CSE 535: Theory of Optimization and Continuous Algorithms **Winter 2019**

Homework 2

Lecturer: Yin Tat Lee *Due Date: 31 Jan 2018*

You are allowed to discuss with others. If you discuss with others, please list your collaborators for each question. In any case, you must write your own solutions.

1. Gradient Descent (20 marks)

Given a convex function such that $\mu I \preceq \nabla^2 f(x) \preceq LI$ for all $x \in \mathbb{R}^n$.

- (5 marks) Show that for any symmetric matrix H such that $\mu I \preceq H \preceq LI$, we have

$$H \succeq \frac{\mu L}{\mu + L} I + \frac{1}{\mu + L} H^2.$$

- (5 marks) Show that for any $x, y \in \mathbb{R}^n$, we have

$$\nabla f(x) - \nabla f(y) = H(x - y)$$

for some symmetric matrix H such that $\mu I \preceq H \preceq LI$.

- (5 marks) Using previous parts, show that

$$\langle \nabla f(x) - f(y), x - y \rangle \geq \frac{\mu L}{\mu + L} \|x - y\|^2 + \frac{1}{\mu + L} \|\nabla f(x) - \nabla f(y)\|^2.$$

- (5 marks) Consider the gradient descent $x^{(k+1)} = x^{(k)} - \frac{1}{L} \nabla f(x^{(k)})$. Using previous parts, show that

$$\|x^{(k)} - x^*\|^2 \leq (1 - \frac{2\mu}{\mu + L})^k \|x^{(0)} - x^*\|^2.$$

2. p-order Taylor Descent (35 marks)

Given a convex function such that $\|\nabla^{p+1} f(x)\|_{\text{op}} \leq 1$ for all $x \in \mathbb{R}^n$ where $p \geq 2$. Namely, the $(p+1)^{th}$ directional derivative of f on directions $h_1, h_2, \cdots, h_{p+1}$

$$\left| D^{p+1} f(x)[h_1, h_2, \cdots, h_{p+1}] \right| \leq 1 \text{ for any unit vectors } h_1, h_2, \cdots, h_{p+1}.$$

Let $f_x(y)$ be the p^{th} order Taylor expansion of f at x, namely,

$$f_x(y) = f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2}(y - x) \nabla^2 f(x)(y - x) + \cdots + \frac{1}{p!} D^p f(x)[y - x, y - x, \cdots, y - x].$$

- (10 marks) Show that $|f_x(y) - f(y)| \leq \frac{1}{(p+1)!} \|x - y\|_2^{(p+1)}$, $\|\nabla f_x(y) - \nabla f(y)\|_2 \leq \frac{1}{p!} \|y - x\|_2^p$ and $\|\nabla^2 f_x(y) - \nabla^2 f(y)\|_{\text{op}} \leq \frac{1}{(p-1)!} \|y - x\|_2^{p-1}$.
Hints: Taylor Theorem shows that for any function ϕ, we have

$$\phi(x) = \phi(a) + \phi'(a)(x - a) + \frac{\phi''(a)}{2}(x - a)^2 + \cdots + \frac{\phi^{(p)}(a)}{p!}(x - a)^p + \int_a^x \frac{\phi^{(p+1)}(a)}{p!}(x - t)^p dt.$$

- (5 marks) Show that for any x, the function $\overline{f}_x(y) = f_x(y) + \frac{1}{(p-1)!(p+1)} \|y - x\|_2^{p+1}$ is convex.
Hints: show that $\nabla_{zz}^2 \phi(z) \succeq (p+1) \|z\|_2^{p-1} \cdot I$ for the function $\phi(z) = \|z\|_2^{p+1}$.

- (5 marks) Consider the algorithm
$$x^{(k+1)} \leftarrow \arg\min_y f_{x^{(k)}}(y) + \frac{1}{(p-1)!(p+1)}\|y - x^{(k)}\|_2^{p+1}.$$
Show that $\|\nabla f(x^{(k+1)})\|_2 \leq \frac{p+1}{p!}\|x^{(k+1)} - x^{(k)}\|_2^p$. (Use the optimality condition $\nabla f_{x^{(k)}}(x^{(k+1)}) + \frac{1}{(p-1)!}\|x^{(k+1)} - x^{(k)}\|^{p-1}(x^{(k+1)} - x^{(k)}) = 0$.)
- (5 marks) For the same algorithm above, show that
$$\begin{aligned} f(x^{(k+1)}) &\leq f(x^{(k)}) - \frac{p-1}{(p+1)!}\|x^{(k+1)} - x^{(k)}\|_2^{p+1} \\ &\leq f(x^{(k)}) - \frac{1}{5p}\|\nabla f(x^{(k+1)})\|_2^{1+\frac{1}{p}} \end{aligned}$$
Hints: Use the fact that $|f_x(y) - f(y)| \leq \frac{1}{(p+1)!}\|x - y\|_2^{(p+1)}$.
- (10 marks) Let $D = \min_{f(x)\leq f(x^{(0)})} \|x - x^*\|_2$ and $\epsilon_k = f(x^{(k)}) - f^*$, For the same algorithm above, show that
$$\epsilon_{k+1} \leq \epsilon_k - \frac{1}{5p}\left(\frac{\epsilon_k}{D}\right)^{1+\frac{1}{p}}$$
and hence shows that
$$f(x^{k)}) - f^* \leq (5p^2)^p \frac{D^{p+1}}{k^p}.$$
Hints: Consider $\frac{1}{\epsilon_{k+1}^{1/p}} - \frac{1}{\epsilon_k^{1/p}}$ and used the fact that $(1-t)^{1/p} \leq 1 - \frac{t}{p}$.

3. Calculus Exercise (20 marks)

Given a convex function $f(x, y)$ in $\mathcal{C}^1$ where $x \in \mathbb{R}^{n_1}$ and $y \in \mathbb{R}^{n_2}$.

- (5 marks) For any x, let $y_x = \arg\min_y f(x, y)$ and $g(x) = \min_y f(x, y)$ (equivalently, $g(x) = f(x, y_x)$). Prove that
$$g(x') \geq g(x) + \langle \nabla_x g(x, y_x), x' - x\rangle \text{ for any } x, x'$$
where $\nabla_x g$ is the gradient of g with respect to the variable x.
- (5 marks) Show that
$$Dy_x[h] = -(\nabla_{yy}^2 f(x, y_x))^{-1}\nabla_{yx}^2 f(x, y_x)h.$$
Hints: Look at the optimality condition $\nabla_y f(x, y_x) = 0$ and take derivative on both side.
- (10 marks) Uses the previous two parts to show that
$$\nabla^2 g(x) = \nabla_{xx}^2 f(x, y_x) - \nabla_{xy}^2 f(x, y_x)\left(\nabla_{yy}^2 f(x, y_x)\right)^{-1}\nabla_{yx}^2 f(x, y_x).$$
Hints: Look at the optimality condition $\nabla f(x, y_x) = 0$ and take derivative on both side to get the derivative of y_x.

4. Recover the Hidden Image (25 marks)

Download the image from https://www.dropbox.com/s/awovylm9b9qaicr/noisy.png. Rescale the image such that the value of each pixel is between 0 and 1. This image is generated from a hidden image with the Salt-and-pepper noise. Namely, each pixel has 90% probability in turning to 0 or 1. You can assume the original image has no pixel with value exactly equals to 0 or exactly equals to 1. Design a convex function to model how good is a picture. Recover the image using gradient descent on the convex function you proposed.

- (10 marks) Write down the optimization problem that to recover the image.
Hints: If the pixel value is between $(0, 1)$ in the noisy image, we know that the value is correct because of the noise only change the value to 0 or 1. Therefore, you may want to find a smooth image with value matched with what we already know.

- (15 marks) Explain what you did. Show the recovered image.
  Hints: To solve a problem of the form $\min_{x_i = c_i \text{ for } i \in I} f(x)$ for some set I, the gradient descent on this function is
  $$x_i = \begin{cases} x_i - h \frac{\partial}{\partial x_i} f(x) & i \notin I \\ c_i & i \in I \end{cases}.$$
  Hints: When you program, it is very easy to make mistake for the gradient. So, it is always a good practice to test your gradient computation numerically by finite difference or use some auto differentiation library.
- (Alternative 15 marks for people who do not write program) Find 15 typos/mistakes/suggestions in the lecture note. 1 mark for each. (You cannot collaborate with other people for this question.)