1 [Q1]

Last week, we talked about the gradient and the hessian. Gradient is the direction which the function descends fastest. And we also talked about convergence analysis of convex optimziation. Especially, we talked about gradient descent and analyzed 2 special cases: M-smooth and strong convexity.

2 [Q2]

done

3 [Q3]

3.1 (a)

Assume $f(\boldsymbol{x}), f(\boldsymbol{y})$ are convex, then we have,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \langle \boldsymbol{x} - \boldsymbol{y}, \nabla f(\boldsymbol{y}) \rangle$$

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{y} - \boldsymbol{x}, \nabla f(\boldsymbol{x}) \rangle$$

Combine the 2 inequalities together and we can get,

$$\begin{aligned} f(\boldsymbol{x}) + f(\boldsymbol{y}) &\geq f(\boldsymbol{y}) + f(\boldsymbol{x}) + \langle \boldsymbol{x} - \boldsymbol{y}, \nabla f(\boldsymbol{y}) \rangle + \langle \boldsymbol{y} - \boldsymbol{x}, \nabla f(\boldsymbol{x}) \rangle \\ - \langle \boldsymbol{x} - \boldsymbol{y}, \nabla f(\boldsymbol{y}) \rangle - \langle \boldsymbol{y} - \boldsymbol{x}, \nabla f(\boldsymbol{x}) \rangle &\geq 0 \\ \langle \boldsymbol{y} - \boldsymbol{x}, \nabla f(\boldsymbol{y}) \rangle + \langle \boldsymbol{y} - \boldsymbol{x}, -\nabla f(\boldsymbol{x}) \rangle &\geq 0 \\ \langle \boldsymbol{y} - \boldsymbol{x}, \nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x}) \rangle &\geq 0 \end{aligned}$$

3.2 (b)

Since ∇f is monotone, let's define

$$\phi(\alpha) = f(\boldsymbol{x} + \alpha(\boldsymbol{y} - \boldsymbol{x}))$$

Then, we have $\phi'(\alpha) \geq \phi'(0)$ for all $\alpha > 0$ and $\alpha \in \boldsymbol{dom}\phi$ where,

$$\phi'(\alpha) = \nabla f(\boldsymbol{x} + \alpha(\boldsymbol{y} - \boldsymbol{x}))^T (\boldsymbol{y} - \boldsymbol{x})$$

Hence,

$$\begin{aligned} f(\boldsymbol{y}) = \phi(1) &= \phi(0) + \int_0^1 \phi'(\alpha) dt \\ &\geq \phi(0) + \phi'(0) \\ &= f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}) \end{aligned}$$

3.3 (c)

If $x > x^\star$, $f'(x) > 0$ and if $x < x^\star$, $f'(x) < 0$.

4 [Q4]

4.1 (a)

- In first part, I showed that (ii) $\rightarrow$ (i). since, $g(\boldsymbol{x})$ is convex in $\boldsymbol{domf}$, we can have,

$$\begin{aligned}
g(\boldsymbol{y}) &\geq g(\boldsymbol{x}) + \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle \\
f(\boldsymbol{y}) - \frac{m}{2}\|\boldsymbol{y}\|_2^2 &\geq f(\boldsymbol{x}) - \frac{m}{2}\|\boldsymbol{x}\|_2^2 + \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x}) - m\boldsymbol{x}\rangle \\
f(\boldsymbol{y}) &\geq f(\boldsymbol{x}) + \frac{m}{2}(\|\boldsymbol{y}-\boldsymbol{x}\|_2^2) + \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle - m\langle \boldsymbol{y}-\boldsymbol{x}, \boldsymbol{x}\rangle \\
f(\boldsymbol{y}) &\geq f(\boldsymbol{x}) + \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle + \frac{m}{2}\langle \boldsymbol{y}-\boldsymbol{x}, \boldsymbol{y}-\boldsymbol{x}\rangle \\
f(\boldsymbol{y}) &\geq f(\boldsymbol{x}) + \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle + \frac{m}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2
\end{aligned}$$

- In second part, I showed that (i) $\rightarrow$ (ii).

$$\begin{aligned}
f(\boldsymbol{y}) &\geq f(\boldsymbol{x}) + \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle + \frac{m}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 \\
f(\boldsymbol{y}) &\geq f(\boldsymbol{x}) + \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle + \frac{m}{2}\langle \boldsymbol{y}-\boldsymbol{x}, \boldsymbol{y}-\boldsymbol{x}\rangle \\
f(\boldsymbol{y}) &\geq f(\boldsymbol{x}) + \frac{m}{2}(\|\boldsymbol{y}-\boldsymbol{x}\|_2^2) + \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle - m\langle \boldsymbol{y}-\boldsymbol{x}, \boldsymbol{x}\rangle \\
f(\boldsymbol{y}) - \frac{m}{2}\|\boldsymbol{y}\|_2^2 &\geq f(\boldsymbol{x}) - \frac{m}{2}\|\boldsymbol{x}\|_2^2 + \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x}) - m\boldsymbol{x}\rangle
\end{aligned}$$

  If we define $g(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{m}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2$, then we can have,

$$g(\boldsymbol{y}) \geq g(\boldsymbol{x}) + \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle$$

4.2 (b)

- In first part, I showed that (ii) $\rightarrow$ (iii).
  Since $g(x)$ is convex, we can have,

$$f(\boldsymbol{y}) - \frac{m}{2}\|\boldsymbol{y}\|_2^2 \geq f(\boldsymbol{x}) - \frac{m}{2}\|\boldsymbol{x}\|_2^2 + \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x}) - m\boldsymbol{x}\rangle$$

  Simiarly, we can have,

$$f(\boldsymbol{x}) - \frac{m}{2}\|\boldsymbol{x}\|_2^2 \geq f(\boldsymbol{y}) - \frac{m}{2}\|\boldsymbol{y}\|_2^2 + \langle \boldsymbol{x}-\boldsymbol{y}, \nabla f(\boldsymbol{y}) - m\boldsymbol{y}\rangle$$

  Combine the inequalities together, we can have that,

$$\begin{aligned}
\langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x}) - m\boldsymbol{x}\rangle + \langle \boldsymbol{x}-\boldsymbol{y}, \nabla f(\boldsymbol{y}) - m\boldsymbol{y}\rangle &\leq 0 \\
\langle \boldsymbol{x}-\boldsymbol{y}, (\nabla f(\boldsymbol{x}) - m\boldsymbol{x}) - (\nabla f(\boldsymbol{y}) - m\boldsymbol{y})\rangle &\geq 0 \\
\langle \boldsymbol{x}-\boldsymbol{y}, \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\rangle &\geq \langle \boldsymbol{x}-\boldsymbol{y}, m(\boldsymbol{x}-\boldsymbol{y})\rangle \\
\langle \boldsymbol{x}-\boldsymbol{y}, \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\rangle &\geq m\|\boldsymbol{x}-\boldsymbol{y}\|_2^2
\end{aligned}$$

- In second part, I showed that (iii) $\rightarrow$ (ii).

$$\begin{aligned}
\langle \boldsymbol{x}-\boldsymbol{y}, \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\rangle &\geq m\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \\
\langle \boldsymbol{x}-\boldsymbol{y}, \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\rangle &\geq \langle \boldsymbol{x}-\boldsymbol{y}, m(\boldsymbol{x}-\boldsymbol{y})\rangle \\
\langle \boldsymbol{x}-\boldsymbol{y}, (\nabla f(\boldsymbol{x}) - m\boldsymbol{x}) - (\nabla f(\boldsymbol{y}) - m\boldsymbol{y})\rangle &\geq 0 \\
\langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x}) - m\boldsymbol{x}\rangle + \langle \boldsymbol{x}-\boldsymbol{y}, \nabla f(\boldsymbol{y}) - m\boldsymbol{y}\rangle &\leq 0
\end{aligned}$$

Hence, we can have that,

$$f(\boldsymbol{y})-\frac{m}{2}\|\boldsymbol{y}\|_2^2+f(\boldsymbol{x})-\frac{m}{2}\|\boldsymbol{x}\|_2^2 \geq f(\boldsymbol{x})-\frac{m}{2}\|\boldsymbol{x}\|_2^2++f(\boldsymbol{y})-\frac{m}{2}\|\boldsymbol{y}\|_2^2+\langle \boldsymbol{x}-\boldsymbol{y}, \nabla f(\boldsymbol{y})-m\boldsymbol{y}\rangle$$

Apparently, it indicates 2 inequalities, which follows,

$$f(\boldsymbol{y})-\frac{m}{2}\|\boldsymbol{y}\|_2^2 \geq f(\boldsymbol{x})-\frac{m}{2}\|\boldsymbol{x}\|_2^2+\langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})-m\boldsymbol{x}\rangle$$

And the second inequality just swaps x with y. And $g(\boldsymbol{x})$ is convex.

5 [Q5]

5.1 (a)

Acoording to Cauchy-Schwartz inequality, we have,

$$\begin{aligned}\langle \boldsymbol{x}-\boldsymbol{y}, \nabla f(\boldsymbol{x})-\nabla f(\boldsymbol{y})\rangle &\leq \|\boldsymbol{x}-\boldsymbol{y}\|_2 * \|\nabla f(\boldsymbol{x})-\nabla f(\boldsymbol{y})\|_2 \\ &\leq \|\boldsymbol{x}-\boldsymbol{y}\|_2 * M\|\boldsymbol{x}-\boldsymbol{y}\|_2 \\ &\leq M\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\end{aligned}$$

5.2 (b)

$$\langle \boldsymbol{x}-\boldsymbol{y}, \nabla f(\boldsymbol{x})-\nabla f(\boldsymbol{y})\rangle \leq M\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \tag{1}$$

$$\langle \boldsymbol{x}-\boldsymbol{y}, \nabla f(\boldsymbol{x})-\nabla f(\boldsymbol{y})-M\boldsymbol{x}+M\boldsymbol{y}\rangle \leq 0 \tag{2}$$

If we define $g(\boldsymbol{x})=\frac{M}{2}\|\boldsymbol{x}\|_2^2-f(\boldsymbol{x})$, then we can have,

$$\nabla g(\boldsymbol{x})=M\boldsymbol{x}-\nabla f(\boldsymbol{x})$$

And from (2) we can achieve,

$$\begin{aligned}\langle \boldsymbol{x}-\boldsymbol{y}, \nabla g(\boldsymbol{y})-\nabla g(\boldsymbol{x})\rangle &\leq 0 \\ \langle \boldsymbol{y}-\boldsymbol{x}, \nabla g(\boldsymbol{y})-\nabla g(\boldsymbol{x})\rangle &\geq 0\end{aligned}$$

5.3 (c)

Since $g(\boldsymbol{x})$ is convex, we can have,

$$g(\boldsymbol{y}) \geq g(\boldsymbol{x})+\langle \boldsymbol{y}-\boldsymbol{x}, \nabla g(\boldsymbol{x})\rangle \tag{3}$$

$$\frac{M}{2}\|\boldsymbol{y}\|_2^2-f(\boldsymbol{y}) \geq \frac{M}{2}\|\boldsymbol{x}\|_2^2-f(\boldsymbol{x})+\langle \boldsymbol{y}-\boldsymbol{x}, M\boldsymbol{x}-\nabla f(\boldsymbol{x})\rangle \tag{4}$$

$$f(\boldsymbol{y}) \leq \frac{M}{2}\|\boldsymbol{y}\|_2^2-\frac{M}{2}\|\boldsymbol{x}\|_2^2+f(\boldsymbol{x})-\langle \boldsymbol{y}-\boldsymbol{x}, M\boldsymbol{x}\rangle+\langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle \tag{5}$$

$$f(\boldsymbol{y}) \leq \frac{M}{2}\langle \boldsymbol{y}-\boldsymbol{x}, \boldsymbol{y}+\boldsymbol{x}\rangle-M\langle \boldsymbol{y}-\boldsymbol{x}, \boldsymbol{x}\rangle+\langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle+f(\boldsymbol{x}) \tag{6}$$

$$f(\boldsymbol{y}) \leq \frac{M}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2+\langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle+f(\boldsymbol{x}) \tag{7}$$

5.4 (d)

since $f(x)$ is convex, we can have,

$$f(\boldsymbol{x}) + \langle \boldsymbol{z} - \boldsymbol{x}, \nabla f(\boldsymbol{x}) \rangle \leq f(\boldsymbol{z}) \tag{8}$$

Besides, (iv) tells us,

$$f(\boldsymbol{z}) \leq f(\boldsymbol{y}) + \langle \boldsymbol{z} - \boldsymbol{y}, \nabla f(\boldsymbol{y}) \rangle + \frac{M}{2} \|\boldsymbol{z} - \boldsymbol{y}\|_2^2 \tag{9}$$

If we combine the 2 inequalities (8)(9) together, then we can have,

$$f(\boldsymbol{x}) - f(\boldsymbol{y}) \leq \langle \boldsymbol{x} - \boldsymbol{z}, \nabla f(\boldsymbol{x}) \rangle + \langle \boldsymbol{z} - \boldsymbol{y}, \nabla f(\boldsymbol{y}) \rangle + \frac{M}{2} \|\boldsymbol{z} - \boldsymbol{y}\|_2^2 \tag{10}$$

If we take the gradient with respect to $\boldsymbol{z}$, then we can have,

$$\boldsymbol{z} = \frac{\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})}{M} + \boldsymbol{y} \tag{11}$$

Plug this back into (10) expression and we can get an upper bound for $f(\boldsymbol{x}) - f(\boldsymbol{y})$,

$$\begin{aligned}
(*) &\leq \left\langle \boldsymbol{x} - \frac{\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})}{M} - \boldsymbol{y}, \nabla f(\boldsymbol{x}) \right\rangle + \left\langle \frac{\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})}{M} + \boldsymbol{y}, \nabla f(\boldsymbol{y}) \right\rangle \\
&\quad + \| \frac{\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})}{M} \|_2^2 \\
&\leq \langle \boldsymbol{x} - \boldsymbol{y}, \nabla f(\boldsymbol{x}) \rangle - \frac{1}{M} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2^2 + \frac{1}{2M} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2^2 \\
&\leq \langle \boldsymbol{x} - \boldsymbol{y}, \nabla f(\boldsymbol{x}) \rangle - \frac{1}{2M} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2^2
\end{aligned}$$

Hence, we can have,

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{y} - \boldsymbol{x}, \nabla f(\boldsymbol{x}) \rangle + \frac{1}{2M} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2^2$$

5.5 (e)

(v) tells us,

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{y} - \boldsymbol{x}, \nabla f(\boldsymbol{x}) \rangle + \frac{1}{2M} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2^2 \tag{12}$$

Simiarly, we can have,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \langle \boldsymbol{x} - \boldsymbol{y}, \nabla f(\boldsymbol{y}) \rangle + \frac{1}{2M} \|\nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x})\|_2^2 \tag{13}$$

Combine the 2 inequalities (12) (13) together and we can have,

$$\langle \boldsymbol{x} - \boldsymbol{y}, \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}) \rangle \geq \frac{1}{M} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2^2$$

5.6 (f)

$$\begin{aligned}
\langle \boldsymbol{x} - \boldsymbol{y}, \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}) \rangle &\geq \frac{1}{M} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2^2 \\
M \langle \boldsymbol{x} - \boldsymbol{y}, \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}) \rangle &\geq \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2^2 \\
M \|\boldsymbol{x} - \boldsymbol{y}\|_2 &\geq \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2
\end{aligned}$$

6 [Q6]

6.1 (a)

Since we have $\{x_k\}_{k=0}^{\infty}$ which satisfies $|x_{k+1} - x^\star| \leq \beta |x_k - x^\star|$, we can get,

$$\begin{aligned} |x_1 - x^\star| &\leq \beta |x_0 - x^\star| \\ |x_2 - x^\star| &\leq \beta |x_1 - x^\star| \\ &\vdots \\ |x_k - x^\star| &\leq \beta |x_{k-1} - x^\star| \end{aligned}$$

If we combine these inequalities together we can have,

$$|x_k - x^\star| \leq \beta^k |x_0 - x^\star| \tag{14}$$

Besides, we know that,

$$k \geq \frac{\log\left(\frac{|x_0 - x^\star|}{\epsilon}\right)}{\log\left(\frac{1}{\beta}\right)}$$

we can achieve in addition,

$$\frac{1}{\beta^k} \geq \frac{|x_0 - x^\star|}{\epsilon} \tag{15}$$

$$\epsilon \geq \beta^k |x_0 - x^\star| \tag{16}$$

Put (16) back to (14) we can achieve,

$$|x_k - x^\star| \leq \epsilon$$

6.2 (b)

$$\lim_{k\to\infty} \frac{x_{k+1} - x^\star}{x_k - x^\star} = \lim_{k\to\infty} \frac{k}{k+1} \tag{17}$$

$$= \beta = 1 \tag{18}$$

Hence, it's sublinear.
Besides, we can get

$$\begin{aligned} |x_1 - x^\star| &= \frac{1}{2} |x_0 - x^\star| \\ |x_2 - x^\star| &= \frac{2}{3} |x_1 - x^\star| \\ &\vdots \\ |x_k - x^\star| &= \frac{k}{k+1} |x_{k-1} - x^\star| \end{aligned}$$

Combine these inequalities together, we can have,

$$|x_k - x^\star| \leq \frac{1}{k+1}|x_0 - x^\star|$$

To make sure $|x_k - x^\star| \leq \epsilon$, we can get,

$$\frac{1}{k+1}|x_0 - x^\star| \geq \epsilon \tag{19}$$

$$k \leq \frac{|x_0 - x^\star|}{\epsilon} - 1 \tag{20}$$

since k is integer, I think it needs to round up.

$$k \leq \lceil \frac{|x_0 - x^\star|}{\epsilon} - 1 \rceil$$

6.3 (c)

For $q = 1$, since $x_k = \frac{1}{2^k}$, $x^\star$ must be 0 when $k \to \infty$, thus we can have,

$$\lim_{k\to\infty} \frac{|x_{k+1} - x^\star|}{|x_k - x^\star|} = \lim_{k\to\infty} \frac{\frac{1}{2^{k+1}}}{\frac{1}{2^k}} \tag{21}$$

$$= \frac{1}{2} \tag{22}$$

Similarly, for $q = 2$, $z_k = \frac{1}{2^{2^k}}$, $x^\star$ must be 0 when $k \to \infty$, and we can have,

$$\lim_{k\to\infty} \frac{|z_{k+1} - z^\star|}{|z_k - z^\star|^2} = \frac{\frac{1}{2^{2^{k+1}}}}{\frac{1}{\left(2^{2^k}\right)^2}}$$

$$= 1$$

7 [Q7]

7.1 (a)

Since the monotone gradient property of convex functions, the function is monotonly increasing or decreasing, bisection in the optimal solution, the gradient of convex function is 0. Only when x_k is close enought to optimal solution $x^\star$ can the iteration stop. Else, the gradient is either greater than 0 or less than 0, choosing the midpoint can shorten interval. And finally, it's guaranteed that the the optimal solution or quasi-optimal solution will be found.

7.2 (b)

In first step, we have,

$$|x_0 - x^\star| \leq \frac{x_u - x_l}{2}$$

If we use bisection algorithm, we compute and have $x_1, x_{u1}, x_{l1}, x_2, x_{u2}, x_{l2}, \cdots$, then,

$$|x_k - x^\star| \leq \frac{x_{uk} - x_{lk}}{2}$$

Since the interval is halved at each step, we have

$$x_{uk} - x_{lk} = \frac{x_{u(k-1)} - x_{l(k-1)}}{2} = \cdots = \frac{x_u - x_l}{2^{k+1}}$$

Hence, we achieve

$$|x_k - x^\star| \leq \frac{x_u - x_l}{2^{k+1}}$$

Thus, the conclusion here is the 'maximum error' possible between x_k and $x^\star$ decreases as k increases, which means x_k converges to $x^\star$.

7.3 (c)

Apparently it converges linearly. The convergence rate β is

$$\beta = \lim_{k\to\infty} \frac{|x_k - x^\star|}{|x_{k-1} - x^\star|} \tag{23}$$

$$= \lim_{k\to\infty} \frac{\frac{x_{uk} - x_{lk}}{2}}{\frac{x_{u(k-1)} - x_{l(k-1)}}{2}} \tag{24}$$

$$= \lim_{k\to\infty} \frac{x_{uk} - x_{lk}}{x_{u(k-1)} - x_{l(k-1)}} \tag{25}$$

Since the interval is halved at each step, we have

$$\beta = \lim_{k\to\infty} \frac{x_{uk} - x_{lk}}{x_{u(k-1)} - x_{l(k-1)}} = \frac{1}{2}$$

7.4 (d)

First we need to find the bound. Note that in convexity, any local minimum is also a global minimum. Hence we can find the bound x_l, x_r which satisfies $f'(x_l) > 0$ and $f'(x_r) < 0$. Hence, we can make sure the bound is $[x_l, x_r]$. Then we can treat it as an constrained optimziation problem and use the method above to solve it.

8 [Q8]

8.1 (a)

if $g(\boldsymbol{a}^T\boldsymbol{x} + b) \geq \frac{1}{2}$, we can have

$$\frac{1}{1 + e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}} \geq \frac{1}{2} \tag{26}$$

$$e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)} \leq 1 \tag{27}$$

$$\boldsymbol{a}^T\boldsymbol{x} + b \geq 0 \tag{28}$$

And vice versa, namely we can deduce $g(\boldsymbol{a}^T\boldsymbol{x}+b) \geq \frac{1}{2}$ from $\boldsymbol{a}^T\boldsymbol{x}+b \geq 0$. Thus, we can conclude,

$$g(\boldsymbol{a}^T\boldsymbol{x}+b) \geq \frac{1}{2} \iff \boldsymbol{a}^T\boldsymbol{x}+b \geq 0$$

Similarly, we can have

$$g(\boldsymbol{a}^T\boldsymbol{x}+b) \leq \frac{1}{2} \iff \boldsymbol{a}^T\boldsymbol{x}+b \leq 0$$

In conclusion, $h(x)$ can equivalently be written as

$$h(x) = \begin{cases} 1 & \text{if } \boldsymbol{a}^T\boldsymbol{x}+b \geq 0 \\ 0 & \text{if } \boldsymbol{a}^T\boldsymbol{x}+b < 0 \end{cases}$$

8.2 (b)

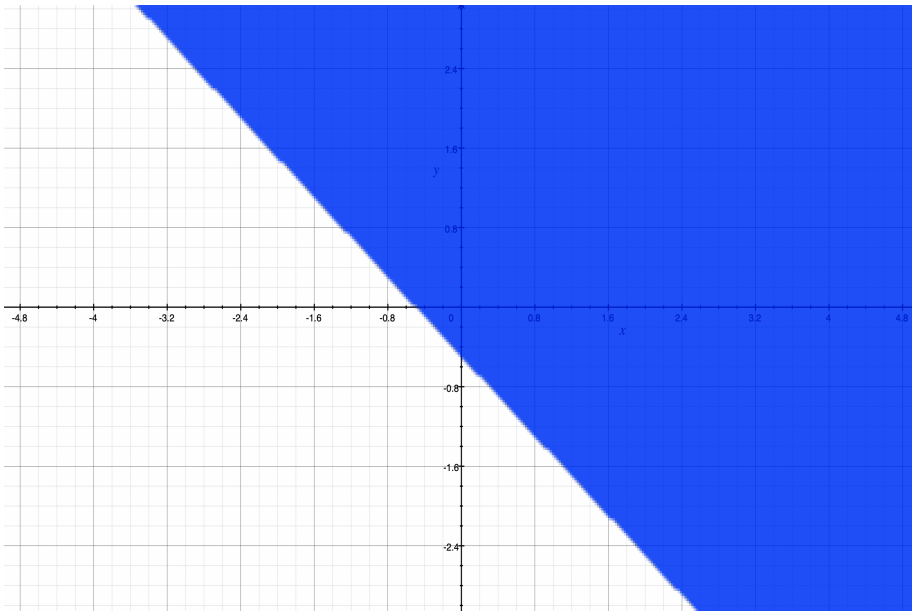

Figure 1: Illustration of Set

Hence, **the blue region is where the set lies**

8.3 (c)

$$\sum_{n=1}^{N} y_n \log(g(\boldsymbol{a}^T\boldsymbol{x}+b)) + (1-y_n)\log(1-g(\boldsymbol{a}^T\boldsymbol{x}+b)) \tag{29}$$

$$= \sum_{n=1}^{N} y_n \log\left(\frac{1}{1+e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}}\right) + (1-y_n)\log\left(1-\frac{1}{1+e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}}\right) \tag{30}$$

$$= \sum_{n=1}^{N} -y_n \log\left(1+e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}\right) + (1-y_n)\log\left(\frac{e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}}{1+e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}}\right) \tag{31}$$

$$= \sum_{n=1}^{N} -y_n \log\left(1+e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}\right) + \log\left(\frac{e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}}{1+e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}}\right) - y_n \log\left(\frac{e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}}{1+e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}}\right) \tag{32}$$

$$= \sum_{n=1}^{N} -y_n \log\left[\left(1+e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}\right)\left(\frac{e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}}{1+e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}}\right)\right] + \log\left(\frac{1}{1+e^{\boldsymbol{a}^T\boldsymbol{x}+b}}\right) \tag{33}$$

$$= \sum_{n=1}^{N} -y_n \log\left(e^{-(\boldsymbol{a}^T\boldsymbol{x}+b)}\right) - \log\left(1+e^{\boldsymbol{a}^T\boldsymbol{x}+b}\right) \tag{34}$$

$$= \sum_{n=1}^{N} y_n \left(\boldsymbol{a}^T\boldsymbol{x}+b)\right) - \log\left(1+e^{\boldsymbol{a}^T\boldsymbol{x}+b}\right) \tag{35}$$

Hence, the origin problem is equivalent to

$$\underset{\boldsymbol{a},b}{\text{maximize}} \sum_{n=1}^{N} y_n \left(\boldsymbol{a}^T\boldsymbol{x}+b)\right) - \log\left(1+e^{\boldsymbol{a}^T\boldsymbol{x}+b}\right)$$

8.4 (d)

$$\nabla f(\boldsymbol{\theta}) = \sum_{n=1}^{N}\left(y_n\tilde{\boldsymbol{x}}_n - \frac{e^{\boldsymbol{\theta}^T\tilde{\boldsymbol{x}}_n}}{1+e^{\boldsymbol{\theta}^T\tilde{\boldsymbol{x}}_n}}\tilde{\boldsymbol{x}}_n\right) \tag{36}$$

$$= \sum_{n=1}^{N}\tilde{\boldsymbol{x}}_n\left(y_n - \frac{e^{\boldsymbol{\theta}^T\tilde{\boldsymbol{x}}_n}}{1+e^{\boldsymbol{\theta}^T\tilde{\boldsymbol{x}}_n}}\right) \tag{37}$$

$$\nabla^2 f(\boldsymbol{\theta}) = -\sum_{n=1}^{N}\frac{e^{-\boldsymbol{\theta}^T\tilde{\boldsymbol{x}}_n}}{(1+e^{-\boldsymbol{\theta}^T\tilde{\boldsymbol{x}}_n})^2}\tilde{\boldsymbol{x}}_n\tilde{\boldsymbol{x}}_n^T$$

8.5 (e)

It's M-smooth for some finite value of M and strongly convex.

8.6 (f)

Here, I set the stopping criteria as the maximum value of θ_k **is less than 0.01**.
In details, I code it in python as below. The first part is to load data.

```
import numpy as np
from sklearn import datasets
import math

np.random.seed(2020) # Set random seed so results are repeatable
x,y = datasets.make_blobs(n_samples=100,n_features=2,centers=2,
                                        cluster_std=6.0)
```

The second part is to define a function in gradient so that i can make my life easier.

```
# define logit function
def logit(x):
    return 1./(1+np.exp(-x))
```

The third part is to iterativly find the optimal value

```
# add 1 for every row of x
x_tilde = np.ones([100,3])
for i in range(len(x)):
    x_tilde[i] = np.r_[x[i], np.ones(1)]

X_train = np.matrix(x_tilde) #100*3
y_train = np.matrix(y).T #100*1

# initialize theta 3*1
theta = np.zeros(X_train.shape[1])
theta = np.matrix(theta).T

k=0
dk = np.ones((3,1))
alpha = 0.001

# iteration number
while max(abs(dk)) > 0.01:
    error = y_train - logit(X_train * theta) #100*1
    dk = X_train.T * error #3*1
    theta += alpha * dk
    k = k + 1
    print("current dk is {} and iteration is {}".format(dk.T, k))
print("theta is {}".format(theta.T))
```

I found in this condition, the it breaks the loop in **2695**. As is shown in the picture,

```
current dk is [[-0.00089327 -0.00122307  0.01033231]] and iteration is 2678
current dk is [[-0.0008915  -0.00122065  0.01031184]] and iteration is 2679
current dk is [[-0.00088973 -0.00121824  0.0102914 ]] and iteration is 2680
current dk is [[-0.00088797 -0.00121583  0.01027101]] and iteration is 2681
current dk is [[-0.00088621 -0.00121342  0.01025066]] and iteration is 2682
current dk is [[-0.00088446 -0.00121102  0.01023034]] and iteration is 2683
current dk is [[-0.0008827  -0.00120862  0.01021007]] and iteration is 2684
current dk is [[-0.00088096 -0.00120623  0.01018984]] and iteration is 2685
current dk is [[-0.00087921 -0.00120385  0.01016965]] and iteration is 2686
current dk is [[-0.00087747 -0.00120146  0.0101495 ]] and iteration is 2687
current dk is [[-0.00087573 -0.00119909  0.01012939]] and iteration is 2688
current dk is [[-0.000874   -0.00119671  0.01010932]] and iteration is 2689
current dk is [[-0.00087227 -0.00119435  0.01008928]] and iteration is 2690
current dk is [[-0.00087054 -0.00119198  0.01006929]] and iteration is 2691
current dk is [[-0.00086881 -0.00118962  0.01004934]] and iteration is 2692
current dk is [[-0.00086709 -0.00118727  0.01002943]] and iteration is 2693
current dk is [[-0.00086538 -0.00118492  0.01000956]] and iteration is 2694
current dk is [[-0.00086366 -0.00118258  0.00998973]] and iteration is 2695
theta is [[-0.28047652 -0.45702889  2.20888147]]
```

Figure 2: Illustration of Code

And the final optimal d_k, namely gradient of $f(\theta)$ is [[-0.00086366 -0.00118258 0.00998973]] and corresponding iteration is 2695. And the final θ is [[-0.28047652 -0.45702889 2.20888147]]

8.7 (g)

Here, in outer loop I set the stopping criteria as the maximum absolute value of θ_k **is less than 0.01**.

In inner loop, I set the stopping criteria as $\alpha_l - \alpha_r > 1e - 10$. In details, I code it in python as below. The first and second part is the same as before.

```
import numpy as np
from sklearn import datasets
import math

np.random.seed(2020) # Set random seed so results are repeatable
x,y = datasets.make_blobs(n_samples=100,n_features=2,centers=2,
                                        cluster_std=6.0)

# define logit function
def logit(x):
    return 1./(1+np.exp(-x))
```

And the third part is to iterativly find the optimal value. There are 2 loops. The Inner loop is to find optimal α in each iteration while the outer loop is to find the optimal solution.

```
# add 1 for every row of x
x_tilde = np.ones([100,3])
for i in range(len(x)):
    x_tilde[i] = np.r_[x[i], np.ones(1)]

X_train = np.matrix(x_tilde) #100*3
y_train = np.matrix(y).T #100*1

# initialize theta 3*1
theta = np.zeros(X_train.shape[1])
theta = np.matrix(theta).T

k=0
dk = np.ones((3,1))

alpha_left = 0
alpha_right = 1
# iteration number
while max(abs(dk)) > 0.01:

    m = 0
    alpha_left = 0
    alpha_right = 1
    h_prime = 1
    while abs(alpha_left - alpha_right) > 1e-10:
        alpha = (alpha_left + alpha_right) / 2
        h_prime = (X_train.T * (- y_train + logit(X_train * (theta +
                                            alpha * dk)))).T * dk
        if h_prime > 0:
            alpha_right = alpha
        else:
            alpha_left = alpha
        m += 1
        print("optimal alpha is {} and iteration is {}".format(alpha,
                                                m))

    error = y_train - logit(X_train * theta) #100*1
    dk = X_train.T * error #3*1
    theta += alpha * dk
```

```
        k = k + 1
        print("current dk is {} and iteration is {}".format(dk.T, k))
    print("theta is {} and total iteration is {}".format(theta.T, m*k))
```

The final results are shown in picture below,

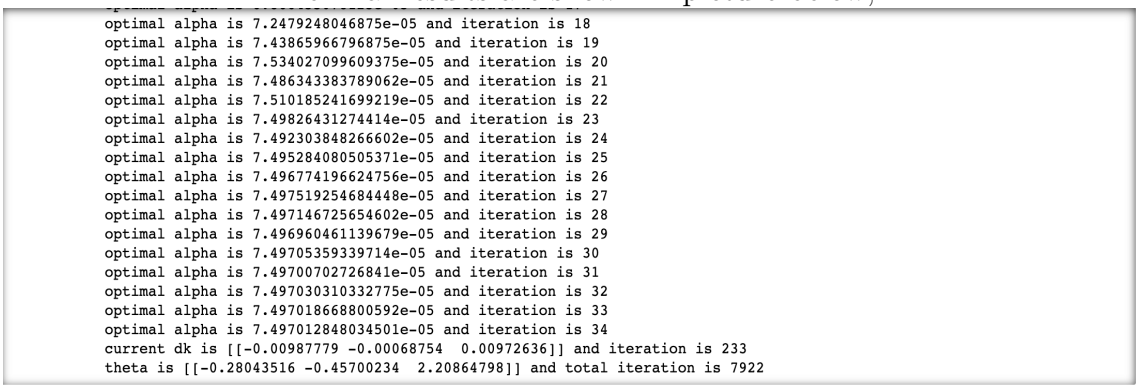

Figure 3: Illustration of Code

When using this strategy to find the "optimal" step size at each iteration, 34 iterations are needed. And total number of iterations taken by the combined bisection searches is 7922. And θ is [[-0.28043516 -0.45700234 2.20864798]]. I don't think it's worth the additional costs since it has more iterations.