Gradient Descent

Last time

↳ GD on least-squares

$$\vec{x}_{k+1} - \underbrace{(A^TA)^{-1}A^T\vec{b}}_{\vec{x}^*} = (I - 2\eta A^TA)^{k+1}(\vec{x}_0 - \underbrace{(A^TA)^{-1}A^T\vec{b}}_{\vec{x}^*}))$$

↳ in order to converge $|\text{eig. values } (I - 2\eta A^TA)| < 1$

↳ η: step size

↳ want to choose η s.t. this converges

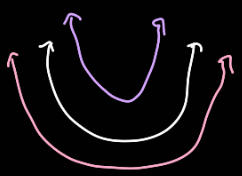
Q/ How do we generalize? (227C sneak peek ;))

- Gradient Descent for μ-strongly convex & L-smooth
  - μ-strongly convex: $\forall \vec{x}, \vec{y} \in$ domain

$$f(\vec{y}) \geq f(\vec{x}) + \nabla f(\vec{x})^T(\vec{y} - \vec{x}) + \frac{\mu}{2}\|\vec{y} - \vec{x}\|_2^2$$

lower bound

↳ trying to say gradients are changing "fast enough"



↳ saying that there is a quadratic lower bound to this fcn

  - L-smoothly convex:

$$f(\vec{y}) \leq f(\vec{x}) + \nabla f(\vec{x})^T(\vec{y} - \vec{x}) + \frac{L}{2}\|\vec{y} - \vec{x}\|_2^2$$

upper bound

↳ gradients are not changing too fast

Thm:

If: $\min_{\vec{x} \in \mathbb{R}} f(\vec{x}) = \vec{x}^*$ L-smooth, μ-strongly conv.

↳ optimum point

$f: \mathbb{R}^n \to \mathbb{R}$

[Recall: GD

$x_{t+1} = x_t - \eta \nabla f(x_t)$

$\lim_{t \to \infty} \vec{x}_t = \vec{x}_p$]

Then: $\|\vec{x}_{t+1} - \vec{x}_*\|_2^2 \leq C^{t+1}\|\vec{x}_0 - \vec{x}_*\|_2^2$

↳ can choose an η s.t. this bound is true

① Lemma: f : L-smooth, Then $\|\nabla f(\vec{x})\|_2^2 \leq 2L(f(\vec{x}) - f(\vec{x}_*))$

Proof:

$$f(\vec{y}) \leq f(\vec{x}) + \nabla f(\vec{x})^T(\vec{y}-\vec{x}) + \frac{L}{2}\|\vec{y}-\vec{x}\|_2^2$$

↳ gives a bound on the gradient, dep on how far away you are from the optimum

$$f(\vec{x}_*) \leq f(\vec{x}) + \nabla f(\vec{x})^T(\vec{x}_*-\vec{x}) + \frac{L}{2}\|\vec{x}-x^*\|_2^2$$

$$x_{t+1} = x_t - \eta \nabla f(x_t), \quad \vec{x} = \vec{x}_{t-1} - \eta \nabla f(\vec{x}_{t-1})$$

$$f(\vec{x}_*) - f(\vec{x}) \leq \nabla f(\vec{x})^T(\vec{x}_* - \vec{x}) + \frac{L}{2}\|\vec{x} - \vec{x}^*\|_2^2$$

A. A attempt

Hint: consider $f(\vec{x}) \overset{?}{} f(\vec{x} + \frac{\nabla f(\vec{x})}{L})$

$$f(\vec{x} + \frac{\nabla f(\vec{x})}{L}) \leq f(\vec{x}) + \nabla f(\vec{x})^T(\frac{\nabla f(\vec{x})}{L}) + \frac{L}{2}\|\frac{\nabla f(\vec{x})}{L}\|_2^2$$

$f(\vec{x}^*)$ is the minimum $\Rightarrow f(\vec{x}^*) \leq f(\vec{x})$

$$f(\vec{x}^*) \leq f(\vec{x} - \frac{\nabla f(\vec{x})}{L})$$

$$f(\vec{x} - \frac{\nabla f(\vec{x})}{L}) \leq f(\vec{x}) + \nabla f(\vec{x})^T\left(-\frac{\nabla f(\vec{x})^T}{L}\right) + \frac{L}{2}\|-\frac{\nabla f(\vec{x})}{L}\|_2^2$$

$$= f(\vec{x}) - \frac{1}{L}\|\nabla f(\vec{x})\|_2^2 + \frac{1}{2L}\|\nabla f(\vec{x})\|_2^2$$

$$= f(\vec{x}) - \frac{1}{2L}\|\nabla f(\vec{x})\|_2^2$$

$$f(\vec{x}^*) \leq f(\vec{x}) - \frac{1}{2L}\|\nabla f(\vec{x})\|_2^2$$

$$\Rightarrow \|\nabla f(\vec{x})\|_2^2 \leq 2L(f(\vec{x}) - f(\vec{x}_*))$$

- Rewrite μ-strong convexity ($\vec{y} = \vec{x}_*$):

Lemma ②:

$$f(\vec{x}_*) \geq f(\vec{x}) + \nabla f(\vec{x})^T(\vec{x}_* - \vec{x}) + \frac{\mu}{2}\|\vec{x}_* - \vec{x}\|_2^2$$

$$f(\vec{x}_*) - f(\vec{x}) - \frac{\mu}{2}\|\vec{x}_* - \vec{x}\|_2^2 \geq \nabla f(\vec{x})^T(\vec{x}_* - \vec{x})$$

$$-f(\vec{x}_*) + f(\vec{x}) + \frac{\mu}{2}\|\vec{x}_* - \vec{x}\|_2^2 \leq \nabla f(\vec{x})^T(\vec{x} - \vec{x}_*)$$

Proof (Main thm):

↳ want to say smth about distance at time (t+1) from our optimum → $\|\vec{x}_{t+1} - \vec{x}_*\|_2^2$ ← want to relate this to previous timestep

$$\vec{x}_{t+1} = \vec{x}_t + \eta \nabla f(\vec{x}_t)$$

$$\|\vec{x}_{t+1} - \vec{x}_*\|_2^2 = \|\vec{x}_t + \eta \nabla f(\vec{x}_t) - \vec{x}_*\|_2^2$$

$$= \| \vec{x}_t - \vec{x}_* + \eta \nabla F(\vec{x}_t) \|_2^2$$

$$= \| \vec{x}_t - \vec{x}_* \|_2^2 + \eta^2 \| \nabla F(x_t) \|_2^2 - 2\eta \langle \nabla F(\vec{x}_t), (\vec{x}_t - \vec{x}_*) \rangle$$

$$= \| \vec{x}_t - \vec{x}_* \|_2^2 + \eta^2 \| \nabla F(x_t) \|_2^2 - 2\eta \nabla F(\vec{x}_t)^T (\vec{x}_t - \vec{x}_*)$$

↳ want to bound this (say it's ≤ something)

Lemma 1 !

$$\leq \| \vec{x}_t - \vec{x}_* \|_2^2 + \eta^2 \cdot 2L(F(\vec{x}_t) - F(\vec{x}_*)) - 2\eta(F(\vec{x}_t) - F(\vec{x}_*) + \frac{\mu}{2} \| \vec{x}_t - \vec{x}_* \|_2^2)$$

$$= (1 - \eta\mu) \| \vec{x}_t - \vec{x}_* \|_2^2 + (2\eta^2 L - 2\eta)(F(\vec{x}_t) - F(\vec{x}_*))$$

↳ choose η to make this term disappear

$$\eta = \frac{1}{L}$$

$$\leq (1 - \frac{\mu}{L}) \| \vec{x}_t - \vec{x}_* \|_2^2$$

recursing

$$\rightarrow \| \vec{x}_{t+1} - \vec{x}_* \|_2^2 \leq (1 - \frac{\mu}{L})^{t+1} \| \vec{x}_0 - \vec{x}_* \|_2^2$$