Lecture 13 Basic Optimization

Friday, October 14, 2016 1:20 PM

Goal: $\min f(x) \leftarrow$ objective

$x \in B \leftarrow$ constraint

usually $B \subseteq \mathbb{R}^n$, $f: \mathbb{R}^n \to \mathbb{R}$

- Convexity
  - idea: if x, y are valid, $\alpha x + (1-\alpha) y$ is also valid ($\alpha \in [0,1]$)
  - set B is convex, if $\forall x, y \in B$
    $\alpha x + (1-\alpha) y \in B$ ($\alpha \in [0,1]$)
  - function $f(x)$ is convex, if $\forall x, y \in B$
    $f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$
    ($\alpha \in [0,1]$)
  - convex optimization: both $f(x)$ and B are convex

- Basic Algorithm: Gradient Descent
  - recap: Gradient $\nabla f(x) \in \mathbb{R}^n$, 1st order derivative

$$(\nabla f(x))_i = \frac{\partial}{\partial x_i} f(x)$$

Hessian: 2nd order derivative $\nabla^2 f(x) \in \mathbb{R}^{n\times n}$

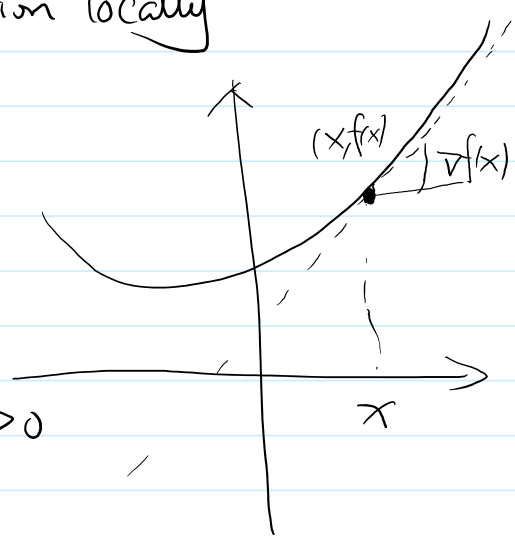
$$(\nabla^2 f(x))_{i,j} = \frac{\partial^2}{\partial x_i \partial x_j} f(x)$$

- General idea in optimization

  approximate the objective function locally

- if $f(x)$ is convex, then

  $f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle$



→ intuition: if $\langle \nabla f(x), y-x \rangle > 0$

$f(y)$ is always worse.

needs an upperbound to guarantee decrease.

- Lipschitz Gradient / "Smoothness"

Def: $f(x)$ is L-smooth (L-Lipschitz Gradient)

if $\forall x, y$ $\quad \|\nabla f(x) - \nabla f(y)\| \leq L \|x-y\|$

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

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

- Analyzing Gradient descent

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

solution: $y = x - \frac{1}{L}\nabla f(x)$

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

(need to show $\|\nabla f(x)\|$ large to make progress)

$$x^{k+1} = x^k - \eta \nabla f(x^k) \qquad (\eta \in (0, \tfrac{2}{L}])$$

Let $r_k = \|x^k - x^*\|$, first show never gets further

$$r_{k+1}^2 = \| x^k - x^* - \eta \nabla f(x^k)\|^2$$
$$= r_k^2 - 2\eta \langle \nabla f(x^k), x^k - x^* \rangle + \eta^2 \|\nabla f(x^k)\|^2$$

$\downarrow (*), \nabla f(x^*) = 0$

$\frac{1}{L}\|\nabla f(x^k)\|^2$

$$\le r_k^2 - \eta\left(\frac{2}{L} - \eta\right)\|\nabla f(x^k)\|^2 \le r_k^2$$

$\Longrightarrow$ always move closer!

Let $\Delta_k = f(x^k) - f(x^*)$, then

$$\underset{\text{convexity}}{\Delta_k \le} \langle \nabla f(x^k), x^k - x^* \rangle \le r_k \|\nabla f(x^k)\|$$
$$\le r_0 \|\nabla f(x^k)\|$$

$$\Delta_{k+1} \le \Delta_k - \underbrace{\eta\left(1 - \frac{L\eta}{2}\right)}_{\omega} \|\nabla f(x^k)\|^2 \le \Delta_k - \frac{\omega}{r_0^2}\Delta_k^2$$

$\uparrow$ same as (**)

idea: if $\Delta_k = \frac{r_0^2}{\cdots}$, $\Delta_{k+1} \le \frac{r_0^2}{\cdots}$

$w(t+1)$

do induction carefully,

$$f(x^k) - f(x^*) \le \frac{2L\|x^0 - x^*\|^2}{k+4}. \qquad \square$$

Strong convexity: better lowerbound

Def: f is μ-strongly convex if $\forall x, y$

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

Lemma: If f is L-smooth, μ-strongly convex, then

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

Theorem: Choose $\eta = \frac{2}{\mu + L}$, then

$$\|x^k - x^*\| \le \left(\frac{L/\mu - 1}{L/\mu + 1}\right)^k \|x^0 - x^*\|$$