Coursework (3) for *Introductory Lectures on Optimization*

Xiaoyu Wang
3220104364

October 26, 2024

Excercise 1. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is L-smooth ($C_L^{1,1}$), and μ-PL, that is

$$\mu\text{-PL:}\ \ \frac{1}{2}\|\nabla f(\boldsymbol{x})\|_2^2 \geq \mu\left(f(\boldsymbol{x}) - f(\boldsymbol{x}^*)\right),$$

then GD iterates with step size $h_k = 1/L$ converge linearly, i.e.

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \leq \left(1 - \frac{\mu}{L}\right)^k \left(f(\boldsymbol{x}_0) - f(\boldsymbol{x}^*)\right).$$

Proof of Excercise 1: Proof: According to conditions we have:

1. $f : \mathbb{R}^n \to \mathbb{R}$ is L-smooth ($C_L^{1,1}$), and μ-PL, that is

$$\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2 \leq L\|\boldsymbol{x} - \boldsymbol{y}\|_2 \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

and in the class of week2,we learned that:

$$|f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle| \leq \frac{L}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2$$

$$\therefore f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle + \frac{L}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2$$

2. f satisfies the μ-PL condition:

$$\frac{1}{2}\|\nabla f(\boldsymbol{x})\|_2^2 \geq \mu\left(f(\boldsymbol{x}) - f(\boldsymbol{x}^*)\right)$$

where $\boldsymbol{x}^*$ is the minimizer of f.

We consider $\boldsymbol{y} = \boldsymbol{x} - h\nabla f(\boldsymbol{x})$

$$\begin{aligned} f(\boldsymbol{y}) &\leq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle + \frac{L}{2}\|\boldsymbol{y} - \boldsymbol{x}\|^2 && (1) \\ &= f(\boldsymbol{x}) - h\|\nabla f(\boldsymbol{x})\|^2 + \frac{h^2}{2}L\|\nabla f(\boldsymbol{x})\|^2 && (2) \\ &= f(\boldsymbol{x}) - h\left(1 - \frac{h}{2}L\right)\|\nabla f(\boldsymbol{x})\|^2 && (3) \end{aligned}$$

now $h = h_k = \frac{1}{L}$

$$\therefore f(\boldsymbol{y}) \leq f(\boldsymbol{x}) - \frac{1}{2L}\|\nabla f(\boldsymbol{x})\|^2$$

We start by analyzing the change in the function value from one iteration to the next.
Use $\boldsymbol{y} = \boldsymbol{x}_k \& \boldsymbol{x} = \boldsymbol{x}_{k-1}$ we have:

$$f(\boldsymbol{x}_k) \leq f(\boldsymbol{x}_{k-1}) - \frac{1}{2L} \|\nabla f(\boldsymbol{x}_{k-1})\|_2^2$$

Using the μ-PL condition,we have:

$$f(\boldsymbol{x}_k) \leq f(\boldsymbol{x}_{k-1}) - \frac{1}{2L} \cdot 2\mu \left(f(\boldsymbol{x}_{k-1}) - f(\boldsymbol{x}^*) \right)$$

$$\therefore f(\boldsymbol{x}_k) \leq f(\boldsymbol{x}_{k-1}) - \frac{\mu}{L} \left(f(\boldsymbol{x}_{k-1}) - f(\boldsymbol{x}^*) \right).$$

$$\therefore f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}_{k-1}) - f(\boldsymbol{x}^*) - \frac{\mu}{L} \left(f(\boldsymbol{x}_{k-1}) - f(\boldsymbol{x}^*) \right) = \left(1 - \frac{\mu}{L} \right) f(\boldsymbol{x}_{k-1}) - f(\boldsymbol{x}^*)$$

By recursively applying this inequality, so we get:

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \leq \left(1 - \frac{\mu}{L} \right) \left(f(\boldsymbol{x}_{k-1}) - f(\boldsymbol{x}^*) \right) \leq \left(1 - \frac{\mu}{L} \right)^2 \left(f(\boldsymbol{x}_{k-1}) - f(\boldsymbol{x}^*) \right) \leq \cdots \leq \left(1 - \frac{\mu}{L} \right)^k \left(f(\boldsymbol{x}_0) - f(\boldsymbol{x}^*) \right)$$

$$\therefore f(\boldsymbol{x}_{k-1}) - f(\boldsymbol{x}^*) \leq \left(1 - \frac{\mu}{L} \right)^k \left(f(\boldsymbol{x}_0) - f(\boldsymbol{x}^*) \right)$$

So, the proof is complete. $\square$