EE 5239 Nonlinear Optimization Homework 1 Cover Sheet

Instructor name: Mingyi Hong & Jiaxiang Li

Student name: Alex Ojemann

- Date assigned: Monday 9/09/2024
- Date due: Friday 9/20/2024, at mid-night
- This cover sheet must be signed and submitted along with the homework answers on additional sheets.
- By submitting this homework with my name affixed above,
  - I understand that late homework will not be accepted,
  - I acknowledge that I am aware of the University's policy concerning academic misconduct (appended below),
  - I attest that the work I am submitting for this homework assignment is solely my own, and
  - I understand that suspiciously similar homework submitted by multiple individuals will be reported to the Dean of Students Office for investigation.
- Academic Misconduct in any form is in violation of the University's Disciplinary Regulations and will not be tolerated. This includes, but is not limited to: copying or sharing answers on tests or assignments, plagiarism, having someone else do your academic work or working with someone on homework when not permitted to do so by the instructor. Depending on the act, a student could receive an F grade on the test/assignment, F grade for the course, and could be suspended or expelled from the University.

1 Reading

- Reading: Textbook Section 1.1
- Appendix A.

2 Problems

1. Let $x \in \mathbb{R}^n$ be a vector. Explain why the following two quantities are norms

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|, \quad \|x\|_\infty = \max_i |x_i|.$$

Vector norms are functions from the vector space to the positive real numbers satisfy the following properties:

Non-negativity: $\|x\| \geq 0$ and $\|x\| = 0$ only if x is the zero vector.

Scalar multiplication: $\|ax\| = |a|\|x\|$

Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$

$\|x\|_1$ is non-negative because it is the sum of absolute values which can not be negative and are only 0 when the initial value is 0, it satisfies scalar multiplication because $\sum_{i=1}^{n} |ax_i| = |a| \sum_{i=1}^{n} |x_i|$, and it satisfies the triangle inequality because $\| \sum_{i=1}^{n} |x_i| + \sum_{i=1}^{n} |y_i| \| \leq \| \sum_{i=1}^{n} |x_i| \| + \| \sum_{i=1}^{n} |y_i| \|$ because $\|x_i + y_i\| \leq \|x_i\| + \|y_i\|$ is true for each element of the vectors x and y by the triangle inequality.

$\|x\|_\infty$ is non-negative because it is the maximum of absolute values which can not be negative and are only 0 when the initial value is 0, it satisfies scalar multiplication because $\max_i |a * x_i| = |a| \max_i |x_i|$, and it satisfies the triangle inequality because $\| \max_i |x_i| + \max_i |y_i| \| \leq \| \max_i |x_i| \| + \| \max_i |y_i| \|$ because $\|x_i + y_i\| \leq \|x_i\| + \|y_i\|$ is true for each element of the vectors x and y by the triangle inequality.

2. Let $x, y \in \mathbb{R}^n$. Suppose that x and y are orthogonal, i.e., $\langle x, y \rangle = x^T y = 0$. Show that the following **Pythagorean Theorem** is true (note $\langle x, x \rangle = \|x\|_2^2$)

$$\|x + y\|_2^2 = \|x\|_2^2 + \|y\|_2^2.$$

$$\|x + y\|_2^2 = \sum_{i=1}^{n} (x_i + y_i)^2 = \sum_{i=1}^{n} x_i^2 + 2x_i y_i + y_i^2$$

$x_i y_i = 0$ because x and y are orthogonal. Thus,

$$\|x + y\|_2^2 = \sum_{i=1}^{n} x_i^2 + y_i^2 = \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 = \|x\|_2^2 + \|y\|_2^2$$

3. Let $x, y \in \mathbb{R}^n$. Show that the following vector z is orthogonal to x

$$z = y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x.$$

$$\langle z, x\rangle = \langle y - \frac{\langle x, y\rangle}{\langle x, x\rangle}x, x\rangle = \langle y, x\rangle - \frac{\langle x, y\rangle}{\langle x, x\rangle}\langle x, x\rangle = \langle y, x\rangle - \langle y, x\rangle = 0$$

Since the dot product of z and x is 0, they are orthogonal.

4. Show that the Cauchy-Schwarz inequality is true, i.e., for any $x, y \in \mathbb{R}^n$, the following is true

$$\langle x, y\rangle \le \|x\|_2\|y\|_2.$$

(Hint: Giving x, y, construct z using the previous problem. Then Applying the Pythagorean Theorem to the vector x and z, i.e., $\|y\|_2^2 = \left\| z - \frac{\langle x,y\rangle}{\langle x,x\rangle}x \right\|_2^2$. Then expand the right hand side.)

Let $z = y - \frac{\langle x,y\rangle}{\langle x,x\rangle}x$

$$\|y\|_2^2 = \left\| z + \frac{\langle x, y\rangle}{\langle x, x\rangle}x \right\|_2^2 = \|z\|_2^2 + \left\| \frac{\langle x, y\rangle}{\langle x, x\rangle}x \right\|_2^2 = \|z\|_2^2 + \frac{\langle x, y\rangle^2}{\|x\|_2^2}$$

$\|z\|_2^2 \ge 0$, so

$$|y\|_2^2 \ge \frac{\langle x, y\rangle^2}{\|x\|_2^2}$$

$$\|x\|_2^2|y\|_2^2 \ge \langle x, y\rangle^2$$

$$\|x\|_2|y\|_2 \ge \langle x, y\rangle$$

5. Please find the gradient and the Hessian for the following functions

$$f(x) = \frac{1}{2}\|\mathbf{x} - \mathbf{b}\|^2,\ f(x) = \frac{1}{2}\|\mathbf{a}^T\mathbf{x} - \mathbf{b}\|^2,\ f(x) = \exp\{\mathbf{b}^T\mathbf{x} + \mathbf{c}\}.$$

Gradient of $f(x) = \frac{1}{2}\|\mathbf{x} - \mathbf{b}\|^2$: $|\mathbf{x} - \mathbf{b}\|$
Hessian of $f(x) = \frac{1}{2}\|\mathbf{x} - \mathbf{b}\|^2$: I
Gradient of $f(x) = \frac{1}{2}\|\mathbf{a}^T\mathbf{x} - \mathbf{b}\|^2$: $\|\mathbf{a}^T\mathbf{x} - \mathbf{b}\| * \frac{d}{d\mathbf{x}}\|\mathbf{a}^T\mathbf{x} - \mathbf{b}\| = \|\mathbf{a}^T\mathbf{x} - \mathbf{b}\| * \mathbf{a}$
Hessian of $f(x) = \frac{1}{2}\|\mathbf{a}^T\mathbf{x} - \mathbf{b}\|^2$: $\mathbf{a} * \mathbf{a}^T$
Gradient of $f(x) = \exp\{\mathbf{b}^T\mathbf{x} + \mathbf{c}\}$: $\exp\{\mathbf{b}^T\mathbf{x} + \mathbf{c}\} * \mathbf{b}$
Hessian of $f(x) = \exp\{\mathbf{b}^T\mathbf{x} + \mathbf{c}\}$: $\exp\{\mathbf{b}^T\mathbf{x} + \mathbf{c}\} * \mathbf{b} * \mathbf{b}^T$

6. Please prove the first and second-order Mean Value Theorem. That is, suppose $f(\cdot)$ is a smooth function. Then for any x, y in the domain of $f(\cdot)$, the following holds true:

$$f(x) - f(y) = f'(x^{\text{mid}})(x - y),$$
$$f(x) - f(y) = f'(y)(x - y) + \frac{1}{2}(x - y)^2 f''(x^{\text{mid}}),$$

for some x^{mid} that lies between the line segments between x, y.

First-Order Mean Value Theorem:

The first order mean value theorem states that:

$$f(x) - f(y) = f'(x^{\text{mid}})(x - y)$$

for some $x^{\text{mid}} \in (y, x)$.

Define a function $g(t)$ as follows:

$$g(t) = f(t) - f(y) - \frac{f(x) - f(y)}{x - y}(t - y)$$

This function satisfies $g(y) = g(x) = 0$, because:

$$g(y) = f(y) - f(y) = 0$$

$$g(x) = f(x) - f(y) - \frac{f(x) - f(y)}{x - y}(x - y) = 0$$

Since $g(t)$ is continuous on the closed interval $[y, x]$, differentiable on the open interval (y, x), and satisfies $g(y) = g(x) = 0$, we can apply Rolle's theorem. Rolle's theorem guarantees that there exists some $x^{\text{mid}} \in (y, x)$ such that:

$$g'(x^{\text{mid}}) = 0$$

Now, differentiate $g(t)$ with respect to t:

$$g'(t) = f'(t) - \frac{f(x) - f(y)}{x - y}$$

Setting this equal to zero at $t = x^{\text{mid}}$ (from Rolle's theorem), we get:

$$f'(x^{\text{mid}}) = \frac{f(x) - f(y)}{x - y}$$

Multiplying both sides by $(x - y)$, we obtain the desired first-order Mean Value Theorem:

$$f(x) - f(y) = f'(x^{\text{mid}})(x - y)$$

Second-Order Mean Value Theorem:

The second-order Mean Value Theorem states that:

$$f(x) - f(y) = f'(y)(x - y) + \frac{1}{2}(x - y)^2 f''(x^{\text{mid}})$$

for some $x^{\text{mid}} \in (y, x)$.

Define the following function $h(t)$:

$$h(t) = f(t) - f(y) - f'(y)(t - y) - \frac{f(x) - f(y) - f'(y)(x - y)}{(x - y)^2}(t - y)^2$$

This function satisfies $h(y) = h(x) = 0$, because:

$$h(y) = f(y) - f(y) = 0$$

$$h(x) = f(x) - f(y) - f'(y)(x-y) - \frac{f(x) - f(y) - f'(y)(x-y)}{(x-y)^2}(x-y)^2 = 0$$

Since $h(t)$ is continuous on $[y, x]$, differentiable on (y, x), and $h(y) = h(x) = 0$, Rolle's theorem guarantees the existence of $x^{\text{mid}} \in (y, x)$ such that:

$$h'(x^{\text{mid}}) = 0$$

Differentiating $h(t)$ with respect to t:

$$h'(t) = f'(t) - f'(y) - \frac{2(t-y)}{(x-y)^2}(f(x) - f(y) - f'(y)(x-y))$$

At $t = x^{\text{mid}}$, we have:

$$f'(x^{\text{mid}}) = f'(y) + \frac{2(x^{\text{mid}} - y)}{(x-y)^2}(f(x) - f(y) - f'(y)(x-y))$$

Simplifying this expression yields the second-order Mean Value Theorem:

$$f(x) - f(y) = f'(y)(x-y) + \frac{1}{2}(x-y)^2 f''(x^{\text{mid}})$$

Note: The following two questions can be answered after finishing Lecture 1.

7. Exercise 1.1.1 in the textbook

$$\frac{\partial f}{\partial x} = 2x + \beta y + 1 = 0$$

$$\frac{\partial f}{\partial y} = 2y + \beta x + 2 = 0$$

$$x = -\frac{\beta y + 1}{2}$$

$$2y + \beta\left(-\frac{\beta y + 1}{2}\right) + 2 = 0$$

$$2y - \frac{\beta^2 y + \beta}{2} + 2 = 0$$

$$4y - (\beta^2 y + \beta) + 4 = 0$$

$$(4 - \beta^2)y + 4 - \beta = 0$$

$$y = \frac{\beta - 4}{4 - \beta^2}$$

$$x = -\frac{\beta\left(\frac{\beta-4}{4-\beta^2}\right)+1}{2}$$

These are the coordinates of the stationary point. For the stationary points to be global minima, the hessian must be positive definite.

$$H = \begin{bmatrix} 2 & \beta \\ \beta & 2 \end{bmatrix}$$

$$\det(H) = 4 - \beta^2$$

The hessian is positive definite when $|\beta| < 2$.

8. Exercise 1.1.2 (a), (b), (c), (d) in the textbook

   a.

$$\frac{\partial f}{\partial x} = 4(x^2 - 4)x = 0$$

$$\frac{\partial f}{\partial y} = 2y = 0$$

$$x = 0, x = 2, x = -2, y = 0$$

$$f(2,0) = f(-2,0) = (2^2 - 4)^2 + 0^2 = 0$$

$$f(0,0) = (0^2 - 4)^2 + 0^2 = 16$$

   Thus, (2,0) and (-2,0) are global minima and (0,0) is a stationary point that is neither a maximum nor a minimum.

   b.

$$\frac{\partial f}{\partial x} = x + \cos y = 0$$

$$\frac{\partial f}{\partial y} = -x \sin y = 0$$

   The stationary points are $((0, \frac{\pi}{2})), ((0, \frac{3\pi}{2})), ((-1, 0)), ((1, \pi))$.

   c.

$$\frac{\partial f}{\partial x} = \cos x + \cos(x + y) = 0$$

$$\frac{\partial f}{\partial y} = \cos y + \cos(x + y) = 0$$

This only occurs at $(0,0)$

d.

$$\frac{\partial f}{\partial x} = -4x(y-x^2) - 2x = 0$$

$$\frac{\partial f}{\partial y} = 2(y-x^2) = 0$$

This occurs at $(0,0)$.

H = $\begin{bmatrix} -4(y-x^2)+8x^2-2 & -4x \\ -4x & 2 \end{bmatrix}$

The eigenvalues of this matrix when x = y = 0 are -2 and 2. Since one is positive and one is negative, this is a saddle point.

Note 1: All problems are referred using version 2 of the textbook. For those who use version 3 of the book, please see the scanned version of the HW assignment posted on the canvas.
Note 2: You are encouraged to type the solution of HW 1. You can use either Latex or Word. The Latex file for this problem has been provided.