

## Lecture 7 (Sep/19)

HW2 is out (Due Friday 9/29 6PM)

Scribe?

Last time

- ▷ Non convex smooth guarantees
- ▷ Characterization of L-smooth convex f

Today

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

Here we continue proving the lemma characterizing  $\beta$ -smooth functions.

Proof continued.

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

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

↑  
↓

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

(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  $\nabla g(y) = 0$ .  
Then, for  $z = x - \frac{1}{L} \nabla g(x)$  we have

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

Then writing everything in terms of  $x$  and  $y$

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

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

# Better guarantees for convex functions

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

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

→

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

$$T = T_1 + T_2$$

$\uparrow$   
uses convex bound (1)

$\uparrow$   
uses general bound (2)

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

$$\min_{T_1 \leq k \leq T} \|\nabla f(x_k)\|^2 \leq \frac{2L(f(x_T) - \min f)}{T_2} \quad (2)$$

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

$$T_1 = T_2 = \frac{2L\|x_0 - x^*\|^2}{\epsilon} \rightarrow \leq \epsilon^2.$$

$\Rightarrow$  we find  $x$  s.t  $\|\nabla f(x)\| \leq \epsilon$  after  $\sqrt{\frac{L}{\epsilon}}$ .  
 square improvement.

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

$$\begin{aligned} \|\bar{x}_{k+1} - \bar{x}^*\|^2 &= \|x_k - \frac{1}{L}\nabla f(x_k) - x^*\|^2 \\ &= \|x_k - x^*\|^2 - \frac{2}{L} \langle x_k - x^*, \nabla f(x_k) \rangle \\ &\quad + \frac{1}{L^2} \|\nabla f(x_k)\|^2 \\ \text{Using (4) of previous Lemma } \swarrow &\leq \|x_k - x^*\|^2 - \frac{2}{L} \|\nabla f(x_k)\|^2 + \frac{1}{L^2} \|\nabla f(x_k)\|^2 \\ &\leq \|x_k - x^*\|^2. \end{aligned}$$

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

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

By convexity

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

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

Then combining (i) and (b)

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

Multiply

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

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

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

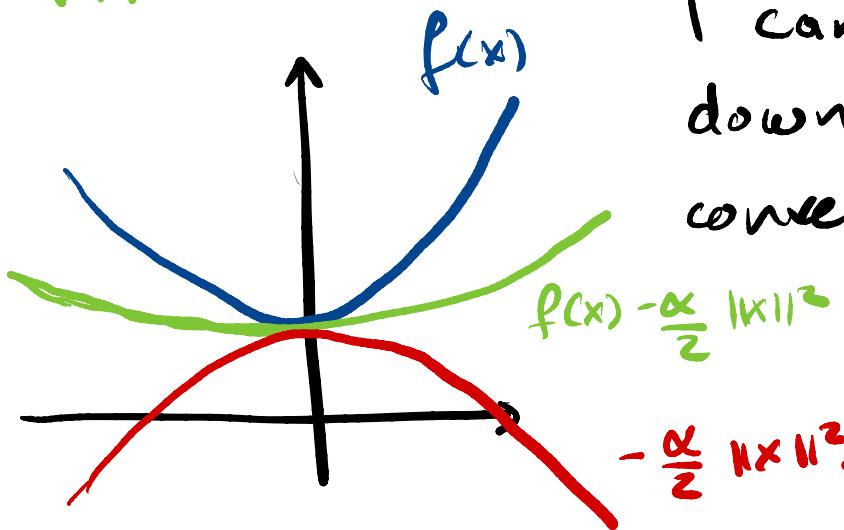
□

We saw before that additional curvature yields faster convergence.

Def: A function  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  is  $\mu$ -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  $\mu$ -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  $C^2$

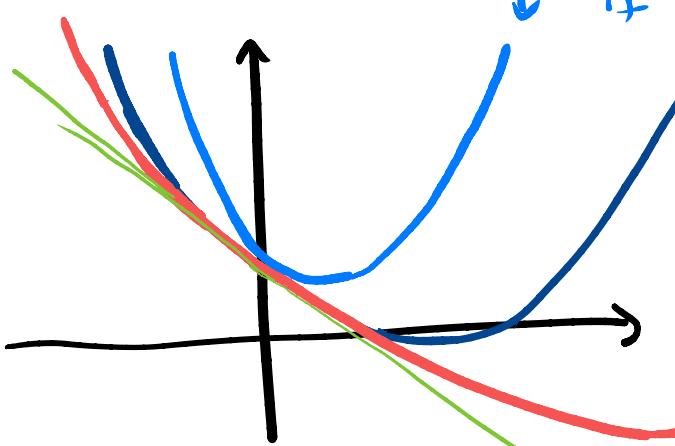
$$(4) \quad \nabla f^2(x) \succeq \mu I \quad \forall x.$$

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

Check!

□

## Intuition



Quadratic upper bound  
if L-Lipschitz gradient

Quadratic lower  
bound if  
 $\mu$ -strongly  
convex

Linear lower bound  
if convex.

Lemma: Any  $f$  L-smooth,  $\mu$ -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  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  be  $\mu$ -strongly convex with L-Lipschitz gradient. Then GD with  $\alpha_k = \frac{2}{\mu+L}$ , we have

$$\checkmark \approx 1 - 2\kappa^{-1}$$

$$f(x_k) - \min f \leq \frac{L}{2} \left( \frac{\kappa - 1}{\kappa + 1} \right)^{2K} \|x_0 - \bar{x}\|^2.$$

$\kappa = L/\mu$  condition number

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

$$\begin{aligned} \|x_{k+1} - \bar{x}^*\|^2 &= \|x_k - \bar{x}^* - \frac{2}{\mu+L} \nabla f(\bar{x}_k)\|^2 \\ &= \|x_k - \bar{x}^*\|^2 - \frac{4}{\mu+L} \langle \nabla f(\bar{x}_k), \bar{x}_k - \bar{x}^* \rangle \\ &\quad + \frac{4}{(\mu+L)^2} \|\nabla f(\bar{x}_k)\|^2 \quad (\star) \end{aligned}$$

Let us upper bound the inner product

$$\begin{aligned} \langle \nabla f(\bar{x}_k) - \nabla f(x^*), \bar{x}_k - \bar{x}^* \rangle &\geq \frac{1}{L+\mu} \|\nabla f(\bar{x}_k)\|^2 + \\ &\quad \frac{\mu L}{L+\mu} \|x_k - x^*\|^2 \quad (\ddot{\sigma}) \end{aligned}$$

Then applying  $(\ddot{\sigma})$  in  $(\star)$  gives

$$\begin{aligned} \|x_{k+1} - \bar{x}^*\|^2 &\leq \|x_k - \bar{x}^*\|^2 + \frac{4}{(\mu+L)^2} \|\nabla f(\bar{x}_k)\|^2 \\ &\quad - \frac{4}{(\mu+L)} \left( \frac{1}{L+\mu} \|\nabla f(\bar{x}_k)\|^2 + \right. \\ &\quad \left. \underbrace{\left( \frac{\kappa-1}{\kappa+1} \right)^2}_{\text{green}} \frac{\mu L}{L+\mu} \|x_k - x^*\|^2 \right) \\ &= \left( 1 - \frac{4\mu^2}{(\mu+L)^2} \right) \|x_k - x^*\|^2. \end{aligned}$$

To prove the result for the function value

$$\begin{aligned} f(x_{k+1}) - f(x^*) &\stackrel{\text{Taylor}}{\leq} \frac{\mathbb{L}}{2} \|\bar{x}_{k+1} - x^*\|^2 \\ &\leq \frac{\mathbb{L}}{2} \left( \frac{k-1}{k+1} \right)^{2k} \|x_0 - x^*\|^2. \end{aligned}$$

□