Notes for Lecture 3

Scribe: Tingkai Jia

1 Equivalent characterizations of L-smoothness

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex and differentiable function. Then the following properties are equivalent characterizations of L-smoothness of f:

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq L\|\mathbf{x} - \mathbf{y}\|_2 \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \tag{A}$$

$$\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y}\rangle \leq L\|\mathbf{x} - \mathbf{y}\|_2^2 \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \tag{B}$$

$$f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle \leq \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \tag{C}$$

$$f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle \geq \frac{1}{2L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2 \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \tag{D}$$

$$\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y}\rangle \geq \frac{1}{L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2 \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \tag{E}$$

Proof $A \Rightarrow B$: By Cauchy-Schwartz, we have

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

Proof $B \Rightarrow C$: Define the function $G : [0, 1] \to \mathbb{R}$ as

$$G(t) := f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), t(\mathbf{y} - \mathbf{x})\rangle,$$

so that $G(0) = 0$ and $G(1) = f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle$. By the fundamental theorem of calculus, we have

$$\begin{aligned}
G(1) - G(0) &= \int_0^1 G'(t)\,\mathrm{d}t = \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle \,\mathrm{d}t. \\
&= \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), t(\mathbf{y} - \mathbf{x})\rangle \,\frac{1}{t}\,\mathrm{d}t. \\
&\leq L\|\mathbf{y} - \mathbf{x}\|_2^2 \int_0^1 t\,\mathrm{d}t = \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2.
\end{aligned}$$

Proof $C \Rightarrow D$: We begin with a useful auxiliary lemma:

Lemma 1. *Consider a differentiable function $g : \mathbb{R}^d \to \mathbb{R}$ satisfying condition (C) and with its global minimum achieved at some $\mathbf{v}^*$. Then*

$$g(\mathbf{v}) - g(\mathbf{v}^*) \geq \frac{1}{2L}\|\nabla g(\mathbf{v})\|_2^2 \quad \textit{for all } \mathbf{v} \in \mathbb{R}^d.$$

Proof. We have

$$\begin{aligned}
g(\mathbf{v}^*) = \inf_{\mathbf{u} \in \mathbb{R}^d} g(\mathbf{u}) &\leq \inf_{\mathbf{u} \in \mathbb{R}^d} \left\{ g(\mathbf{v}) + \langle \nabla g(\mathbf{v}), \mathbf{u} - \mathbf{v}\rangle + \frac{L}{2}\|\mathbf{v} - \mathbf{u}\|_2^2 \right\} \\
&= g(\mathbf{v}) - \frac{1}{2L}\|\nabla g(\mathbf{v})\|_2^2,
\end{aligned}$$

where the last step follows by showing that the minimum of the quadratic program over $\mathbf{u}$ is achieved at $\mathbf{u}^* = \mathbf{v} - \frac{1}{L}\nabla g(\mathbf{v})$, and then performing some algebra.

Note: This lemma and its proof are of independent interest, as they show how gradient descent with step size $1/L$ can be thought of as minimizing a linear approximation along with a quadratic regularization term scaled by $L/2$.

Let us now show that $C \Rightarrow D$. For a fixed $\mathbf{x} \in \mathbb{R}^d$, define the function

$$g_x(\mathbf{z}) = f(\mathbf{z}) - \langle \nabla f(\mathbf{x}), \mathbf{z} \rangle.$$

Note that g_x is convex, differentiable and minimized when $\mathbf{z} = \mathbf{x}$, and it satisfies our smoothness condition. Hence, the preceding lemma with $\mathbf{v}^* = \mathbf{x}$ and $\mathbf{v} = \mathbf{y}$ implies that

$$g_x(\mathbf{y}) - g_x(\mathbf{x}) \geq \frac{1}{2L}\|\nabla g_x(\mathbf{y})\|_2^2 = \frac{1}{2L}\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|_2^2.$$

A little bit of calculation shows that

$$g_x(\mathbf{y}) - g_x(\mathbf{x}) = f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle,$$

which completes the proof. □

Proof $D \Rightarrow E$: We have

$$f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq \frac{1}{2L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2 \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d,$$

$$f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \leq \frac{1}{2L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2 \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d.$$

Adding these inequalities yields E.

Proof $E \Rightarrow A$: By Cauchy-Schwartz, we have

$$\langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \leq \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|_2 \|\mathbf{x} - \mathbf{y}\|_2.$$

2 Equivalent characterizations of μ-strong convexity

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex and differentiable function. Then the following properties are equivalent characterizations of μ-strong convexity of f:

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \geq \mu\|\mathbf{x} - \mathbf{y}\|_2 \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \tag{A}$$

$$\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq \mu\|\mathbf{x} - \mathbf{y}\|_2^2 \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \tag{B}$$

$$f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq \frac{\mu}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \tag{C}$$

$$f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \leq \frac{1}{2\mu}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2 \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \tag{D}$$

$$\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \leq \frac{1}{\mu}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2 \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \tag{E}$$

Note that all of these conditions can be obtained from the L-smoothness conditions by:

- flipping all the inequality signs, and
- replacing L by μ everywhere