CMO Assignment 2

Adarsh Shah, 19473

September 29, 2021

Problem 1

1.a

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

1.b

From the given $g(x)$,

$$g(x) = g(x_0) + \nabla g(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)Q(x - x_0) \text{ and } \nabla g(x_0) = Qx_0 + b$$

Given $\nabla^2 f(x) \succ 0 \rightarrow f(x)$ is convex function.
Let $t(x) = g(x) - f(x) \geq 0$, it can be verified that t(x) is a convex function and is minimum at $x = x_0$ as $f(x_0) = g(x_0)$.

$$\begin{aligned} t(\lambda x + (1-\lambda)y) &= g(\lambda x + (1-\lambda)y) - f(\lambda x + (1-\lambda)y) \\ &\leq \lambda g(x) + (1-\lambda)g(y) - \lambda f(x) - (1-\lambda)f(y) \\ &= \lambda(g(x) - f(x)) + (1-\lambda)(g(y) - f(y)) \\ &= \lambda t(x) + (1-\lambda)t(y) \end{aligned}$$

Hence, expanding second order taylor series of $t(x)$ at $x = x_0$, we get

$$t(x) = (Qx_0 + b - \nabla f(x_0))^T (x - x_0) + \frac{1}{2}(x - x_0)(Q - \nabla^2 f(x))(x - x_0) \geq 0 \text{ for some } x \in \mathbf{R}^4$$

At $x = x_0$, $t(x)$ is minimum. Hence,

$$Q - \nabla^2 f(x) \succeq 0 \rightarrow Q \succeq 25\mathbf{I} \text{ and } \nabla t(x_0) = Qx_0 + b - \nabla f(x_0) = 0 \rightarrow b = \nabla f(x_0) - Qx_0$$

Let, $Q = 25\mathbf{I}$, we get $b = [-92, 4, 54, -23]^T$. From $c = g(x_0) - \frac{1}{2}x_0^T Q x_0 - b^T x0$, we get $c = 242.5$.

1.c

For a quadratic function, we know

$$x^* = -Q^{-1}b \text{ where } Q = 25\mathbf{I} \text{ and } b = [-92, 4, 54, -23]^T$$

Therefore, $x^* = [3.68, -0.16, -2.16, 0.92]$ and $g(x^*) = 3.99$.

Problem 2

2.a

Let $f(x^{(k+1)}) = \Phi(\alpha) = f(x^k) - \alpha g_k^T g_k + \alpha^2 \frac{g_k^T Q g_k}{2}$.
The function $\Phi(\alpha)$ is of quadratic form and the minimum value of $ax^2 - bx + c$ is obtained at $x = \frac{b}{2a}$. Therefore,

$$\alpha^* = \underset{\alpha > 0}{\text{argmin}} \Phi(\alpha) = \frac{g_k^T g_k}{g_k^T Q g_k} \in [\frac{1}{\lambda_{max}}, \frac{1}{\lambda_{min}}]$$

where λ_{max},λ_{min} are the largest and smallest eigenvalues of Q. Taking constant α in this interval ensures decrease after each iteration.

2.b

Given, $\underset{x \in \mathbf{R}^2}{f(x)} = x^t \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} x + [5, 6]^T x + 7$
Here, $Q = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$, the maximum and minimum eigenvalues of Q are $\lambda_{max} = 5$ and $\lambda_{min} = 1$.
Therefore, $\alpha^* \in [\frac{1}{5}, 1]$.

Problem 4

4.a

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

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

$$\therefore f^{''}(x) = \frac{1-x^2}{(x^2+1)^2} < 0 \text{ for some } x$$

Therefore, $f(x)$ is not a convex function.

4.b

Let, $f'(x) = \frac{x}{x^2+1}$ and $f'(y) = \frac{y}{y^2+1}$.

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

By showing $\frac{\|f'(x)-f'(y)\|}{\|x-y\|} \leq L$, we can find the Lipshictz constant.

$$\begin{aligned}
\therefore \frac{(f'(x)-f'(y))^2}{(x-y)^2} &= \frac{(y-x)^2(xy-1)^2}{(x^2+1)^2(y^2+1)^2(x-y)^2} \\
&= \frac{(xy-1)^2}{(x^2+1)^2(y^2+1)^2} \\
&\leq 1
\end{aligned}$$

Therefore, the Lipchitz constant L is 1 and $f \in C_1^1$.

4.c

We have proved that $f \in C_1^1$, $f(x) \geq 0$ and $\lim_{|x|\to\infty} f(x) \to \infty$. Hence, $f(x)$ is coercive and has global minimum at $x = 0$. These condition are sufficient for the gradient descent algorithm to converge according to global convergence theorem.

Problem 5

5.a

$f'(0) = 0$ and $f''(0) = 1 \geq 0$. Therefore, by second order sufficient conditions $x^* = 0$.

5.b

Newton Iterates : $x^{k+1} = x^k - \frac{f'(x^k)}{f''(x_k)}$ and $x^* = 0$

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

Therefore, $f(x) \in C^3$ and $f'(x) \in C^2$

$$x^{k+1} = x^k - \frac{f'(x^k)}{f''(x_k)}$$
$$\therefore x^{k+1} - x^* = x^k - x^* - \frac{f'(x^k) - f'(x^*)}{f''(x^k)}$$
$$\therefore x^{k+1} = -\frac{f'(x^k) - f'(x^*) - f''(x^k)(x^k - x^*)}{f''(x^k)}$$

We know, $f'(x^*) = f'(x^k) + \frac{1}{2}f''(x^k)(x^* - x^k) + \frac{1}{2}f'''(x)(x^* - x^k)^2$ for some $x \in [0, x^k]$

$$\therefore x^{k+1} = \frac{f'''(x)}{2f''(x^k)}(x^k)^2$$

For Newton method to converge, $|\frac{x^k f'''(x)}{2f''(x^k)}| < 1$,

$$\therefore |\frac{2x^2(x^2-3)(1+x^2)^2}{2(1+x^2)^3(1-x^2)}| < 1$$
$$\therefore |\frac{x^4 - 3x^2}{1-x^4}| < 1$$
$$\therefore 3x^2 - x^4 < 1 - x^4$$
$$\therefore x^2 < \frac{1}{3}$$

Hence, $x^0 \in (-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ if Newton method converges.

Problem 6

6.a

Given $x = Sy$, $x^* = Sy^*$ and $g(y) = f(Sy)$,
Let $x^* = \underset{x}{min} f(x)$, we know

$$f(x) \geq f(x^*), \forall x$$
$$\therefore f(Sy) \geq f(x^*) = f(Sy^*)$$
$$\therefore g(y) \geq g(y^*), \forall y$$

6.b

Since S is a non singular matrix, $B = SS^T$ is a positive definite matrix.
Let $u^k = -SS^T \nabla f(x^k)$, if $\nabla f(x^k)^T u^k \leq 0$ then u^k is a descent direction.

$$\nabla f(x^k)^T u^k = \nabla f(x^k)^T \cdot (-SS^T \nabla f(x^k)) = -\nabla f(x^k)^T \cdot B \nabla f(x^k)^T = -\lambda_{min} \|\nabla f(x^k)^T\|^2 \leq 0$$

λ_{min} is the minimum positive eigenvalue of SS^T. Hence, proved.

6.c

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

6.d

$x^{k+1} = x^k - (\nabla^2 f(x))^{-1} \nabla f(x)$ is Newton Method.

6.e

Given, $SS^T = (\nabla^2 f(x))^{-1}$

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

The maximum and minimum eigenvalue of I is 1.
Therefore, the condition number of $\nabla^2 g(y) = 0$.

6.f

i.

The optimal step size : $\alpha^* = -\frac{\nabla f(x)^T u}{u^T \nabla^2 f(x) u}$.

ii.

Taking $u = -SS^T \nabla f(x)$ we get,

$$\begin{aligned}\alpha^* &= -\frac{\nabla f(x)^T (-SS^T \nabla f(x))}{(-SS^T \nabla f(x))^T \nabla^2 f(x) (-SS^T \nabla f(x))} \\ &= \frac{\nabla f(x)^T SS^T \nabla f(x)}{\nabla f(x)^T SS^T \nabla^2 f(x) SS^T \nabla f(x)}\end{aligned}$$

iii

Iterations : 29

iv

$\beta = \dfrac{1}{200}$, $\;$ **iter** $= 21$

$(7.00000303, 1.00011404, 1.00001178, 1.00001178, 1.00001178, 1.00001178, 1.00001178, 1.00001178, 1.00001178, 1.00001178)$

$\beta = \dfrac{1}{700}$, $\;$ **iter** $= 10$ $Best$

$(6.99999672, 1.00032732, 1.0000353, 1.0000353, 1.0000353, 1.0000353, 1.0000353, 1.0000353, 1.0000353, 1.0000353)$

$\beta = \dfrac{1}{2000}$, $\;$ **iter** $= 15$

$(6.9999959, 1.00001664, 1.00000178, 1.00000178, 1.00000178, 1.00000178, 1.00000178, 1.00000178, 1.00000178, 1.00000178)$