GEORGIA INSTITUTE OF TECHNOLOGY
School of Electrical and Computer Engineering

ECE 6270
Midterm Exam

Monday, March 8, 2021

Name: Ge Xueren
Last, First

TIME LIMIT: 120 minutes. MUST BE SUBMITTED BY 12:00pm

Start time: 9:08 EST End time: 11:08 EST

Academic Integrity Statement

I pledge that I have neither given nor received any unauthorized aid on this quiz.

Xueren Ge
Signature

- Open notes. You may use any of the course materials during this quiz.
- "Closed internet". While you may consult online course materials on or linked to from my website or canvas, the entire internet is not a resource. Do not search for similar problems (or posting these problems) on Chegg, stackexhcange, etc. I do not think you would find anything helpful anyway, but since the internet is a big place and my own searches don't always reveal everything that might be available, I am asking in the interest of fairness that no one consult any outside resources. And of course, you may not ask for help from others.
- This quiz will be conducted under the rules and guidelines of the Georgia Tech Honor Code. **Please sign below the statement above.**
- Please submit all work. You may perform your work on a printed copy of the quiz itself (preferred), or on your own scratch paper. In either case, be sure to take pictures/scan all of your work before uploading. If working on your own paper, **please work each question on a separate sheet of paper and clearly identify your answer by drawing a box around it** (where appropriate). Try to mirror the structure of the quiz itself as much as possible in terms of which pages/where to place answers.
- You **can** use resources like calculators, Wolfram Alpha, Python, MATLAB, etc. However, be sure to document your work as clearly as possible. I cannot give you partial credit if you write nothing down but an answer if it is incorrect.

Problem 1 (16 pts): For each of the following functions on $\mathbb{R}^N$, indicate if it is convex, concave, both, or neither by circling the appropriate answer.

1. $f(\boldsymbol{x}) = \sqrt{\sum_{n=1}^{N} x_n^2}$.

   Concave Convex Both Neither

2. $f(\boldsymbol{x}) = \left(\sum_{n=1}^{N} \sqrt{|x_n|}\right)^2$.

   Concave Convex Both Neither

3. $f(\boldsymbol{x}) = \max_{m=1,\ldots,M} \boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{x}$, where $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M \in \mathbb{R}^N$ are fixed vectors and can be arbitrary.

Concave | Convex | Both | Neither

4. $f(\boldsymbol{x}) = \min_{m=1,\ldots,M} \boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{x}$, where $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M \in \mathbb{R}^N$ are fixed vectors and can be arbitrary.

Concave | Convex | Both | Neither

Problem 2 (20 pts): Consider the following functions on $\mathbb{R}$. For each, answer the following questions:

- Is f convex on $\mathbb{R}$?
- Is f strictly convex on $\mathbb{R}$?
- Is f strongly convex on $\mathbb{R}$?
- Is f M-smooth on $\mathbb{R}$ for some $M < \infty$?

Indicate your answer by circling the properties that f satisfies. Show your work, and indicate m and M where appropriate.

1. $f(x) = |x|$

Convex	Strictly convex	Strongly convex $m =$	M-smooth $M =$

no lower bound and higher bound

$|y| \le |x| + \langle y-x, \nabla f(x) \rangle + \frac{M}{2} \|y-x\|_2^2$

not strongly convex and M-smooth.

2. $f(x) = x^2$

Convex	Strictly convex	Strongly convex $m =$	M-smooth $M =$

$\nabla^2 f(x) = 2$

$\therefore m = 2$
$M = 2$

3. $f(x) = x^3$

Convex | Strictly convex | Strongly convex $m =$ | M-smooth $M =$

$\nabla^2 f(x) = 6x$

$6x \in [-\infty, +\infty]$ is not bounded.

$f(x) = x^3$ is not convex

not strongly convex and M-smooth

4. $f(x) = |x|^3$

Convex | Strictly convex | Strongly convex $m =$ | M-smooth $M =$

$\nabla^2 f(x) = 6|x| \geq 0$

it's strictly convex

but not strongly convex and not M-smooth.

Problem 3 (14 pts): In analyzing the convergence of various algorithms we have generally focused mostly on guarantees of the form

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \leq \epsilon.$$

A natural question is whether or not this tells us anything about $\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2$.[1]

1. Show that if f is strongly convex with parameter m then $\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 \leq \frac{2}{m}\left(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)\right)$.

$$f(y) \geq f(x) + \langle y - x, \nabla f(x) \rangle + \frac{m}{2} \|y - x\|_2^2$$

$$f(x_k) \geq f(x^*) + \langle x_k - x^*, \nabla f(x^*) \rangle + \frac{m}{2} \|x_k - x^*\|_2^2$$

$$f(x_k) - f(x^*) \geq \langle x_k - x^*, 0 \rangle + \frac{m}{2} \|x_k - x^*\|_2^2$$

$$f(x_k) - f(x^*) \geq \frac{m}{2} \|x_k - x^*\|_2^2$$

$$\frac{2}{m}\left(f(x_k) - f(x^*)\right) \geq \|x_k - x^*\|_2^2$$

2. Show that if f is M-smooth then $\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 \geq \frac{2}{M}\left(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)\right)$.

$$f(y) \leq f(x) + \langle y - x, \nabla f(x) \rangle + \frac{M}{2} \|y - x\|_2^2$$

$$f(x_k) \leq f(x^*) + \langle x_k - x^*, \nabla f(x^*) \rangle + \frac{M}{2} \|x_k - x^*\|_2^2$$

$$f(x_k) \leq f(x^*) + \langle x_k - x^*, 0 \rangle + \frac{M}{2} \|x_k - x^*\|_2^2$$

$$f(x_k) - f(x^*) \leq \frac{M}{2} \|x_k - x^*\|_2^2$$

$$\|x_k - x^*\|_2^2 \geq \frac{2}{M}\left(f(x_k) - f(x^*)\right)$$

[1] The solutions to these problems are short. If you do not see it relatively quickly, skip them and come back later.

Problem 4 (10 pts): Consider the quadratic function

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}, \tag{1}$$

where $\boldsymbol{Q} \in \mathbb{S}_{++}^{N}$ and $\boldsymbol{b} \in \mathbb{R}^{N}$. In class we derived an explicit formula for the exact minimizer of the one-dimensional function $\phi(\alpha) = f(\boldsymbol{x} + \alpha\boldsymbol{d})$. Specifically, the α that minimizes ϕ is given by

$$\alpha^{\star} = \frac{\boldsymbol{d}^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{Q}\boldsymbol{x})}{\boldsymbol{d}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{d}}.$$

Recall the Armijo condition for sufficient decrease:

$$f(\boldsymbol{x} + \alpha\boldsymbol{d}) \leq f(\boldsymbol{x}) + c_1\alpha\langle\boldsymbol{d}, \nabla f(\boldsymbol{x})\rangle.$$

For what values of c_1 does $\alpha^{\star}$ satisfy the Armijo condition?

To save you some time, I will point out that $f(\boldsymbol{x} + \alpha\boldsymbol{d}) - f(\boldsymbol{x}) = \frac{1}{2}\alpha^2\boldsymbol{d}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{d} + \alpha(\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b})^{\mathrm{T}}\boldsymbol{d}$.

Answer:

$(0, \frac{1}{2}]$

$$f(x+\alpha d) - f(x) = \frac{1}{2}\alpha^2 d^T Q d + \alpha(Qx-b)^T d$$

$$\frac{1}{2}\alpha^2 d^T Q d + \alpha(Qx-b)^T d \leq c_1\alpha \langle d, Qx-b\rangle$$

$$\frac{1}{2}\alpha^2 d^T Q d + \alpha(Qx-b)^T d \leq c_1\alpha(Qx-b)^T d$$

$$\frac{1}{2}\alpha^2 d^T Q d + (1-c_1)(Qx-b)^T d \leq 0$$

$$\frac{1}{2} d^T(b-Qx) \leq (c_1-1)(Qx-b)^T d$$

$$c_1 - 1 \leq -\frac{1}{2}$$

$$c_1 \leq \frac{1}{2}$$

Since $c_1 > 0$

hence $0 < c_1 \leq \frac{1}{2}$

Problem 5 (40 pts): Consider the function on $\mathbb{R}^2$ given by

$$f(\boldsymbol{x}) = f(x_1, x_2) = 2x_1^2 + x_2^2 - 2x_1(x_2 + 1).$$

In the following problems we will explore different ways to think about solving

$$\underset{\boldsymbol{x}\in\mathbb{R}^2}{\text{minimize}}\, f(\boldsymbol{x}). \tag{2}$$

1. Calculate the gradient $\nabla f(\boldsymbol{x})$.

$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 4x_1 - 2(x_2+1) \\ 2x_2 - 2x_1 \end{bmatrix}$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 4x_1 - 2(x_2+1) \\ 2x_2 - 2x_1 \end{bmatrix}$$

2. Calculate the Hessian $\nabla^2 f(\boldsymbol{x})$.

$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix}$

3. Is f convex? Justify your answer.

Circle one: Yes No

Justification: $D_1 = 4 > 0 \quad D_2 = 2 > 0$

$D_3 = 4 \times 2 - (-2)^2 = 4 > 0$

4. Find (analytically) the $\boldsymbol{x}$ that solves (2).

$\boldsymbol{x}^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$\nabla f(x) = 0$

$$\begin{cases} 4x_1 = 2(x_2+1) \\ 2x_2 = 2x_1 \end{cases}$$

$$\begin{cases} x_1 = 1 \\ x_2 = 1 \end{cases}$$

5. Suppose that you wish to solve this problem using gradient descent. If $\boldsymbol{x}_0 = \mathbf{0}$, what will the first step direction $\boldsymbol{d}_0$ be?

$\boldsymbol{d}_0 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

$$d = -\nabla f(x_0) = -\begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

6. With $\boldsymbol{x}_0 = \boldsymbol{0}$ and the $\boldsymbol{d}_0$ calculated in the previous problem, find the step size α that minimizes $f(\boldsymbol{x}_0 + \alpha \boldsymbol{d}_0)$, and calculate $\boldsymbol{x}_1$ using this α.

$\alpha_0 = \frac{1}{4}$

$\boldsymbol{x}_1 = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix}$

$f(x_0 + \alpha d_0) = f\left(\begin{bmatrix} 2\alpha \\ 0 \end{bmatrix}\right)$

$\nabla \phi(\alpha) = 16\alpha - 4 = 0 \qquad \alpha = \frac{1}{4}$

$x_1 = x_0 + \frac{1}{4} d_0 = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix}$

7. Repeat this process for one more step, i.e., compute $\boldsymbol{d}_1$, find the optimal α, and then compute $\boldsymbol{x}_2$.

$\boldsymbol{d}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$\alpha_1 = \frac{1}{2}$

$\boldsymbol{x}_2 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$

$x_2 = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$f(x_2) = \alpha^2 - \alpha + \frac{3}{2}$

$\nabla f = 2\alpha - 1 = 0 \qquad \alpha = \frac{1}{2}$

$x_2 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$

8. Now consider the heavy ball method where an exact line search is used at each iteration to choose α. The first iteration of the heavy ball method is identical to that of gradient descent, but the step from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$ will be different. Calculate the $\boldsymbol{x}_2$ that would result if using the heavy ball method where $\beta = \frac{1}{5}$ (which is the choice suggested by theory), but where α is chosen using an exact line search, i.e., α is chosen so as to minimize $f(\boldsymbol{x}_1 + \boldsymbol{p}_1 - \alpha\nabla f(\boldsymbol{x}_1))$.

$\alpha_1 = \frac{3}{5}$

$\boldsymbol{x}_2 = \begin{bmatrix} \frac{3}{5} \\ \frac{3}{5} \end{bmatrix}$

$$P_1 = \beta_1(X_1 - X_0) = \begin{bmatrix} \frac{1}{10} \\ 0 \end{bmatrix}$$

$$f(X_1 + P_1 - \alpha \nabla f(X_1))$$

$$= f\left(\begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{10} \\ 0 \end{bmatrix} - \alpha \begin{bmatrix} 0 \\ -1 \end{bmatrix}\right) = f\left(\begin{bmatrix} \frac{3}{5} \\ \alpha \end{bmatrix}\right)$$

$$\nabla f = 2\alpha - 2 \cdot \frac{3}{5}\alpha = 0 \quad \Rightarrow \quad \alpha = \frac{3}{5}$$

$$X_2 = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} + \frac{1}{5}\begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} - \frac{3}{5}\begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} \\ \frac{3}{5} \end{bmatrix}$$

9. Now consider the BFGS algorithm where, again, an exact line search is used at each iteration to choose α. Once again, the first iteration is identical to the previous methods, but $\boldsymbol{x}_2$ will be different. Calculate the $\boldsymbol{x}_2$ that would result for BFGS assuming that $\boldsymbol{H}_0^{-1} = \mathbf{I}$.

$\boldsymbol{H}_1^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$

$\alpha_1 = 1$

$\boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$x_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad H_0^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad g_0 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$

$d_0 = -H_0^{-1} g_0 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

$f(x_0 + \alpha_0 d_0) = f(\begin{bmatrix} 2\alpha_0 \\ 0 \end{bmatrix}) = 8\alpha_0^2 - 4\alpha_0$

$\nabla f = 16\alpha_0 - 4 = 0 \Rightarrow \alpha_0 = \frac{1}{4}$

$g_1 = \nabla f(x_1) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

$x_1 = x_0 + \alpha_0 d_0 = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} \quad s = x_1 - x_0 = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} \quad y = g_1 - g_0 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$

$a = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \quad p = 1 \quad H_1^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix} \quad d_1 = -H_1^{-1} g_1 = \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix}$

$f(x_1 + \alpha_1 d_1) = 2\left(\frac{\alpha_1 + 1}{2}\right)^2 + \alpha_1^2 - 2 \cdot \frac{\alpha_1 + 1}{2}(\alpha_1 + 1)$

$\nabla f = 4 \cdot \frac{\alpha_1 + 1}{2} \cdot \frac{1}{2} + 2\alpha_1 - 2(\alpha_1 + 1) = 0$

$\alpha_1 = 1$

$x_2 = x_1 + \alpha_1 d_1 = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

10. Now consider Newton's method. Start at $\boldsymbol{x}_0 = \boldsymbol{0}$ and use Newton's method to compute $\boldsymbol{x}_1$ assuming a fixed step size of $\alpha = 1$.

$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$d_0 = -(\nabla^2 f(0))^{-1} \nabla f\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$x_1 = \alpha d_0 + x_0$$

$$= 1 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$