Lecture 7

HW 2 is out

Last time

- Nonconvex smooth guarantees
- Characterization of L-smooth convex f

Today

- Finish proof
- Better guarantees for convex f
- Strongly convex

Proof continued

Here we continue proving the Lemma characterizing β-smooth functions.

Proof continued.

(4) $\Rightarrow$ (1) By Cauchy-Schwarz

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

$$\Updownarrow$$

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

(1) $\Rightarrow$ (3) Taylor Approximation Theorem.

(3) $\Rightarrow$ (4) For any x, y, define

$$g(\cdot) = f(\cdot) + \langle \nabla f(y), \cdot \rangle.$$

Note that g is convex and satisfies (3) and $\nabla g(y) = 0$.

Then, for $z = x - \frac{1}{L}\nabla g(x)$ we have

$$g(x) + \underbrace{\langle \nabla g(x), z - x\rangle + \frac{L}{2}\|z - x\|^2}_{-\frac{1}{2L}\|\nabla g(x)\|^2} \geq g(z) \geq g(y)$$

Then writing everything in terms of x an y

$$f(x) - \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 - \langle \nabla f(y), x\rangle \geq f(y) - \langle \nabla f(y), y\rangle$$

Reversing the role of $x \leftrightarrow y$

$$f(y) - \langle \nabla f(x), y\rangle - \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \geq f(x) - \langle \nabla f(x), x\rangle$$

$+$

Adding $\quad \langle \nabla f(y) - \nabla f(x), y - x\rangle \geq \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2 \quad \square$

Better guarantees for convex functions

Theorem: Let $f: \mathbb{R}^d \to \mathbb{R}$ be convex with L-Lipschitz grad. Let $\bar{x}^* \in \arg\min f$. Then, GD with $\alpha_k = 1/L$ produces

$$f(\bar{x}_{k+1}) - \min f \leq \frac{2L \|\bar{x}_0 - \bar{x}^*\|^2}{k}.$$

$\dashv$

How does this compare?

Our result for general L-smooth functions said we find $x \in \mathbb{R}^d$ s.t. $\|\nabla f(x)\| \leq \varepsilon$ after $\Omega(\frac{1}{\varepsilon^2})$ iterations.

One could compare the rates by

$$T = T_1 + T_2$$

T_1: uses convex bound (1)

T_2: uses general bound (2)

$$f(x_{T_1+1}) - \min f \leq \frac{2L \|x_0 - x^*\|^2}{T_1} \qquad (1)$$

$$\min_{T_1 \le k \le T} \|\nabla f(x_k)\|^2 \le \frac{2L\left(f(x_{T_1}) - \min f\right)}{T_2} \qquad (2)$$

$$\le \frac{4L^2 \|x_0 - x^*\|^2}{T_1 T_2}$$

$$T_1 = T_2 = \frac{2L\|x_0 - x^*\|^2}{\varepsilon} \;\rightarrow\; \le \varepsilon^2.$$

$\Rightarrow$ we find x s.t $\|\nabla f(x)\| \le \varepsilon$ after $\Omega(\frac{1}{\varepsilon})$.

square improvement.

Proof: First we prove that $\|x_k - x^*\|$ doesn't grow

$$\|\bar{x}_{k+1} - \bar{x}^*\|^2 = \|x_k - \tfrac{1}{L}\nabla f(x_k) - x^*\|^2$$

$$= \|x_k - x^*\|^2 - \frac{2}{L}\langle x_k - x^*, \nabla f(x_k)\rangle + \frac{1}{L^2}\|\nabla f(x_k)\|^2$$

Using (4) of previous Lemma

$$\le \|x_k - x^*\|^2 - \frac{2}{L^2}\|\nabla f(x_k)\|^2 + \frac{1}{L^2}\|\nabla f(x_k)\|^2$$

$$\le \|x_k - x^*\|^2.$$

Next, we prove the rate. Let $\delta_k = f(x_k) - \min f$.

From DL,

$$\delta_{k+1} \le \delta_k - \frac{1}{2L}\|\nabla f(x_k)\|_2^2 \qquad (\because)$$

By convexity

$$\delta_k \leq \nabla f(x_k)^T (x_k - x^*) \leq \|\nabla f(x_k)\| \|x_k - x^*\|.$$

$$\Rightarrow \frac{\delta_k}{\|x_k - x^*\|} \leq \|\nabla f(x_k)\|^2 \qquad (\heartsuit)$$

Then combining (☺) and ($\heartsuit$)

$$\delta_{k+1} \leq \delta_k - \frac{1}{2L} \frac{\delta_k^2}{\|x_k - x^*\|^2} \leq \delta_k - \frac{1}{2L} \frac{\delta_k \delta_{k+1}}{\|x_0 - x^*\|^2}$$

Multiply $\frac{1}{\delta_k \delta_{k+1}}$

$$\Rightarrow \frac{1}{\delta_k} \leq \frac{1}{\delta_{k+1}} - \frac{1}{2L} \frac{1}{\|x_0 - x^*\|^2}$$

$$\Rightarrow \frac{1}{\delta_{k+1}} \geq \frac{1}{\delta_k} + \frac{1}{2L} \frac{1}{\|x_0 - x^*\|^2} \geq \frac{1}{\delta_0} + \frac{k}{2L\|x_0 - x^*\|^2}$$

$$\Rightarrow \frac{1}{\delta_{k+1}} \geq \frac{k}{2L\|x_0 - x^*\|^2}$$

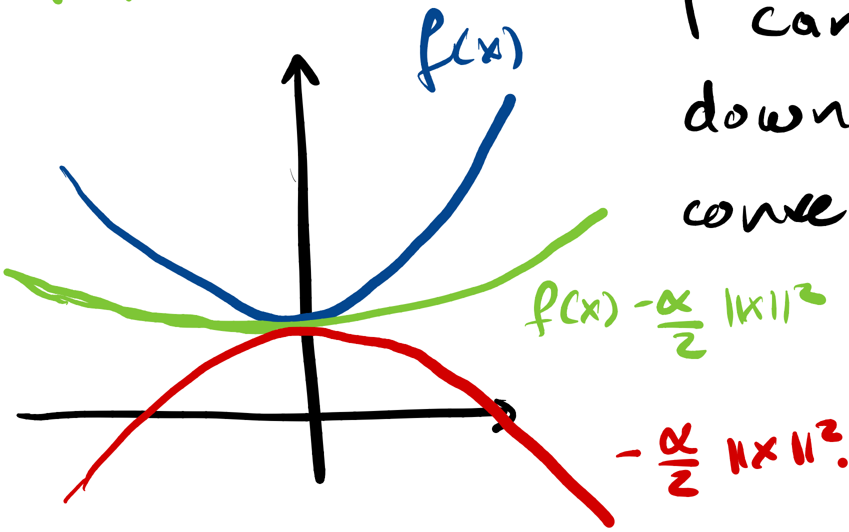
□

We saw before that additional curvature yields faster convergence.

Def: A function $f: \mathbb{R}^d \to \mathbb{R}$ is μ-strongly convex if $x \mapsto f(x) - \frac{\mu}{2}\|x\|^2$ is convex.

⊣

Intuition



I can curve my function down and it is still convex.

Lemma: For $f \in C^1$, the following are equivalent

(1) f is μ-strongly convex.

(2) $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y-x\|^2 \quad \forall x, y.$

(3) $\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \mu \|y-x\|^2 \quad \forall x, y.$

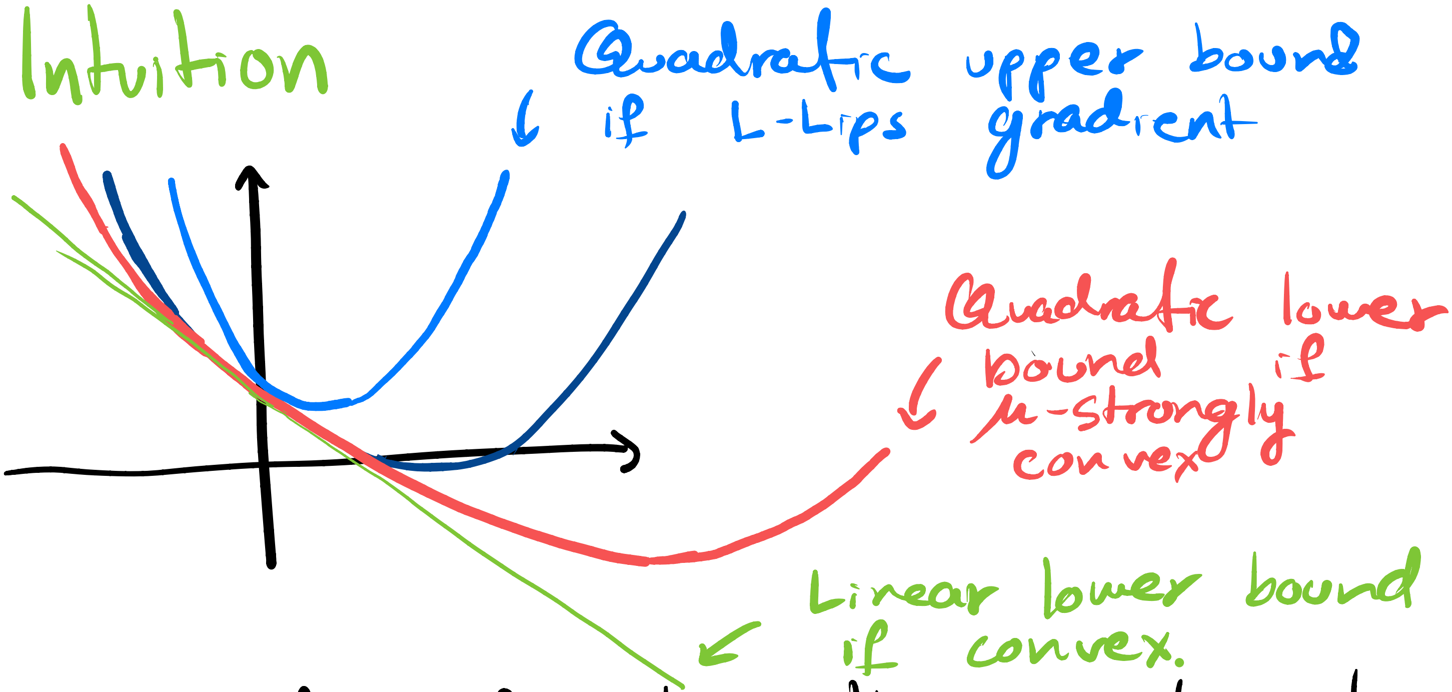
If f is C^2

(4) $\nabla f^2(x) \succeq \mu I \quad \forall x.$

Proof: These follow from the characterization of convexity for $f(\cdot) + \frac{\mu}{2}\|\cdot\|^2$.

Check! $\square$

Intuition



Lemma: Any f L-smooth, μ-strongly convex satisfies that

$$\forall x, y \quad \langle \nabla f(y) - \nabla f(x), y - x\rangle \geq \frac{\mu L}{\mu + L}\|x - y\|^2 + \frac{1}{\mu+L}\|\nabla f(y) - \nabla f(x)\|^2$$

Proof: Exercise.

Hint: Consider $g(x) = f(x) - \frac{\mu}{2}\|x\|^2$ and show that it is $(L-\mu)$-smooth. $\square$

Theorem: Let $f: \mathbb{R}^d \to \mathbb{R}$ be μ-strongly convex with L-Lipschitz gradient. Then GD with $\alpha_k = \frac{2}{\mu+L}$, we have

$$f(x_k) - \min f \le \frac{L}{2}\Big(\overbrace{\frac{K-1}{K+1}}^{\approx 1 - 2K^{-1}} \Big)^{2k} \|\bar{x}_0 - \bar{x}^*\|^2$$

$K = L/\mu$ condition number

Proof: Goal: show that $\|x_k - x^*\|$ contracts

$$\|\bar{x}_{k+1} - \bar{x}^*\|^2 = \|\bar{x}_k - \bar{x}^* - \frac{2}{\mu+L} \nabla f(\bar{x}_k)\|^2$$

$$= \|\bar{x}_k - \bar{x}^*\|^2 - \frac{4}{\mu+L} \langle \nabla f(\bar{x}_k), \bar{x}_k - \bar{x}^* \rangle + \frac{4}{(\mu+L)^2} \|\nabla f(\bar{x}_k)\|^2 \quad (*)$$

Let us upper bound the inner product

$$\langle \nabla f(\bar{x}_k) - \nabla f(x^*), \bar{x}_k - \bar{x}^* \rangle \ge \frac{1}{L+\mu} \|\nabla f(x_k)\|^2 + \frac{\mu L}{L+\mu} \|x_k - x^*\|^2 \quad (\ddot{o})$$

Then applying $(\ddot{o})$ in $(*)$ gives

$$\|\bar{x}_{k+1} - \bar{x}^*\|^2 \le \|\bar{x}_k - \bar{x}^*\|^2 + \frac{4}{(\mu+L)^2} \|\nabla f(\bar{x}_k)\|^2 - \frac{4}{(\mu+L)} \Big(\frac{1}{(\mu+L)} \|\nabla f(x_k)\|^2 + \frac{\mu L}{L+\mu} \|x_k - x^*\|^2 \Big)$$

$$= \overbrace{\Big(1 - \frac{4\mu L}{(L+\mu)^2}\Big)}^{\left(\frac{K-1}{K+1}\right)^2} \|x_k - x^*\|^2.$$

To prove the result for the function value

$$f(x_{k+1}) - f(x^*) \overset{\text{Taylor}}{\leq} \frac{L}{2} \|\bar{x}_{k+1} - x^*\|^2$$

$$\leq \frac{L}{2} \left(\frac{K-1}{K+1}\right)^{2K} \|x_0 - x^*\|^2.$$

$\square$