Optimization Methods in Machine Learning
Homework Assignment 1

Solve the problems in LaTeX format and submit a PDF file before the deadlines.

Soft Deadline: +1 week after the release date: 7 Nov 2024, 23:59
Hard Deadline: +2 weeks after the release date (but with a 25% penalty in points): 14 Nov 2024, 23:59

Preliminaries

Notations: $\mathbb{S}^n := \left\{A \in \mathbb{R}^{n\times n} \mid A = A^\top\right\}$, $\mathbb{S}^n_{++} := \left\{A \in \mathbb{R}^{n\times n} \mid A = A^\top, A \succcurlyeq 0\right\}$,
$\mathbb{S}^n_{++} := \left\{A \in \mathbb{R}^{n\times n} \mid A = A^\top, A \succ 0\right\}$, $\|A\|_F = \sqrt{\langle A, A\rangle}$ is the Frobenius norm, $I_n \in \mathbb{R}^{n\times n}$ is the identity matrix, $\|\cdot\|_{\mathrm{op}}$ is the operator norm w.r.t. the standard Euclidean norm $\|\cdot\|$.

In Problem 3, you will need the definition and theorem below.

Definition 1. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *L-smooth*, if there exists a constant $L \geq 0$ such that for all $x, y \in \mathbb{R}^d$,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

We can prove the theorem below (it is not required to prove it).

Theorem 2. *A twice differentiable function f is L–smooth if and only if $\left\|\nabla^2 f(x)\right\|_{\mathrm{op}} \leq L$ for all $x \in \mathbb{R}^n$.*

Problem 1

Let f be one of the following functions:

(a) $f : \mathbb{R}^n \to \mathbb{R}$, function $f(x) = \frac{1}{2}\|xx^T - A\|_F^2$, where $A \in \mathbb{S}^n$.

(b) $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$, function $f(x) = \frac{\langle Ax, x\rangle}{\|x\|^2}$, where $A \in \mathbb{S}^n$.

(c) $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$, function $f(x) = \langle x, x\rangle^{\langle x, x\rangle}$.

(d) $f : \mathbb{R}^n \to \mathbb{R}$, function $f(x) = \log\left(\sum\limits_{i=1}^{m} e^{\langle a_i, x\rangle}\right)$, where $a_1, \ldots, a_m \in \mathbb{R}^n$.

For each of the specified cases, compute the first and second derivatives df and $d^2 f$, the gradient vector ∇f, and the Hessian matrix $\nabla^2 f$. [3 points]

Problem 2

Let f be one of the following functions:

(a) $f : E \to \mathbb{R}$, function $f(t) = \det(A - tI_n)$, where $A \in \mathbb{R}^{n\times n}$, $E := \{t \in \mathbb{R} : \det(A - tI_n) \neq 0\}$.

(b) $f : \mathbb{R}_{++} \to \mathbb{R}$, function $f(t) = \|(A + tI_n)^{-1}b\|$, where $A \in \mathbb{S}^n_+$, $b \in \mathbb{R}^n$.

For each of the specified cases, compute the first derivative f'. [2 points]

Problem 3

Considering Definition 1 and Theorem 2, solve the problems:

- Let $g : \mathbb{R} \to \mathbb{R}$ be a differentiable function with the second derivative such that $|g''(x)| \le L$ for all $x \in \mathbb{R}$. Let $a \in \mathbb{R}^n$, $b \in \mathbb{R}$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be the function $f(x) = g(\langle a, x \rangle + b)$. Show that the function f is $L\|a\|^2$–smooth.
- Let $f : \mathbb{R}^n \to \mathbb{R}$ be the logistic regression loss function:

$$f(x) = \frac{1}{m} \sum_{i=1}^{m} \ln(1 + e^{\langle a_i, x \rangle}),$$

  where $a_1, \dots, a_m \in \mathbb{R}^n$. Show that f is $\frac{1}{4m} \sum\limits_{i=1}^{m} \|a_i\|^2$–smooth. [2 points]

Problem 4

Solve Exercise 1 from the lecture notes. [2 points]

Problem 5

We have four methods with the oracle complexities

$$\sqrt{\frac{LR^2}{\varepsilon}}, \quad \frac{LR^2}{\varepsilon}, \quad \frac{L}{\mu} \log \frac{R^2}{\varepsilon}, \quad \text{and} \quad \sqrt{\frac{L}{\mu}} \log \frac{R^2}{\varepsilon}$$

accordingly, where $L \ge 1$, $R \ge 1$, $0 \le \mu \le 1$, and $0 < \varepsilon \le 1$. For each possible combination of R, L, μ, and ε, which method should we prefer over others? [1 point]