LINMA 2471 – Optimization Models and Methods II
Homework II

Derivative-Free Optimization with Finite Differences

1. Suppose that $f \in C^{1,1}_L(\mathbb{R}^n)$ and that $f(\cdot)$ is bounded from below by $f_{low} \in \mathbb{R}$. Given $h > 0$ and $x \in \mathbb{R}^n$, defined $g_h(x) \in \mathbb{R}^n$ by

$$[g_h(x)]_i = \frac{f(x + he_i) - f(x)}{h}, \quad i = 1, \dots, n.$$

Consider the following *derivative-free method* to minimize $f(\cdot)$:

$$x_{k+1} = x_k - \frac{1}{L} g_h(x_k), \quad \forall k \geq 0. \tag{1}$$

(a) If $\|\nabla f(x_k)\| > \epsilon$ and $h \leq \dfrac{\epsilon}{L\sqrt{n}}$, show that

$$\|g_h(x_k)\| \geq \frac{\|\nabla f(x_k)\|}{2}.$$

(b) If $\|\nabla f(x_k)\| > \epsilon$ and $h = \dfrac{\epsilon}{4L\sqrt{n}}$, show that

$$f(x_k) - f(x_{k+1}) \geq \frac{1}{8L} \|\nabla f(x_k)\|^2.$$

(c) Given $\epsilon > 0$, consider method (1) with $h = \dfrac{\epsilon}{4L\sqrt{n}}$ and let $T(\epsilon) \leq +\infty$ be the first iteration index such that $\|\nabla f(x_{T(\epsilon)})\| \leq \epsilon$. Prove that

$$T(\epsilon) \leq 8L\,(f(x_0) - f_{low})\,\epsilon^{-2}.$$

2. Consider the function $f_0 : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f_0(x) = \sin(x_1 + x_2) + (x_1 - x_2)^2 - 1.5x_1 + 2.5x_2 + 1. \tag{2}$$

(a) Show that $\nabla f_0(\cdot)$ is L-Lipschitz continuous and give explicitly the Lipschitz constant.

(b) Implement the gradient method with constant stepsize $1/L$ and the derivative-free method (1) with $h = \dfrac{\epsilon}{4L\sqrt{n}}$. Apply these two methods to minimize $f_0(\cdot)$ using the starting point

$$x_0 = \begin{bmatrix} -3 \\ 4 \end{bmatrix},$$

and the stopping criterion

$$\|\nabla f_0(x_k)\|_2 \leq \epsilon. \tag{3}$$

In a single table, display the number of iterations required by each method to generate x_k satisfying (3) for the values $\epsilon \in \left\{10^{-3}, 10^{-5}, 10^{-7}, 10^{-9}, 10^{-11}, 10^{-13}\right\}$. If a method fails for some ϵ, fill the corresponding entry of the table with "F".

(c) Provide a detailed explanation in case a failure is observed.

Nonsmooth Convex Optimization

3. Assume that $f : \mathbb{R}^n \to \mathbb{R}$ has a minimizer x^*, and that $f(\cdot)$ is M-Lipschitz continuous. Let $\{x_k\}_{k\geq 0}$ be generated by the subgradient method

$$x_{k+1} = x_k - \alpha_k g(x_k), \quad g(x_k) \in \partial f(x_k),$$

with $\alpha_k = \epsilon/\|g(x_k)\|^2$. Show that if

$$N \geq M^2\|x_0 - x^*\|^2\epsilon^{-2},$$

then

$$\min_{k=0,\ldots,N} \{f(x_k)\} - f(x^*) \leq \epsilon.$$

4. Consider the ℓ_1-regularized logistic regression problem

$$\min_{x\in\mathbb{R}^{n+1}} \; F(x) \equiv -\sum_{i=1}^{m} \left[b^{(i)} \log(m_x(a^{(i)})) + (1 - b^{(i)}) \log(1 - m_x(a^{(i)}))\right] + \lambda\|x\|_1, \tag{4}$$

where $\lambda > 0$, $\left\{(a^{(i)}, b^{(i)})\right\}_i^m \subset \mathbb{R}^{n+1} \times \{0, 1\}$ is the dataset (with $a_1^{(i)} = 1$ for $i = 1, \ldots, m$) and $m_x(a) \equiv 1/\left(1 + e^{-\langle a,x\rangle}\right)$ is the logistic model.

(a) Implement the subgradient method with stepsize $\alpha_k = \epsilon/\|g(x_k)\|^2$ (from **Exercise 3**), the proximal gradient method and the accelerated proximal gradient method (see **Lecture 4**) to solve (4).

(b) Consider problem (4) with $\lambda = 5$, and the full dataset **Iris**, which is freely available from the UC Irvine Machine Learning Repository

`https://archive.ics.uci.edu/dataset/53/iris`

More specifically, consider the type *Iris-virginica* to be the positive class and the other types to be the negative class. Apply your codes to the corresponding optimization problem. Use $\epsilon = 0.9$ (for the subgradient method), $x_0 = 0$, and execute 500 iterations of each code. In the same cartesian plane, plot the curves corresponding to $(k, \min_{i=0,\ldots,k} F(x_i))$ for each method. Use different colors to distinguish the methods.

(c) Is the performance of each method in accordance with the theoretical predictions? Provide detailed explanations for each method.