NYCU CS 535520 Fall 2024: Optimization Algorithms (**Due: 2024/09/18, 21:00**)

Homework 0: Fundamentals

Instructor: Ping-Chun Hsieh

Submission Guidelines: Please compress all your write-ups into one PDF file (photos/scanned copies are acceptable; please make sure that the electronic files are of good quality and reader-friendly) and submit the file via E3.

Problem 1 (Optimality Conditions) (10+10=20 points)

(a) To ensure that we are fully comfortable with various optimality conditions, let's answer the following True-False questions. Please mark each subproblem as "True" of "False" and provide explanations / justifications / counterexamples.

- (i) Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable in the ϵ-neighborhood of x^*, and x^* is a local minimizer of f. Then, we have $\nabla f(x^*) = 0$.
- (ii) If $f : X \to \mathbb{R}$ is a convex function, then $\nabla f(x^*) = 0$ is a sufficient condition for $x^* \in X$ to be a strict global minimizer.
- (iii) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex function. Then $\nabla f(x^*) = 0$ is both necessary and sufficient for x^* to be a global minimizer.
- (iv) Let $f : X \to \mathbb{R}$ be continuously differentiable and let $A \subseteq X$ be the feasible set (induced by the constraints). If x^* is a local minimizer of f over A, then we have $\nabla f(x^*)^\top (x - x^*) \geq 0$, for all $x \in A$.

(b) Please prove the second-order necessary condition for optimality: Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable on an ϵ-neighborhood of $x^* \in \mathbb{R}^n$, and x^* is a local minimizer of f. Show that in addition to $\nabla f(x^*) = 0$, we must also have $\nabla^2 f(x^*) \succcurlyeq 0$. (Note: Please carefully justify every step of your proof)

Problem 2 (A Surprising Example of Optimality Conditions) (10+10=20 points)

In this problem, let's study a somewhat surprising example, where reckless intuition could be misleading. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Suppose that a point x^* is a local minimum of f along every line that passes through x^*, i.e., the function $h(t) := f(x^* + td)$ is minimized at $t = 0$ for all $d \in \mathbb{R}^n$.

(a) Show that x^* is a stationary point, i.e., $\nabla f(x^*) = 0$. (Note: Please carefully justify each step of your proof)

(b) Show by constructing a counterexample that x^* is not necessarily a local minimizer of f. (Hint: Consider the function of two variables $f(x_1, x_2) = (x_2 - px_1^2)(x_2 - qx_1^2)$, where $0 < p < q$. Show that $(0, 0)$ is indeed a local minimizer of f along every line that passes through $(0, 0)$. However, if we pick a scalar $m \in (p, q)$, then $f(x_1, mx_1^2) < 0$ if $x_1 \neq 0$ while $f(0, 0) = 0$).

Problem 3 (Convex Sets) (10 points)

In this problem, let's study the feasible set induced by a quadratic inequality, which will frequently show up in our subsequent lectures. Let $C \subseteq \mathbb{R}^n$ be the solution set of a quadratic inequality as follows:

$$C = \{x \in \mathbb{R}^n : x^\top A x + b^\top x + c \leq 0\}, \tag{1}$$

where A is real symmetric square matrix, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Show that C is a convex set if $A \succcurlyeq 0$. Moreover, is the converse statement true?

Problem 4 (Convex Functions) (8+8+10=26 points)

(a) Recall from Lecture 1 that the log-sum-exp function $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(x) := \log(\exp(x_1) + \cdots + \exp(x_n))$ is a popular function used in deep learning. Please verify that the log-sum-exp function $f(x)$ is a convex function.

(b) There are several operations that can preserve convexity. One widely-used operation is the "*pointwise maximum of convex functions.*" Specifically, let Θ be a parameter set (for simplicity, let's assume it is finite) and let $f(x;\theta) : \mathbb{R}^n \to \mathbb{R}$ be a convex function of x, for all $\theta \in \Theta$. Define the pointwise maximum of convex functions as

$$F(x) := \max_{\theta \in \Theta} f(x;\theta). \tag{2}$$

Please show that $F(x)$ remains a convex function of x.

(c) Another very useful property of convexity is that "*compositions of functions could still preserve convexity.*" Consider the composition $f(x) = h(g(x))$, where $g : \mathbb{R}^n \to \mathbb{R}^k$, $h : \mathbb{R}^k \to \mathbb{R}$, and $f : \mathbb{R}^n \to \mathbb{R}$, and g, h are both twice differentiable. Without loss of generality, we can restrict ourselves to the case $n = 1$ since convexity is determined by the behavior of a function on arbitrary lines that intersect its domain. As a result, we can rewrite the composition as

$$f(x) = h(g(x)) = h(g_1(x), \cdots, g_k(x)), \tag{3}$$

where $x \in \mathbb{R}$. In this case, the second-order derivative of f can be obtained by

$$f''(x) = g'(x)^\top \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^\top g''(x). \tag{4}$$

Please verify the following two properties:

- If h is convex and <u>non-decreasing</u> in each input argument, and all g_i are <u>convex</u>, then f is also convex.
- If h is convex and <u>non-increasing</u> in each input argument, and all g_i are <u>concave</u>, then f is also convex.

Problem 5 (Subgradients) (10 points)

As will be discussed in Lecture 2, we would like to establish an optimality condition for non-differentiable functions by using "subgradients." Specifically, for a function $f : X \to \mathbb{R}$, a vector g is called a subgradient of f at some $x \in X$ if $f(z) \geq f(x) + g^\top(z - x)$, for all $z \in X$. The set of all subgradients at $x \in X$, denoted by $\partial f(x)$, is called the "subdifferential" of f at x. Suppose f is a convex function. Show that if f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$. (Hint: Prove this by contradiction)

Problem 6 (Duality Gap and Strong Duality) (8+8+8=24 points)

As will be discussed in Lecture 2, in this problem, we practice how to derive the duality gap and verify that strong duality does not necessarily hold for convex problems. Consider the following optimization problem:

$$\min_{x,y} \quad \exp(-x), \quad \text{subject to } x^2/y \leq 0, \tag{5}$$

with variables x and y, and domain $\mathcal{D} = \{(x, y) : y > 0\}$.

(a) Verify that this is a convex optimization problem and find the optimal value.

(b) Write down the Lagrange dual problem (with Lagrange multiplier λ), and then find the optimal dual solution λ^* and the corresponding dual optimal value d^*. What is the duality gap?

(c) Does Slater's condition hold for this problem?