

SI251 - Convex Optimization, 2024 Spring  
Homework 3

Name: **Zhou Shouchen**

Student ID: 2021533042

Due 23:59 (CST), Apr. 24, 2024

1. **(50 pts) L-smooth functions.** Suppose the function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is convex and differentiable. Please prove that the following relations holds for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  if  $f$  with an  $L$ -Lipschitz continuous conditions,

$$[1] \Rightarrow [2] \Rightarrow [3]$$

$$[1] \quad \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \leq L \|\mathbf{x} - \mathbf{y}\|^2,$$

$$[2] \quad f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|^2,$$

$$[3] \quad f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{1}{2L} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|^2, \forall \mathbf{x}, \mathbf{y}.$$

**Solution:**

1.  $[1] \Rightarrow [2]$ .

Define

$$g(t) = f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})), t \in [0, 1]$$

So we have  $g(0) = f(\mathbf{x})$  and  $g(1) = f(\mathbf{y})$ , and  $\nabla g(t) = \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))^T (\mathbf{y} - \mathbf{x})$ .

So we have:

$$\begin{aligned} f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) &= g(1) - g(0) - \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) \\ &= \left( \int_0^1 \nabla g(t) dt \right) - \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) \quad (\text{Newton-Leibniz formula}) \\ &= \left( \int_0^1 \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))^T (\mathbf{y} - \mathbf{x}) dt \right) - \left( \int_0^1 \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) dt \right) \\ &= \int_0^1 (\nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}))^T (\mathbf{y} - \mathbf{x}) dt \\ &= \int_0^1 \frac{1}{t} \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), [\mathbf{x} + t(\mathbf{y} - \mathbf{x})] - \mathbf{x} \rangle dt \\ &\leq \int_0^1 \frac{1}{t} L \|\mathbf{x} + t(\mathbf{y} - \mathbf{x}) - \mathbf{x}\|^2 dt \quad ([1] \text{ as the condition}) \\ &= \int_0^1 t L \|\mathbf{y} - \mathbf{x}\|^2 dt \\ &= L \|\mathbf{y} - \mathbf{x}\|^2 \cdot \int_0^1 t dt \\ &= L \|\mathbf{y} - \mathbf{x}\|^2 \cdot \frac{1}{2} \\ &= \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|^2 \end{aligned}$$

So we have proved

$$f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) \leq \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|^2$$

i.e.

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|^2$$

So we have proved that  $[1] \Rightarrow [2]$ .

2.  $[2] \Rightarrow [3]$  :

From [2], we could get that:

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|^2$$

In order to get the form of [3], we can set  $\mathbf{z} = \mathbf{x} + \frac{1}{L}(\nabla f(\mathbf{y}) - \nabla f(\mathbf{x}))$ .

So we have

$$\begin{aligned} f(\mathbf{z}) &\leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{z} - \mathbf{x}) + \frac{L}{2} \|\mathbf{z} - \mathbf{x}\|^2 \\ &= f(\mathbf{x}) + \frac{1}{L} \nabla f(\mathbf{x})^T (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})) + \frac{L}{2} \left\| \frac{1}{L} (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})) \right\|^2 \\ &= f(\mathbf{x}) + \frac{1}{L} \nabla f(\mathbf{x})^T (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})) + \frac{1}{2L} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|^2 \end{aligned}$$

But the direction of inequality's sign is different from what we want, so we need to use the first order's term to construct a negative one, i.e.

$$\begin{aligned} f(\mathbf{z}) &\leq f(\mathbf{x}) + \frac{1}{L} \nabla f(\mathbf{x})^T (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})) + \frac{1}{2L} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|^2 \\ f(\mathbf{z}) - f(\mathbf{x}) &\leq -\frac{1}{L} (-\nabla f(\mathbf{x}))^T (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})) + \frac{1}{2L} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|^2 \\ f(\mathbf{z}) - f(\mathbf{x}) - \frac{1}{L} \nabla f(\mathbf{x})^T (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})) &\leq -\frac{1}{L} (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x}))^T (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})) + \frac{1}{2L} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|^2 \\ f(\mathbf{z}) - f(\mathbf{x}) - \frac{1}{L} \nabla f(\mathbf{x})^T (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})) &\leq -\frac{1}{2L} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|^2 \\ \frac{1}{2L} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|^2 &\leq f(\mathbf{x}) - f(\mathbf{z}) + \frac{1}{L} \nabla f(\mathbf{x})^T (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})) \\ \frac{1}{2L} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|^2 &\leq f(\mathbf{x}) - f(\mathbf{z}) + \nabla f(\mathbf{y})^T (\mathbf{z} - \mathbf{x}) \quad \left( \mathbf{z} = \mathbf{x} + \frac{1}{L} (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})) \right) \end{aligned}$$

Since  $f(x)$  is convex, so we can get the first order Taylor expansion of  $f(\mathbf{z})$  at  $\mathbf{y}$ :

$$f(\mathbf{z}) \geq f(\mathbf{y}) + \nabla f(\mathbf{y})^T (\mathbf{z} - \mathbf{y})$$

i.e.

$$\nabla f(\mathbf{y})^T \mathbf{z} \leq f(\mathbf{z}) - f(\mathbf{y}) + \nabla f(\mathbf{y})^T \mathbf{y}$$

And put it into the above inequality, we can get:

$$\begin{aligned} \frac{1}{2L} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|^2 &\leq f(\mathbf{x}) - f(\mathbf{z}) + \nabla f(\mathbf{y})^T (\mathbf{z} - \mathbf{x}) \\ &= f(\mathbf{x}) - f(\mathbf{z}) + (f(\mathbf{z}) - f(\mathbf{y}) + \nabla f(\mathbf{y})^T \mathbf{y}) - f(\mathbf{y})^T \mathbf{x} \\ &= f(\mathbf{x}) - f(\mathbf{y}) + \nabla f(\mathbf{y})^T (\mathbf{y} - \mathbf{x}) \end{aligned}$$

i.e.

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{1}{2L} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|^2$$

Swap  $\mathbf{x}$  and  $\mathbf{y}$ , we can get:

$$f(\mathbf{x}) \geq f(\mathbf{y}) + \nabla f(\mathbf{y})^T (\mathbf{x} - \mathbf{y}) + \frac{1}{2L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2$$

So we have proved that  $[2] \Rightarrow [3]$ .

So above all, we have proved that  $[1] \Rightarrow [2] \Rightarrow [3]$ .

2. **(50 pts) Backtracking line search.** Please show the convergence of backtracking line search on a  $m$ -strongly convex and  $M$ -smooth objective function  $f$  as

$$f(\mathbf{x}^{(k)}) - p^* \leq c^k \left( f(\mathbf{x}^{(0)}) - p^* \right)$$

where  $c = 1 - \min\{2m\alpha, 2\beta\alpha m/M\} < 1$ .

---

**Algorithm 9.2** *Backtracking line search.*

**given** a descent direction  $\Delta x$  for  $f$  at  $x \in \text{dom } f$ ,  $\alpha \in (0, 0.5)$ ,  $\beta \in (0, 1)$ .

$t := 1$ .

**while**  $f(x + t\Delta x) > f(x) + \alpha t \nabla f(x)^T \Delta x$ ,  $t := \beta t$ .

---

**Solution:**

Since  $f$  is convex and  $M$ -smooth, so we could expand  $f(\mathbf{x} + t\Delta\mathbf{x})$  at  $\mathbf{x}$ :

$$\begin{aligned} f(\mathbf{x} + t\Delta\mathbf{x}) &\leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (t\Delta\mathbf{x}) + \frac{M}{2} \|t\Delta\mathbf{x}\|^2 \\ &= f(\mathbf{x}) + t \nabla f(\mathbf{x})^T \Delta\mathbf{x} + \frac{M}{2} t^2 \|\Delta\mathbf{x}\|^2 \end{aligned}$$

Since the backtracking line search would terminate when

$$f(\mathbf{x} + t\Delta\mathbf{x}) \leq f(\mathbf{x}) + \alpha t \nabla f(\mathbf{x})^T \Delta\mathbf{x}$$

If we let the  $f(\mathbf{x} + t\Delta\mathbf{x})$ 's upper bound  $f(\mathbf{x}) + t \nabla f(\mathbf{x})^T \Delta\mathbf{x} + \frac{M}{2} t^2 \|\Delta\mathbf{x}\|^2$  to be less than  $f(\mathbf{x}) + \alpha t \nabla f(\mathbf{x})^T \Delta\mathbf{x}$ , then it must have been terminated.

So we have:

$$\begin{aligned} f(\mathbf{x}) + t \nabla f(\mathbf{x})^T \Delta\mathbf{x} + \frac{M}{2} t^2 \|\Delta\mathbf{x}\|^2 &\leq f(\mathbf{x}) + \alpha t \nabla f(\mathbf{x})^T \Delta\mathbf{x} \\ \frac{M}{2} t^2 \|\Delta\mathbf{x}\|^2 &\leq (\alpha - 1) t \nabla f(\mathbf{x})^T \Delta\mathbf{x} \\ \frac{M}{2} t \|\Delta\mathbf{x}\|^2 &\leq (\alpha - 1) \nabla f(\mathbf{x})^T \Delta\mathbf{x} \quad (\text{Since } t > 0) \\ t &\leq \frac{2(\alpha - 1) \nabla f(\mathbf{x})^T \Delta\mathbf{x}}{M \|\Delta\mathbf{x}\|^2} \end{aligned}$$

Since  $\alpha \in (0, 0.5)$ , so  $(\alpha - 1) < 0$ , and since  $\Delta\mathbf{x}$  is the descent direction, so  $\nabla f(\mathbf{x})^T \Delta\mathbf{x} < 0$ , so  $t$  is less than a positive number.

Since  $t$  is generated by the backtracking line search, so we have  $t \leq \frac{\beta \cdot 2(\alpha - 1) \nabla f(\mathbf{x})^T \Delta\mathbf{x}}{M \|\Delta\mathbf{x}\|^2}$ . Also, the initial step is  $t = 1$ , so combine the above inequality, we have:

$$t = \min \left\{ 1, \frac{\beta \cdot 2(\alpha - 1) \nabla f(\mathbf{x})^T \Delta\mathbf{x}}{M \|\Delta\mathbf{x}\|^2} \right\}$$

And for this problem, we take the descent direction as  $\Delta \mathbf{x} = -\nabla f(\mathbf{x})$ , so we have:

$$t = \min \left\{ 1, \frac{\beta \cdot 2(\alpha - 1) \nabla f(\mathbf{x})^T (-\nabla f(\mathbf{x}))}{M \|\nabla f(\mathbf{x})\|^2} \right\} = \min \left\{ 1, \frac{2\beta(1 - \alpha)}{M} \right\}$$

Since  $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t\Delta \mathbf{x}$ , we have

$$\begin{aligned} f(\mathbf{x}^{(k+1)}) &= f(\mathbf{x}^{(k)} + t\Delta \mathbf{x}^{(k)}) \\ &= f(\mathbf{x}^{(k)} - t\nabla f(\mathbf{x}^{(k)})) \\ &\leq f(\mathbf{x}^{(k)}) + \alpha t \nabla f(\mathbf{x}^{(k)})^T \Delta \mathbf{x}^{(k)} \\ &= f(\mathbf{x}^{(k)}) - \alpha t \|\nabla f(\mathbf{x}^{(k)})\|^2 \\ &= f(\mathbf{x}^{(k)}) - \alpha \cdot \min \left\{ 1, \frac{2\beta(1 - \alpha)}{M} \right\} \|\nabla f(\mathbf{x}^{(k)})\|^2 \end{aligned}$$

To continuously prove, we need to introduce the Lemma:  $f(\mathbf{x})$  is  $m$ -strongly convex, then

$$p^* \geq f(\mathbf{x}) - \frac{1}{2m} \|\nabla f(\mathbf{x})\|^2$$

proof: Since  $f$  is  $m$ -strongly convex, so we have  $\forall \mathbf{x}, \mathbf{y}$ :

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{m}{2} \|\mathbf{y} - \mathbf{x}\|^2$$

Let

$$g(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{m}{2} \|\mathbf{y} - \mathbf{x}\|^2$$

So  $\nabla g(\mathbf{y}) = \nabla f(\mathbf{x}) + m(\mathbf{y} - \mathbf{x})$ , and  $\nabla^2 g(\mathbf{y}) = mI \succeq 0$ , so

$$g(\mathbf{y})_{\min} = g\left(\mathbf{x} - \frac{1}{m} \nabla f(\mathbf{x})\right) = f(\mathbf{x}) - \frac{1}{2m} \|\nabla f(\mathbf{x})\|^2$$

So

$$f(\mathbf{y}) \geq f(\mathbf{x}) - \frac{1}{2m} \|\nabla f(\mathbf{x})\|^2$$

Change  $\mathbf{y}$  to  $p^*$ , we have  $p^* \geq f(\mathbf{x}) - \frac{1}{2m} \|\nabla f(\mathbf{x})\|^2$ .

So we have proved the Lemma.

So we get that

$$\|\nabla f(\mathbf{x})\|^2 \geq (2m) (f(\mathbf{x}) - p^*)$$

With the Lemma, we can get that

$$\begin{aligned} f(\mathbf{x}^{(k+1)}) &\leq f(\mathbf{x}^{(k)}) - \alpha \cdot \min \left\{ 1, \frac{2\beta(1 - \alpha)}{M} \right\} \|\nabla f(\mathbf{x}^{(k)})\|^2 \\ f(\mathbf{x}^{(k+1)}) - p^* &\leq f(\mathbf{x}^{(k)}) - p^* - \alpha \cdot \min \left\{ 1, \frac{2\beta(1 - \alpha)}{M} \right\} \|\nabla f(\mathbf{x}^{(k)})\|^2 \\ f(\mathbf{x}^{(k+1)}) - p^* &\leq f(\mathbf{x}^{(k)}) - p^* - \alpha \cdot \min \left\{ 1, \frac{2\beta(1 - \alpha)}{M} \right\} (2m) (f(\mathbf{x}^{(k)}) - p^*) \quad (\text{Lemma}) \end{aligned}$$