MATH 173A: Homework #7

Due on Dec 3, 2024 at 23:59pm

Professor Cloninger

Ray Tsai

A16848188

Problem 1

Suppose a function $f : \mathbb{R}^d \to \mathbb{R}$ is L-smooth with $L = 4$ and satisfies the PL-property with parameter $\mu = 2$, i.e.,

$$\frac{1}{2}\|\nabla f(x)\|^2 \geq \mu(f - f^*).$$

Consider the gradient descent method for minimizing f. Let x^* be the global minimum and suppose $x^{(0)}$ is the initialization such that

$$\|x^* - x^{(0)}\| \leq 5.$$

Determine the step size η and the number of steps needed to satisfy

$$|f(x^{(t)}) - f(x^*)| \leq 10^{-4}.$$

Proof. The step size is $\eta = \frac{1}{L} = \frac{1}{4}$. The convergence rate is

$$\begin{aligned} f(x^{(t)}) - f(x^*) &\leq \left(1 - \frac{\mu}{L}\right)^t [f(x^{(0)}) - f(x^*)] \\ &= (0.5)^t [f(x^{(0)}) - f(x^*)]. \end{aligned}$$

Since f is L-smooth and $\|x^* - x^{(0)}\| \leq 5$,

$$\|\nabla f(x^{(0)})\| = \|\nabla f(x^{(0)}) - f(x^*)\| \leq L\|x^* - x^{(0)}\| \leq 4 \cdot 5 = 20.$$

By the PL-condition,

$$f(x^{(0)}) - f(x^*) \leq \frac{1}{2\mu}\|\nabla f(x^{(0)})\|^2 \leq \frac{1}{4} \times 400 = 100.$$

Hence,

$$f(x^{(t)}) - f(x^*) \leq (0.5)^t \times 100 \leq 10^{-4} \implies t \geq 6 \log_2 10 \approx 20.$$

$\square$

Problem 2

Consider the following set in $\mathbb{R}^n$ for an integer $s > 0$:

$$B = \{x \in \mathbb{R}^n \mid x_i \geq 0, \text{ for } i = 1, \dots, n \text{ and } x \text{ has at most } s \text{ nonzeros.}\}.$$

(a) Find an expression for the orthogonal projection of a point $x \in \mathbb{R}^n$ onto B (No need for justification).

Proof. Let $x_i^+ = \max(x_i, 0)$, and let $I_s(x)$ be the index set of the s largest components of x. Note that $|I_s(x)| = s$. Define projection $\Pi_B(x)$ by sending

$$x_i \mapsto \begin{cases} x_i & \text{if } i \in I_s(x^+) \\ 0 & \text{otherwise.} \end{cases}$$

□

(b) For the function

$$f(x) = \frac{1}{2}\|Ax - b\|^2,$$

write a projected gradient descent algorithm to solve

$$\min_{x \in \Omega} f(x)$$

for $\Omega = B$, with B from part (a). You need to specify the gradient formula and the projection formula. You do not need to specify the step size for this problem.

Proof. Let $x^{(0)} \in B$, and let μ be the step size. For $t = 1, \dots,$

1. Set $y^{(t+1)} = x^{(t)} - \mu \nabla f(x^{(t)}) = x^{(t)} - \mu A^T(Ax^{(t)} - b) = (I - \mu A^T A)x^{(t)} + \mu A^T b$.
2. Set $y_i^{(t+1)} = \max(0, y_i^{(t+1)})$ for all i.
3. Calculate $I_s(y^{(t+1)})$.
4. Set $x_i^{(t+1)} = \begin{cases} y_i^{(t+1)} & \text{if } i \in I_s(y^{(t+1)}) \\ 0 & \text{otherwise} \end{cases}$.

□

(c) Consider the function in (b) and suppose

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad s = 1$$

for the set B in (a). Does the projected gradient method converge to the global minimizer for any initialization $x^{(0)}$ if the step size $\mu \leq \frac{1}{8}$? Justify your answer.

Proof. No. Consider initializations $x^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $x^{(0)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

Case 1: $x^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$**.**

Following the steps in (b),

$$y^{(1)} = \begin{bmatrix} 1 \\ \mu \end{bmatrix}.$$

Since $\mu \leq 1$, $x^{(1)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and thus the algorithm converges to $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Case 2: $x^{(0)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$**.**

Following the steps in (b),

$$y^{(1)} = \begin{bmatrix} 4\mu \\ 1 \end{bmatrix}.$$

Since $4\mu \leq 0.5 \leq 1$, $x^{(1)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and thus the algorithm converges to $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. $\square$

But then $f(1,0) = 0.5$ and $f(0,1) = 2$, so the algorithm does converge to the global minimum for all initializations.