19 Gradient Descent (I)

Tuesday, March 25, 2025 13:54

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

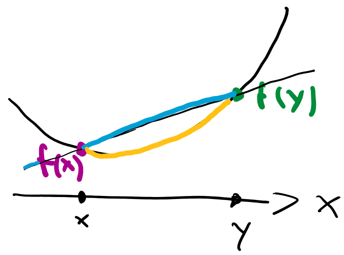
$(\nabla f(x))_i = \frac{d}{dx_i} f(x)$ derivative wrt x_i

$(\nabla^2 f(x))_{ij} = \frac{d}{dx_j}\frac{d}{dx_i} f(x)$

Def: $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is convex if

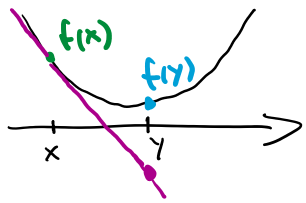
(1) $\forall x,y \in \mathbb{R}^n \ \forall t \in [0,1]$

$\underline{f(t \cdot x + (1-t) \cdot y)} \leq \underline{t \cdot f(x) + (1-t) \cdot f(y)}$



(2) $f(y) \geq f(x) + f'(x) \cdot (y-x)$

$\underline{f(y)} \geq \underline{f(x) + \nabla f(x)^T (y-x)}$



(3) $f''(x) \geq 0$

$\nabla^2 f(x) \succeq 0$ PSD positive semi definite

all eigenvalues are non-negative

\succeq

Problem: Given convex f.

We want to find $\min_x f(x)$

Example: $Ax = b$ $\quad f(x) = \| Ax - b \|$

Gradient Descent

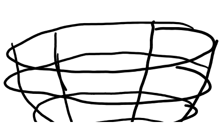
$\nabla f(x) =$ direction of steepest ascend

$-\nabla f(x) =$ direction of steepest desend

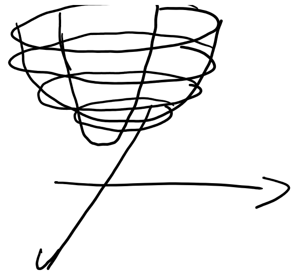
Idea: x^0 start point, repeatedly compute

$x^{t+1} \leftarrow x^t - \eta \cdot \nabla f(x^t)$

$\eta \in \mathbb{R}_{>0}$ stepsize

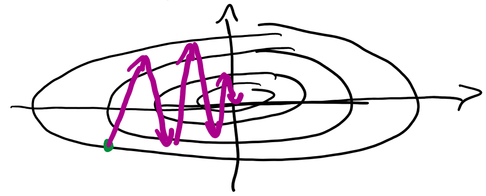


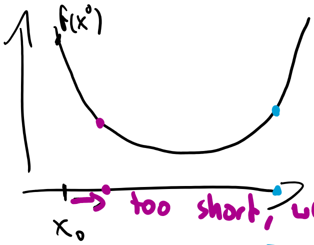
top view



$f: \mathbb{R}^2 \to \mathbb{R}$

top view





too short, we need many steps, algorithm is slow

too long, overshoot optimum

Def: f is L-smooth if $\|\nabla f(x) - \nabla f(y)\| \le L \cdot \|x - y\| \quad \forall x, y$

$L \in \mathbb{R}_{\ge 0}$ $\iff$ $\nabla^2 f(x) \preceq L \cdot I$ all eigenvalues are $\le L$

$|f''(x)| \le L$

$f(x) = 3x^2$

$f'(x) = 6x$

$f''(x) = 6 \implies f$ is 6-smooth

$|f'(x) - f'(y)| = |6x - 6y| = 6 \cdot |x - y|$

Intuition: If Gradient does not change too quickly then we can take longer steps

Theorem: If f is convex and L-smooth and stepsize $\eta \le \frac{1}{L}$

then for $x^{t+1} \leftarrow x^t - \eta \nabla f(x^t)$ for $t = 0, 1, 2, \dots T$

we have $f(x^T) - f(x^*) \le \dfrac{\|x^0 - x^*\|^2}{2 \cdot \eta \cdot T} \le \varepsilon$

optimum $\min_x f(x)$

$T \ge \dfrac{\|x^0 - x^*\|^2}{2\eta\varepsilon}$

Example: $f(x) = \|Ax - b\|^2$ $\quad Ax^* = b$

$f(x^*) = 0$

$\nabla f(x) = 2 \cdot A^T (Ax - b)$

$f(x^*) = 0$

$\nabla f(x) = 2 \cdot A^T (Ax - b)$

while $f(x) > \varepsilon$

$\quad x \leftarrow x - \eta A^T(Ax - b) \cdot 2$

need $\eta \le \frac{1}{L}$

$$\|\nabla f(x) - \nabla f(y)\|$$
$$= \|A^T(Ax-b) - A^T(Ay-b)\|$$
$$= \|A^T A(x-y)\|$$
$$\le \underbrace{\lambda_{max}(A^TA)}_{L} \cdot \|x-y\|$$

largest eigenvalue

$$\# \text{ iterations} \le \frac{\|x^0 - x^*\|^2}{2\varepsilon} \cdot \lambda_{max}(A^TA)$$

$x^* = A^{-1}b \qquad x^0 = 0$

$$\|x^0 - x^*\|^2 = \|A^{-1}b\|^2 \le \lambda_{min}(A) \cdot \|b\|$$

2:48

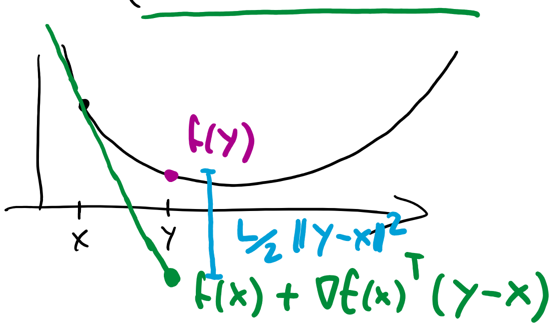
If f is L-smooth then gradient descent converges in $O(\frac{1}{\varepsilon})$

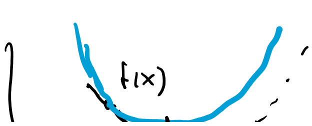
Later: other conditions which guarantee $O(\frac{1}{\varepsilon^2})$ or $O(\log \frac{1}{\varepsilon})$ iterations

---

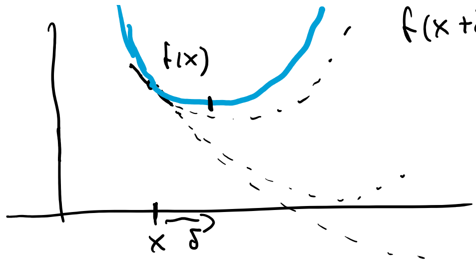
Lemma: If f is L-smooth

$$|f(y) - (f(x) + \nabla f(x)^T(y-x))| \le \frac{L}{2} \cdot \|y-x\|^2$$





$$f(x+\delta) \le f(x) + \nabla f(x)^T \delta + \frac{L}{2}\|\delta\|^2$$



$$f(x+\delta) \le f(x) + \nabla f(x)^T \delta + \frac{L}{2}\|\delta\|^2$$

minimum

Lemma: For stepsize $\eta \le \frac{1}{L}$

$$f(x^t - \eta \nabla f(x^t)) \le f(x^t) - \frac{\eta}{2}\|\nabla f(x^t)\|^2$$

Proof: $f(x^t - \eta \nabla f(x^t)) \le f(x^t) - \nabla f(x^t)^T(\eta \nabla f(x^t)) + \frac{L}{2}\|\eta \nabla f(x^t)\|^2$

1st Lemma for $x = x^t$
$y = x^t - \eta \nabla f(x^t)$

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

$$= f(x^t) - \eta \cdot \|\nabla f(x^t)\|^2 + \frac{L\eta^2}{2}\|\nabla f(x^t)\|^2$$

$$= f(x^t) - \eta \|\nabla f(x^t)\|^2 + L \cdot \eta \cdot \frac{\eta}{2}\|\nabla f(x^t)\|^2$$

$$\le f(x^t) - \frac{\eta}{2}\|\nabla f(x^t)\|^2$$

$$\left| f(y) - \left(f(x) + \nabla f(x)^T(y-x)\right)\right| \le \frac{L}{2}\|y-x\|^2$$

$$\left| f(y) - \left(f(x) + \nabla f(x)^T(y-x)\right)\right|$$

$$= \left| \int_0^1 \nabla f(x + t(y-x))^T (y-x)\, dt - \nabla f(x)^T(y-x) \right|$$

$$\frac{d}{dt} f(x + t\cdot(y-x)) = \nabla f(x + t(y-x))^T (y-x)$$

$$= \left| \int_0^1 \left((\nabla f(x + t(y-x)))^T - \nabla f(x)^T \right)(y-x)\, dt \right|$$

$$\le \int_0^1 \|\nabla f(x + t(y-x)) - \nabla f(x)\| \cdot \|y-x\|\, dt$$

$$\le \int_0^1 \|\nabla f(\underbrace{x+t(y-x)}_{b}) - \nabla f(\underset{\uparrow \atop a}{x})\| \cdot \|y-x\| \, dt$$

$$\le \int_0^1 L \cdot t \cdot \|y-x\| \cdot \|y-x\| \, dt = \frac{L}{2} \|y-x\|^2$$

$\|\nabla f(a) - \nabla f(b)\| \le L \cdot \|a-b\|$ L-smooth