Linear Algebra, Convexity, and Gradient Descent

R. Kyng & M. Probst *Problem Set 2 — Tuesday, March 2nd*

The exercises for this week will not count toward your grade, but you are highly encouraged to solve them all.

Exercise 0.

Prove that if a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is symmetric, then $\|\boldsymbol{A}\| = \max(|\lambda_{\max}(\boldsymbol{A})|, |\lambda_{\min}(\boldsymbol{A})|)$ and give an example of a non-symmetric matrix for which this is not true.

Exercise 1.

Consider a twice continuously differentiable function $f : S \to \mathbb{R}$, where $S \subset \mathbb{R}^n$ is a convex open set. Prove that f is β-gradient Lipschitz if and only if for all $\boldsymbol{x} \in S$ we have $\|\boldsymbol{H}_f(\boldsymbol{x})\| \leq \beta$.

Exercise 2.

Prove that when running Gradient Descent, $\|\boldsymbol{x}_i - \boldsymbol{x}^*\|_2 \leq \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2$ for all i.

Exercise 3.

Prove the following theorem.

Theorem. Let $f : \mathbb{R}^n \to \mathbb{R}$ be an β-gradient Lipschitz, convex function. Let $\boldsymbol{x}_0$ be a given starting point, and let $\boldsymbol{x}^* \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x})$ be a minimizer of f. The Gradient Descent algorithm given by

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \frac{1}{\beta} \boldsymbol{\nabla} f(\boldsymbol{x}_i)$$

ensures that the kth iterate satisfies

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \leq \frac{2\beta \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2^2}{k+1}.$$

Hint: do an induction on $1/gap_i$.

Exercise 4.

1. For each of the following functions answer these questions:

    - Is the function convex?

- Is the function β-gradient Lipschitz for some β?
- If the function is β-gradient Lipschitz give an upper bound on β – the bound should be within a factor 4 of the true value.

(a) $f(x) = |x|^{1.5}$ on $x \in \mathbb{R}$

(b) $f(x) = \exp(x)$ on $x \in \mathbb{R}$

(c) $f(x) = \exp(x)$ on $x \in (-1, 1)$

(d) $f(x, y) = \sqrt{x + y}$ on $(x, y) \in (0, 1) \times (0, 1)$.

(e) $f(x, y) = \sqrt{x + y}$ on $(x, y) \in (1/2, 1) \times (1/2, 1)$.

(f) $f(x, y) = \sqrt{x^2 + y^2}$ on $(x, y) \in \mathbb{R}^2$.

Special Exercise: Strongly Convex Functions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Assume f is twice continuously (Frechét) differentiable and that its first and second (Frechét) derivatives are integrable (basically, don't worry that weird stuff is happening with the derivatives).

Assume that for all $\boldsymbol{x}$, we have for some constant $\mu > 0$, that $\lambda_{\min}(H_f(\boldsymbol{x})) \geq \mu$. When this holds, we say that f is *μ-strongly convex*.

Part A. Prove that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \boldsymbol{\nabla} f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) + \frac{\mu}{2} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 .$$

Part B. Prove that there is value $L \in \mathbb{R}$ such that for all $\boldsymbol{x} \in \mathbb{R}^n$, we have $f(\boldsymbol{x}) \geq L$. In other words, the function is not unbounded below.

Part C. Prove that f is *strictly convex* as per Definition 3.2.8 in Chapter 3. Prove also that the minimizer $\boldsymbol{x}^* \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x})$ of f is unique.

Part D. Let $\boldsymbol{x}_0$ be a given starting point and $\boldsymbol{x}^*$ be the minimizer of f. Suppose we have an algorithm DecentDescent which takes a starting point $\boldsymbol{x}_0$, and a step count $t \in \mathbb{N}$. DecentDescent$(\boldsymbol{x}_0, t)$ runs for t steps and returns $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$ such that

$$f(\tilde{\boldsymbol{x}}) - f(\boldsymbol{x}^*) \leq \frac{\gamma \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2^2}{t + 1}$$

where $\gamma > 0$ is a positive number.

Assume that the *cost* of running DecentDescent for t steps is t. Explain how, with a total cost of at most $\frac{8\gamma}{\mu} \log(\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2 / \delta)$, we can produce a point $\widehat{\boldsymbol{x}} \in \mathbb{R}^n$ such that $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^*\|_2 \leq \delta$ for $\delta > 0$.

Part E. Consider a function $h : \mathbb{R}^n \to \mathbb{R}$ which is both μ-strongly convex and β-gradient Lipschitz. Give an algorithm that returns $\boldsymbol{x}'$ with

$$h(\boldsymbol{x}') - h(\boldsymbol{x}^*) \leq \epsilon$$

by computing the gradient of h at at most $\frac{32\beta}{\mu} \log(2\beta \left\| \boldsymbol{x}_0 - \boldsymbol{x}^* \right\|_2^2 / \epsilon)$ points.