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 $f : \mathbb{R}^n \to \mathbb{R}$ 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,

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

$$\begin{aligned}
&[1]\ \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \le L\|\mathbf{x} - \mathbf{y}\|^2, \\
&[2]\ f(\mathbf{y}) \le f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|^2, \\
&[3]\ f(\mathbf{y}) \ge f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{1}{2L}\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|^2, \forall \mathbf{x}, \mathbf{y}.
\end{aligned}$$

Solution:

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

   Define

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

   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:

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

   So we have proved

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

   i.e.

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

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

2. $[2] \Rightarrow [3]$:
   From [2], we could get that:

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

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

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

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

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

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

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

   i.e.

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

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

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

   i.e.

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

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

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

   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

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

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

Algorithm 9.2 *Backtracking line search.*

given a descent direction Δx for f at $x \in \textbf{dom}\, f$, $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$.
$t := 1$.
while $f(x + t\Delta x) > f(x) + \alpha t \nabla f(x)^T \Delta x, \quad t := \beta t.$

Solution:

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

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

Since the backtracking line search would terminate when

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

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:

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

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 $t \le \frac{\beta \cdot 2(\alpha - 1)\nabla f(\mathbf{x})^T\Delta\mathbf{x}}{M\|\Delta\mathbf{x}\|^2}$.
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 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t\Delta\mathbf{x}^{(k)}$, we have

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

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

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

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

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

Change $f(\mathbf{y})$ 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

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

With the Lemma, we can get that

$$\begin{aligned}
f\left(\mathbf{x}^{(k+1)}\right) &\leq f\left(\mathbf{x}^{(k)}\right) - \alpha \cdot \min\left\{1, \frac{2\beta(1-\alpha)}{M}\right\} \left\|\nabla f\left(\mathbf{x}^{(\mathbf{k})}\right)\right\|^2 \\
f\left(\mathbf{x}^{(k+1)}\right) - p^* &\leq f\left(\mathbf{x}^{(k)}\right) - p^* - \alpha \cdot \min\left\{1, \frac{2\beta(1-\alpha)}{M}\right\} \left\|\nabla f\left(\mathbf{x}^{(\mathbf{k})}\right)\right\|^2 \\
&\leq f\left(\mathbf{x}^{(k)}\right) - p^* - \alpha \cdot \min\left\{1, \frac{2\beta(1-\alpha)}{M}\right\} (2m)\left(f\left(\mathbf{x}^{(\mathbf{k})}\right) - p^*\right) \qquad \text{(Lemma)} \\
&\leq f\left(\mathbf{x}^{(k)}\right) - p^* - 2m\alpha \cdot \min\left\{1, \frac{\beta}{M}\right\}\left(f\left(\mathbf{x}^{(\mathbf{k})}\right) - p^*\right) \qquad (\alpha \in (0, 0.5) \Rightarrow 2(1-\alpha) \geq 1) \\
&\leq \left(1 - 2m\alpha \cdot \min\left\{1, \frac{\beta}{M}\right\}\right)\left(f\left(\mathbf{x}^{(k)}\right) - p^*\right) \\
&= \left(1 - \min\left\{2m\alpha, \frac{2\beta\alpha m}{M}\right\}\right)\left(f\left(\mathbf{x}^{(k)}\right) - p^*\right)
\end{aligned}$$

Let $c = 1 - \min\left\{2m\alpha, \frac{2\beta\alpha m}{M}\right\}<1$.
So we have

$$\begin{aligned} f\left(\mathbf{x}^{(k+1)}\right) &\leq c \cdot \left(f\left(\mathbf{x}^{(k)}\right) - p^*\right) \\ f\left(\mathbf{x}^{(k)}\right) &\leq c \cdot \left(f\left(\mathbf{x}^{(k-1)}\right) - p^*\right) \\ &\vdots \\ f\left(\mathbf{x}^{(1)}\right) &\leq c \cdot \left(f\left(\mathbf{x}^{(0)}\right) - p^*\right) \end{aligned}$$

Multiply all the inequalities starts from the second one, we have

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

So above all, we have proved that the convergence of backtracking line search on a m-strongly convex and M-smooth objective function f as

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

where

$$c = 1 - \min\left\{2m\alpha, \frac{2\beta\alpha m}{M}\right\} < 1$$