Nonlinear Optimization 1, Fall 2024 - Homework 1
Due one hour before lecture on 9/12 (Gradescope)

Your submitted solutions to assignments should be your own work. You are allowed to discuss homework problems with other students, but should carry out the execution of any thoughts/directions discussed independently, on your own. Acknowledge any source you consult.

Problem 1 - Computing gradients

For each of the following functions $f\colon \mathbb{R}^n \to \mathbb{R}$, write down the subset of $\mathbb{R}^n$ where the function is twice differentiable and compute its gradient and Hessian.

(a) $f(x) = \frac{1}{2}x^T Hx$ where $H \in \mathbb{R}^{n\times n}$ is a fixed matrix. What if H is symmetric?

(b) $f(x) = b^T Ax - \frac{1}{2}x^T A^T Ax$, where $A \in \mathbb{R}^{m\times n}$ is a fixed matrix and $b \in \mathbb{R}^m$ is a fixed vector.

(c) $f(x) = \|x\|_2 = \left(\sum_{i=1}^n x_i^2\right)^{1/2}$.

(d) $f(x) = \|Ax - b\|_2$, where $A \in \mathbb{R}^{m\times n}$ is a fixed matrix and $b \in \mathbb{R}^m$ is a fixed vector.

Problem 2 - Approximating functions

Answer all the following questions.

(a) Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a function. Assume that the partial derivatives $\partial f(x)/\partial x_i$ exists for all $i \in \{1, \ldots, d\}$. Prove or disprove that $\nabla f(x) = (\partial f(x)/\partial x_1, \ldots, \partial f(x)/\partial x_d)^\top$.

Now, for the next few questions consider any set $S \subseteq \mathbb{R}^n$ and twice continuously differentiable function $f\colon S \to \mathbb{R}$. Let $x \in S$ and $s \in \mathbb{R}^n$ be such that $x + ts \in S$ for all $t \in [0, 1]$.

(b) By defining $\theta(t) = f(x + ts)$ and using the Fundamental Theorem of Calculus:

$$\theta(1) = \theta(0) + \int_0^1 \theta'(t)\, dt,$$

show that

$$\left|f(x+s) - f(x) - \nabla f(x)^T s\right| \leq \frac{1}{2}L\|s\|_2^2$$

whenever f has an L-Lipschitz continuous gradient on S.

(c) Justify the formula

$$\theta(1) = \theta(0) + \theta'(0) + \int_0^1 \int_0^t \theta''(\alpha)\, d\alpha\, dt.$$

Hence, show that

$$\left|f(x+s) - f(x) - \nabla f(x)^T s - \frac{1}{2}s^T \nabla^2 f(x)s\right| \leq \frac{1}{6}Q\|s\|_2^3$$

whenever f has a Q-Lipschitz continuous Hessian on S with the operator norm (that is, when any $x, y \in S$ have $\|\nabla^2 f(x) - \nabla^2 f(y)\|_{op} \leq Q\|x - y\|_2$ where $\|M\|_{\mathrm{op}} = \sup\{\|Mu\|_2 \mid \|u\|_2 \leq 1\}$).

Problem 3 - Convex functions

Prove the following statement about convex functions.

(a) Show that a function is convex if, and only if, its epigraph is convex.

(b) For any $i \in \{1, \ldots, k\}$, let $f_i \colon \mathbb{R}^d \to \mathbb{R}$ be a convex function and $\alpha_i > 0$ a scalar.

1. Show that the function given by $f(x) = \sum \alpha_i f_i(x)$ is convex.
2. Similarly, show that $f(x) = \max\{f_1(x), \ldots, f_k(x)\}$ also yields a convex function.
3. Show that the composition of two convex functions is not necessarily convex.
4. Consider an affine map $T(x) = Ax + b$ where A is an $n \times d$ matrix and $b \in \mathbb{R}^n$, and a convex function $h \colon \mathbb{R}^n \to \mathbb{R}$. Show that their composition $h \circ T$ is convex.

Problem 4 - Grading scheme

Please read the syllabus defining the linear optimization problem that will be solved to maximize each student's course score. This question asks you to reason about and then write a short program to solve this three-dimensional optimization problem.

(a) Denote the set of all feasible grading rubrics $(H, M, F) \in \mathbb{R}^3$ as

$$\mathcal{P} = \{(H, M, F) \mid H{+}M{+}F \le 100,\ H, M \ge 15,\ F \ge M,\ 50 \le M{+}F \le 80,\ H{+}M{+}F \ge 90\}.$$

Prove the set $\mathcal{P}$ is convex (that is, for every $x, y \in \mathcal{P}$ and $\lambda \in [0, 1]$, $\lambda x + (1 - \lambda)y \in \mathcal{P}$).

(b) Denote the set of ten corners of $\mathcal{P}$ as

$$\mathcal{S} = \{(15, 40, 40),\ (20, 40, 40),\ (50, 25, 25),\ (40, 25, 25),\ (15, 37.5, 37.5),\\ (15, 15, 65),\ (20, 15, 65),\ (50, 15, 35),\ (40, 15, 35),\ (15, 15, 60)\}.$$

Carathéodory's Theorem ensures us that $x \in \mathcal{P}$ can be written as a convex combination of these corners: For any $x \in \mathcal{P}$, there exist coefficients $\lambda_p \ge 0$ for each $p \in \mathcal{S}$ such that

$$\sum_{p \in S} \lambda_p p = x \qquad \text{and} \qquad \sum_{p \in S} \lambda_p = 1.$$

Given this theorem, show that for any student with arbitrary course component grades (C_H, C_M, C_F, C_P), one of these ten corner points in $\mathcal{S}$ maximizes their course score (Hint: an average is smaller than a max).

(c) Knowing some corner must be optimal, write a program that computes the maximum course score for students given their four component gradings. Compute and output the (i) maximum course score and (ii) an optimal corner point in $\mathcal{S}$ for the following three hypothetical students

$$(C_H, C_M, C_F, C_P) = (100, 90, 80, 70),$$

$$(C_H, C_M, C_F, C_P) = (85, 85, 85, 85),$$

$$(C_H, C_M, C_F, C_P) = (70, 80, 90, 100).$$