ECE 490 (Introduction to Optimization) – Homework 2

Due: 11:59pm, Feb. 24

Problem 1. (15 points) Minimization of Quadratic Functions

(a) Suppose $f(x_1, x_2, x_3) = 2x_1^2 + 2x_2^2 + x_3^2 + 2x_2x_3 - 2x_1 - 2x_2 - 2x_3 + 5$. Find the minimum and maximum of f over $\mathbb{R}^3$ if they exist.

(b) Consider the positive definite quadratic minimization problem $\min_{x\in\mathbb{R}^n} \frac{1}{2}x^TQx$ where Q is a positive definite matrix. Apply the gradient method with the stepsize α_k chosen by the direct line search. What is α_k? Write out α_k as a function of Q and x_k.

(c) Consider the ridge regression problem $\min_{x\in\mathbb{R}^n} \frac{1}{n}\sum_{i=1}^n \left\{(a_i^Tx - b_i)^2 + \frac{\lambda}{2}||x||_2^2\right\}$, where $\lambda > 0$ and (a_i, b_i) (for $i = 1, 2, \cdots, n$) are given. What is the optimal solution x^*? is the optimal solution unique?

Problem 2. (15 points) Convexity and Concavity:

(a) Consider $f : \mathbb{R}^n \to \mathbb{R}$. For $y_1, y_2 \in \mathbb{R}^n$, define the function $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = f(x(y_1 - y_2) + y_2)$, for $x \in \mathbb{R}$. Prove that f is convex on $\mathbb{R}^n$ if and only if g is convex on $\mathbb{R}$ for all $y_1, y_2 \in \mathbb{R}^n$.

(b) Is the following set convex?

$$\mathcal{S} = \{x = (x_1, x_2) \in \mathbb{R}^2 : x_1 > 0, x_2 > 0, \text{ and } x_1 \log(x_1) + x_2 \log(x_2) \leq 4\}.$$

(c) Let $g : \mathbb{R}^n \to \mathbb{R}$ be a concave function and $f : \mathbb{R} \to \mathbb{R}$ be a concave increasing function. Prove that $f(g(x))$ is a concave function.

Problem 3. (10 points) Consider the problem of minimizing the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined as:

$$f(x) = f(x_1, x_2) = 2x_1^2 + 2x_2^4$$

. Use steepest descent with Armijo's Rule, with parameters $\tilde{\alpha} = 1$, $\sigma = 0.05$, and $\beta = 0.5$. Find α_k if $x_k = (1, 0)$.

Problem 4. (10 points) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and assume that $f(x) \geq f_{\min}$, for all x. Here $f_{\min}$ is finite. Consider the following version of gradient descent method with constant step size

$$x_{k+1} = x_k - \alpha D\nabla f(x_k),$$

where D is a positive definite matrix. Let $\lambda_{\min} > 0$ and $\lambda_{\max} > 0$ denote the minimum and maximum eigenvalues of D. Assume that f has Lipschitz gradients with Lipschitz constant L. Show that if

$$0 < \alpha < \frac{2\lambda_{\min}}{L\lambda_{\max}^2},$$

then

$$\lim_{k\to\infty} \nabla f(x_k) = 0.$$

Problem 5. Define $f : \mathbb{R}^2 \to \mathbb{R}$ as

$$f(x_1, x_2) = x_1^2 + 2\frac{1-\epsilon}{1+\epsilon}x_1x_2 + x_2^2$$

with $0 < \epsilon < 1$. Now, we consider the minimization problem of f, i.e. $\min_{x_1,x_2} f(x_1, x_2)$.

(a) (5 points) What are the minimizers of f?

(b) (10 points) Find the largest $m > 0$ and the smallest $M > 0$ in terms of ϵ such that

$$mI \preceq \nabla^2 f(x_1, x_2) \preceq MI$$

for all (x_1, x_2), where I is the identity matrix. Find the condition number of $\nabla^2 f$ given by $\kappa := \frac{M}{m}$ in terms of ϵ.

(c) (5 points) How does κ change as ϵ decreases to 0. Do you expect gradient descent to converge faster or slower as ϵ decreases to 0?

(d) (20 points) Write Python code to implement gradient descent with constant step-size $\alpha = \frac{m}{2M^2}$ starting from $(x_1, x_2) = (1, 1)$. Plot the trajectories of gradient descent for $\epsilon = 1, 0.1, 0.01, 0.001$ in the parameter space, i.e., the space of (x_1, x_2), along with the values of $f(x_1, x_2)$ during the optimization. Specifically, for each choice of ϵ, you need to plot two figures: (i) scatter plot showing (x_1, x_2) points for all iterations, where the x-axis if for x_1 and the y-axis is for x_2; you need to mark the initial and final iterations in the plot. (ii) plot the curve of $f(x_1, x_2)$ vs. the number of iterations. Does the variation of your observed rate of convergence with ϵ agree with your expectation in part (c)? What happens with the simulations if you try $\alpha = \frac{1}{M}$? Does the algorithm still converge?

Problem 6. (10 points) Consider a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$.

(a) Assume that f has Lipschitz gradients with Lipschitz constant L. In addition, f is also m-strongly convex. Prove the following inequality holds for all $x, y \in \mathbb{R}^n$

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \frac{mL}{m + L} \|x - y\|^2 + \frac{1}{m + L} \|\nabla f(x) - \nabla f(y)\|^2.$$

(b) Use the above inequality to show that the gradient descent method $x_{k+1} = x_k - \alpha \nabla f(x_k)$ with $\alpha = \frac{1}{L}$ satisfies the following convergence rate bound:

$$\|x_k - x^*\| \leq \left(1 - \frac{m}{L}\right)^k \|x_0 - x^*\|,$$

where x^* is assumed to be the unique global minimizer of f.