CS 726 - Fall 2020

Homework #2

Due : 10/05/2020, 5pm

Zijie Zhang

October 3, 2020

Question 1

Proof. $\Rightarrow$, Let

$$\varphi(\alpha) = \frac{1}{\alpha}\left(f\left((1-\alpha)x + \alpha y\right) - f(x)\right)$$

f is m-strongly convex means,

$$f\left((1-\alpha)x + \alpha y\right) \leqslant (1-\alpha)f(x) + \alpha f(y) - \frac{m}{2}\alpha(1-\alpha)||y-x||^2$$

$$f\left((1-\alpha)x + \alpha y\right) - f(x) \leqslant \alpha\left(f(y) - f(x)\right) - \frac{m}{2}\alpha(1-\alpha)||y-x||^2$$

$$f(y) - f(x) \geqslant \varphi(\alpha) + \frac{m}{2}(1-\alpha)||y-x||^2$$

Let $\alpha \to 0$, we have

$$f(y) \geqslant f(x) + \varphi^{'}(0) + \frac{m}{2}||y-x||^2 = f(x) + \langle \nabla f(x), y-x\rangle + \frac{m}{2}||y-x||^2$$

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

Consider, Taylor Theorem:

$$f(x + \alpha(y-x)) = f(x) + \langle \nabla f(x), \alpha(y-x)\rangle + \frac{\alpha^2}{2}(y-x)^T\nabla^2 f(x + \gamma\alpha(y-x))(y-x)$$

Combine the above two formulas, it gives

$$(y-x)^T\nabla^2 f(x)(y-x) \geqslant m||y-x||^2$$

Thus, we have

$$\nabla^2 f(x) \succeq mI$$

$\Leftarrow$, By Taylor Theorem,

$$f(y) = f(x) + \langle \nabla f(x), y-x\rangle + \frac{1}{2}\nabla^2 f(x + \gamma(y-x))||y-x||^2$$

$\nabla^2 f(x) \succeq mI$ means the smallest eigenvalue of $\nabla^2 f(x)$ is greater than m, therefore

$$\frac{1}{2}\nabla^2 f(x + \gamma(y-x))||y-x||^2 \geqslant \frac{1}{2}m||y-x||^2$$

That is

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

Consider

$$(1-\alpha)f(x) + \alpha f(y) - f((1-\alpha)x + \alpha y)$$

We will have

$$(1-\alpha)f(x) + \alpha f(y) - f((1-\alpha)x + \alpha y) \geqslant \frac{m}{2}||y-x||^2(\alpha - \alpha^2) = \frac{m}{2}\alpha(1-\alpha)||y-x||^2$$

$\square$

Question 2

Proof. Let $x_{k+1} = x_k + \nabla f(x_k)$, we have

$$f(x_{k+1}) - f(x_k) \geqslant \langle \nabla f(x_k), \nabla f(x_k) \rangle + \frac{m}{2} ||\nabla f(x_k)||^2$$

Add them together,

$$f(x_{k+1}) - f(x_0) \geqslant \left(1 + \frac{m}{2}\right) \sum_{i=0}^{k} ||\nabla f(x_i)||^2 \geqslant \left(1 + \frac{m}{2}\right) \left\| \sum_{i=0}^{k} \nabla f(x_i) \right\|^2 = \left(1 + \frac{m}{2}\right) \|x_{k+1} - x_0\|^2$$

The gradient of f can go to ∞, when $\|x_{k+1} - x_0\|$ is large enough.
So, f cannot be Lipschitz continuous on the entire $\mathbb{R}^d$. But it is possible on the unit Euclidean ball. □

Question 3

Proof. By Lemma2.2

$$\begin{aligned} f(x_{k+1}) &\leqslant f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2} \|x_{k+1} - x_k\|_2^2 \\ &= f(x_k) - \alpha_k \langle \nabla f(x_k), \nabla_{i_k} f(x_k) e_{i_k} \rangle + \frac{L}{2} \alpha_k^2 \|\nabla_{i_k} f(x_k) e_{i_k}\|_2^2 \\ &= f(x_k) + \left(\frac{L}{2} \alpha_k - 1 \right) \alpha_k \|\nabla_{i_k} f(x_k) e_{i_k}\|_2^2 \end{aligned}$$

Choose $\alpha_k = \frac{1 + \sqrt{1 - L\beta d^2}}{L}$, then

$$\mathbb{E}[f(x_{k+1}) - f(x_k)] = -\frac{\beta d^2}{2} \mathbb{E}[\|\nabla_{i_k} f(x_k) e_{i_k}\|_2^2] = -\frac{\beta}{2} \|\nabla f(x_k)\|_2^2$$

□

Question 4

Proof.

(i)

$$\nabla \psi(y) = y - x_0$$

$$\begin{aligned} D_\psi(x, y) &= \frac{1}{2} \|x - x_0\|_2^2 - \frac{1}{2} \|y - x_0\|_2^2 - \langle \nabla \psi(y), x - y \rangle \\ &= \frac{1}{2} \|x\|_2^2 + \frac{1}{2} \|y\|_2^2 - \langle x, y \rangle \\ &= \frac{1}{2} \|x - y\|_2^2 \end{aligned}$$

(ii)

$$\nabla \phi(y) = \psi(y) + x_0 = y$$

$$\begin{aligned} D_\phi(x, y) &= \phi(x) - \phi(y) - \langle y, x - y \rangle \\ &= \psi(x) - \psi(y) + \langle x_0 - y, x - y \rangle \\ &= \frac{1}{2} \|x\|_2^2 + \frac{1}{2} \|y\|_2^2 - \langle x, y \rangle \\ &= D_\psi(x, y) \end{aligned}$$

(iii) Left: $D_\psi(x,y) = \frac{1}{2}\|x\|_2^2 + \frac{1}{2}\|y\|_2^2 - \langle x, y\rangle$
Right:

$$
\begin{aligned}
&D_\psi(z,y) + \langle \nabla\psi(z) - \nabla\psi(y), x - z\rangle + D_\psi(x,z) \\
&= \frac{1}{2}\|z\|_2^2 + \frac{1}{2}\|y\|_2^2 - \langle z, y\rangle + \langle z - y, x - z\rangle + \frac{1}{2}\|x\|_2^2 + \frac{1}{2}\|z\|_2^2 - \langle x, z\rangle \\
&= \frac{1}{2}\|x\|_2^2 + \frac{1}{2}\|y\|_2^2 - \langle x, y\rangle
\end{aligned}
$$

Left=Right

(iv) Obviously, $\nabla m_k(v_k) = 0$, thus

$$
\begin{aligned}
D_{m_k}(x, v_k) &= m_k(x) - m_k(v_k) \\
&= \sum_{i=0}^{k} a_i D_{\psi_i}(x, v_k) \\
&= \sum_{i=0}^{k} a_i \left(\frac{1}{2}\|x\|_2^2 + \frac{1}{2}\|v_k\|_2^2 - \langle x, v_k\rangle\right) \\
&= \sum_{i=0}^{k} a_i \frac{1}{2}\|x - v_k\|_2^2 \\
&= \frac{A_k}{2}\|x - v_k\|_2^2
\end{aligned}
$$

So, we have proved

$$
m_{k+1}(x) = m_k(v_k) + a_{k+1}\psi_{k+1}(x) + \frac{A_k}{2}\|x - v_k\|_2^2
$$

□

Question 5

Proof. □