CS/Math/Stat/ISyE 726: Nonlinear Optimization I	Fall 2020
Midterm Exam	
Instructor: Jelena Diakonikolas	October 16, 7–9 PM CT

Guidelines:

- This exam is 2 hours long. You are required to answer each of the FOUR multi-part questions.
- You should submit your midterm exam by uploading it to Canvas. To account for the time it takes you to prepare and upload your files, there is a "grace" period of 30 minutes. This means that if you upload your exam by 9:30 PM CT, there is no late penalty. After that, being late by 1-10 minutes carries a 10% penalty, 11-20 minutes carries a 20% penalty, 21-30 minutes carries a 50% penalty, and after 31 minutes or more of being late (on or after 10:01 PM CT), your exam is not accepted anymore and you get zero points.
- If you have any questions during the exam, please post them **publicly** on Piazza, so that all students have access to questions and answers.
- The questions are **not** ordered by difficulty; they are ordered by topic. It is recommended that you first read all the questions and work from easier towards the more difficult ones.
- You can use any of the results we showed in class or in the homework (unless you are asked to prove it). Clearly state the results you use. Substantiate all other claims you make.
- The exam is open book and open notes, but you are **not** allowed to use any online materials or collaborate with other students.

GOOD LUCK!

Question	Points
1	
2	
3	
4	
Total	

Q 1. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a continuously differentiable function.

(i) Prove that f is m-strongly convex for some $m > 0$ if and only if

$$(\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d): \quad f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{m}{2} \|\mathbf{y} - \mathbf{x}\|_2^2. \qquad [10pt]$$

(ii) Prove that f is m-strongly convex for some $m > 0$ if and only if

$$(\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d): \quad \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq m \|\mathbf{y} - \mathbf{x}\|_2^2. \qquad [10pt]$$

(iii) Recall that if f is m-strongly convex for some $m > 0$, then the following PL condition holds:

$$(\forall \mathbf{x} \in \mathbb{R}^d): \quad f(\mathbf{x}) - f(\mathbf{x}^*) \leq \frac{1}{2m} \|\nabla f(\mathbf{x})\|_2^2, \tag{1}$$

where $\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x})$.

Does the converse to this statement hold, that is, does the PL condition imply strong convexity? [5pt]

Q 2. Suppose that $f : \mathbb{R}^d \to \mathbb{R}$ is a convex and L-smooth function.

Recall (from Homework #3) that f is convex and L-smooth if and only if it satisfies the following inequality:

$$(\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d): \quad \frac{1}{2L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2 \le f(\mathbf{x}) - f(\mathbf{y}) - \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle . \tag{2}$$

Consider the steepest descent algorithm with step size $\frac{1}{L}$, i.e.,

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L}\nabla f(\mathbf{x}_k).$$

Prove that the potential function C_k defined by

$$C_k = \frac{k}{L}\|\nabla f(\mathbf{x}_k)\|_2^2 + f(\mathbf{x}_k)$$

is non-increasing with k. Conclude that, for all $k \ge 1$, we have

$$\|\nabla f(\mathbf{x}_k)\|_2^2 \le \frac{L(f(\mathbf{x}_0) - f(\mathbf{x}_k))}{k}.$$

Hint: It is useful to first show that $\forall k \ge 0 : \|\nabla f(\mathbf{x}_{k+1})\|_2 \le \|\nabla f(\mathbf{x}_k)\|_2$. [20pt]

Q 3. Let $f : \mathbb{R}^d \to \mathbb{R}$ be an L-smooth convex function that is minimized at some $\mathbf{x}^* \in \mathbb{R}^d$. Consider the following algorithm: you start with some $\mathbf{x}_0 \in \mathbb{R}^d$ and run Nesterov's accelerated method for smooth minimization for $k/2$ iterations, where k is an even integer number. Then, starting with the output point of Nesterov's algorithm, for the remaining $k/2$ iterations, you run gradient descent. Argue that the output point $\mathbf{x}_k$ of this algorithm must satisfy:

$$f(\mathbf{x}_k) - f(\mathbf{x}^*) = O\Big(\frac{L\|\mathbf{x}^* - \mathbf{x}_0\|_2^2}{k^2}\Big) \qquad \text{[7pt]}$$

and

$$\|\nabla f(\mathbf{x}_k)\|_2^2 = O\Big(\frac{L^2\|\mathbf{x}^* - \mathbf{x}_0\|_2^2}{k^3}\Big). \qquad \text{[8pt]}$$

Now suppose that f is, in addition, m-strongly convex for some $m > 0$. Consider restarting the algorithm described in this question every time the norm of the gradient halves (i.e., if the algorithm is started at $\mathbf{x}_0$, you restart when $\|\nabla f(\mathbf{x}_k)\|_2 \le \frac{1}{2}\|\nabla f(\mathbf{x}_0)\|_2$). Given $\epsilon > 0$, how many iterations (total) would it take this restarted algorithm to output a point $\mathbf{x}$ with $\|\nabla f(\mathbf{x})\|_2 \le \epsilon$? How about a point $\mathbf{x}$ with $f(\mathbf{x}) - f(\mathbf{x}^*) \le \epsilon$? [15pt]

Hint: Recall that strong convexity implies:

$$(\forall \mathbf{x} \in \mathbb{R}^d): \quad \frac{m}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2 \le f(\mathbf{x}) - f(\mathbf{x}^*) \le \frac{1}{2m}\|\nabla f(\mathbf{x})\|_2^2.$$

Q 4. Suppose you were given a fictional algorithm, AGD-G, which, given an initial point $\mathbf{x}_0 \in \mathbb{R}^d$, accuracy ϵ, and first-order oracle access to an L-smooth convex function f returns a point $\mathbf{x}_{\mathrm{A}}^{\mathrm{out}}$ such that

$$\|\nabla f(\mathbf{x}_{\mathrm{A}}^{\mathrm{out}})\|_2 \leq \epsilon$$

after $O\Big(\sqrt{\frac{L(f(\mathbf{x}_0)-f(\mathbf{x}^*))}{\epsilon}}\Big)$ iterations, where $\mathbf{x}^* \in \mathbb{R}^d$ is a minimizer of f. Note that we are assuming that AGD-G does not need to know L.

Use AGD-G to construct an algorithm AGD-G+ that, given an initial point $\mathbf{x}_0 \in \mathbb{R}^d$, accuracy ϵ, and first-order oracle access to an L-smooth and m-strongly convex function f, returns a point $\mathbf{x}_{\mathrm{A+}}^{\mathrm{out}}$ such that

$$\|\nabla f(\mathbf{x}_{\mathrm{A+}}^{\mathrm{out}})\|_2 \leq \epsilon$$

after $O\Big(\sqrt{\frac{L}{m}}\log(\frac{L\|\nabla f(\mathbf{x}_0)\|_2}{\epsilon})\Big)$ iterations. Your algorithm should work without the knowledge of L or m. [20pt]

Would your algorithm still work if f was not strongly convex, but instead satisfied the PL condition from Eq. (1)? Why or why not? [5pt]