for comparsion are almost the same with proof one, we skip the detail here)

Case III: $Q_{12} < 0$, then denote $a := \frac{Q_{11}}{Q_{12}}$, $b := \frac{Q_{22}}{Q_{12}}$ and further we have $a < 0$, $b < 0$, $ab > 1$ from the positive definiteness of Q. Similarly as Case II. we consider $\hat{Q}$ and its scaled version. Again, the steps for compare the condition number are almost the same with proof one case III. we skip the detail here.

3. (i) # of iterations : ~ 3301

(ii) # of iterations: ~ 3732

(iii) # of iterations: ~ 1271

(iv) # of iterations: ~ 235

(v) # of iterations: ~ 104

we see that the diagonally scaled gradient method with backtracking line Search strategy is the most efficient way to solve this ill-conditioned quadratic problem.

4. (i) Note that

$$\frac{\partial f_1}{\partial x_1} = 1, \quad \frac{\partial f_1}{\partial x_2} = (5 - x_2)x_2 - 2 + (-2x_2 + 5)x_2 = -3x_2^2 + 10x_2 - 2$$

$$\frac{\partial f_2}{\partial x_1} = 1, \quad \frac{\partial f_2}{\partial x_2} = (x_2 + 1)x_2 - 14 + (2x_2 + 1)x_2 = 3x_2^2 + 2x_2 - 14$$

then $$\nabla f = \begin{bmatrix} 2f_1 \frac{\partial f_1}{\partial x_1} + 2f_2 \frac{\partial f_2}{\partial x_1} \\ 2f_1 \frac{\partial f_1}{\partial x_2} + 2f_2 \frac{\partial f_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Solving the above nonlinear system, we obtain three stationary points

$$(5, 4), \left(\frac{53 + 4\sqrt{22}}{3}, \frac{2 + \sqrt{22}}{3}\right), \text{ and } \left(\frac{53 - 4\sqrt{22}}{3}, \frac{2 - \sqrt{22}}{3}\right).$$

To charaterize the above three points, we can either use a graph or compute the Hessian matrix at these points.

For practice purpose, let's compute the Hessian matrix at each point.

First. Since we have

$$H(x_1,x_2)=\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix},$$

where

$$\frac{\partial^2 f}{\partial x_1^2} = 2\left(\frac{\partial f_1}{\partial x_1}\right)^2 + 2f_1 \frac{\partial^2 f_1}{\partial x_1^2} + 2\left(\frac{\partial f_2}{\partial x_1}\right)^2 + 2f_2 \frac{\partial^2 f_2}{\partial x_1^2}$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = 2\frac{\partial f_1}{\partial x_2}\frac{\partial f_1}{\partial x_1} + 2f_1 \frac{\partial^2 f_1}{\partial x_1 \partial x_2} + 2\frac{\partial f_2}{\partial x_2}\frac{\partial f_1}{\partial x_1} + 2f_2 \frac{\partial^2 f_2}{\partial x_1 \partial x_2}$$

$$\frac{\partial^2 f}{\partial x_2^2} = 2\left(\frac{\partial f_1}{\partial x_2}\right)^2 + 2f_1 \frac{\partial^2 f_1}{\partial x_2^2} + 2\left(\frac{\partial f_2}{\partial x_1}\right)^2 + 2f_2 \frac{\partial^2 f_2}{\partial x_2^2}$$

$$\frac{\partial^2 f_1}{\partial x_1^2} = 0, \quad \frac{\partial^2 f_2}{\partial x_1^2} = 0, \quad \frac{\partial^2 f_1}{\partial x_1 \partial x_2} = 0, \quad \frac{\partial^2 f_2}{\partial x_1 \partial x_2} = 0$$

$$\frac{\partial^2 f_1}{\partial x_2^2} = -6x_2 + 10, \quad \frac{\partial^2 f_2}{\partial x_2^2} = 6x_2 + 2$$

then at each stationary point, we have

$$H(5,4) = \begin{bmatrix} 4 & 64 \\ 64 & 3728 \end{bmatrix}, \quad H\left(\frac{53+4\sqrt{22}}{3}, \frac{2+\sqrt{22}}{3}\right) = \begin{bmatrix} 4 & \frac{4154}{193} \\ \frac{4154}{193} & -\frac{18662}{29} \end{bmatrix}$$

$$H\left(\frac{53-4\sqrt{22}}{3}, \frac{2-\sqrt{22}}{3}\right) = \begin{bmatrix} 4 & \frac{-5727}{107} \\ -\frac{5727}{107} & \frac{88385}{98} \end{bmatrix}$$

It is easy to check that $H(5,4)$ is positive definite, $H\left(\frac{53+4\sqrt{22}}{3}, \frac{2+\sqrt{22}}{3}\right)$ is indefinite, and $H\left(\frac{53-4\sqrt{22}}{3}, \frac{2-\sqrt{22}}{3}\right)$ is positive definite. Thus $\left(\frac{53+4\sqrt{22}}{3}, \frac{2+\sqrt{22}}{3}\right)$ is a saddle point, and both $(5,4)$ and $\left(\frac{53-4\sqrt{22}}{3}, \frac{2-\sqrt{22}}{3}\right)$ are strictly local minimum points. Further, we observe that $f(x_1,x_2) \geq 0$ on $\mathbb{R}^2$ and

$$f(5,4) = 0$$

$$f\left(\frac{53-4\sqrt{22}}{3}, \frac{2-\sqrt{22}}{3}\right) = \frac{6221}{127}$$

thus $(5,4)$ is a strict global minimum while $\left(\frac{53-4\sqrt{22}}{3}, \frac{2-\sqrt{22}}{3}\right)$ is a strict local minimum

6ii) $(-50, 7)$.

1. # of iterations: n 2252
2. # of iterations: n 8
3. # of iterations: n 29

converge to $(5,4)$

$(20, 7)$

1. # of iterations: n 2447
2. # of iterations: n 8
3. # of iterations: n 29

converge to $(5,4)$

$(20, -18)$

1. # of iterations: n 2466
2. # of iterations: n 18

converge to $\left(\frac{53-4\sqrt{22}}{3}, \frac{2-\sqrt{22}}{3}\right)$

3. # of iterations: does not converge

$(5, -10)$

1. # of iterations: n 2123 converge to $(5,4)$
2. # of iterations: n 13 converge to $\left(\frac{53-4\sqrt{22}}{3}, \frac{2-\sqrt{22}}{3}\right)$
3. # of iterations: does not converge.

damped Gauss-Newton's method fails to converge when the initial guesses are $(20,-18)$, $(5,-10)$.

For these two initial guesses, the squence $\{\vec{x}_k\}$ attempt to approach $\left(\frac{53-4\sqrt{22}}{3}, \frac{2-\sqrt{22}}{3}\right)$ when implement damped Gauss-Newton's method. However, the condition number of $J'J$ at $\left(\frac{53-4\sqrt{22}}{3}, \frac{2-\sqrt{22}}{3}\right)$ is around 1.36×10^{17} which meaning the inverse of $J'J$ is not accurate at all, then the search direction is not accurate at all. Thus the method fails to converge.