Note Title 25-Jun-19

$$\min_x f(x)$$

① We know the exact form : $f(x) = \frac{1}{2} x^T A x - x^T b$

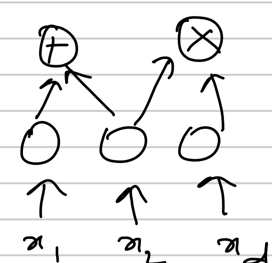
$$x^* = A^{-1} b$$

② $x \xrightarrow[\text{oracle}]{\text{Black box}} f(x)$] Zeroth order oracle

$x \longrightarrow f(x), \nabla f(x)$] First order oracle

$x \longrightarrow f(x), \nabla f(x), \nabla^2 f(x)$] Second order oracle.

Thm: If $f(\cdot)$ is represented by a circuit with simple nodes then the $\nabla f(x)$ can be computed in time linear in the circuit size.



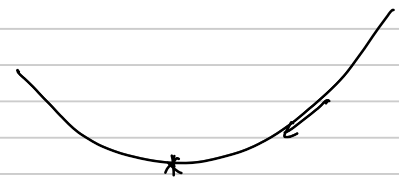
First order oracles can be implemented very efficiently.

We will focus only on first order algorithms.

OUTLINE

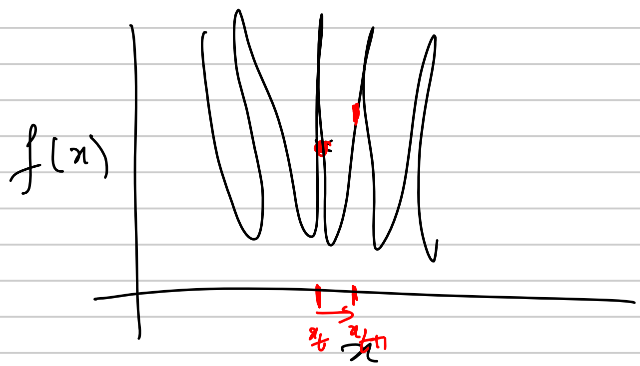
1. CONVEX OPTIMIZATION (DETERMINISTIC)
2. STOCHASTIC CONVEX OPTIMIZATION
3. NONCONVEX OPT. (DET & STOCH.)
4. CONSTRAINED / MINIMAX OPT.

ACCESS TO $x \rightarrow f(x), \nabla f(x)$



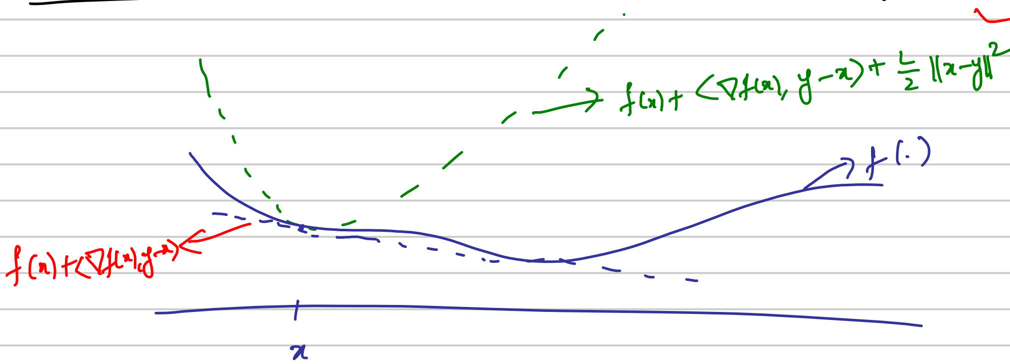
Gradient descent : Start with x_0

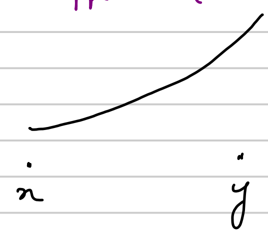
$$x_{t+1} = x_t - \eta \nabla f(x_t).$$



[DEFN.] SMOOTHNESS : A function $f(\cdot)$ is said to be L - Smooth if $\|\nabla f(x) - \nabla f(y)\| \leq L \cdot \|x-y\|$. $\forall x, y$.

LEMMA : If $f(\cdot)$ is L- Smooth then $f(y) \leq \underbrace{f(x) + \langle \nabla f(x), y-x \rangle}_{\text{First order Taylor approximation}} + \underbrace{\frac{L}{2} \|x-y\|^2}_{\text{quadratic upper bound.}} \quad \forall x, y$.





Proof: Given x and y,

$$f(y) = f(x) + \int_{\tau=0}^{1} \langle \nabla f((1-\tau)x + \tau y), y-x \rangle \, d\tau.$$

$$= \boxed{f(x) + \langle \nabla f(x), y-x \rangle} \longrightarrow \text{Linear Taylor expansion}$$

$$+ \underbrace{\int_{\tau=0}^{1} \langle \nabla f((1-\tau)x + \tau y) - \nabla f(x), y-x \rangle \, d\tau}_{S}$$

$$S \leq \int_{\tau=0}^{1} \|\nabla f((1-\tau)x + \tau y) - \nabla f(x)\| \cdot \|y-x\| \, d\tau.$$

$$\overset{\text{(smoothness)}}{\leq} \int_{\tau=0}^{1} L \cdot \| (1-\tau)x + \tau y - x \| \cdot \|y-x\| \, d\tau$$

$$= L \, \|x-y\|^2 \int_{\tau=0}^{1} \tau \, d\tau = \frac{L}{2} \|x-y\|^2$$

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

Thm: If $f(\cdot)$ is L-Smooth and $\boxed{\eta \leq \frac{1}{L}}$ then

$$\min_{t=1,\dots,T} \|\nabla f(x_t)\|^2 \leq 2 \frac{(f(x_0) - f^*)}{\eta T}. \qquad f^* \triangleq \min_x f(x).$$

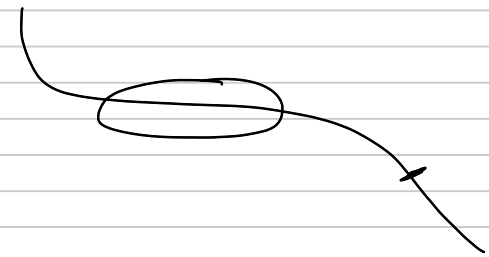
Proof: $f(x_{t+1}) \overset{\text{(smoothness lemma)}}{\leq} f(x_t) + \langle \nabla f(x_t), x_{t+1} - x_t \rangle + \frac{L}{2} \|x_t - x_{t+1}\|^2$

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

$$= f(x_t) - \eta \underbrace{\left(1 - \frac{\eta L}{2}\right)}_{\geq \frac{1}{2}} \|\nabla f(x_t)\|^2$$

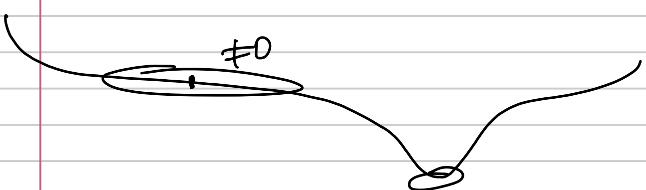
$$\leq f(x_t) - \frac{\eta}{2} \|\nabla f(x_t)\|^2$$

$$f^* \le \underbrace{f(x_{T+1})} \le f(x_0) - \frac{\eta}{2} \sum_{t=0}^{T} \|\nabla f(x_t)\|^2$$



Reorganizing proves the theorem $\square$

Consider a special class of functions where $\nabla f(x) = 0 \Rightarrow x$ is global optimum.



$$\boxed{\mathcal{F}} : \begin{cases} ① \ \nabla f(x) = 0 \Rightarrow x \text{ is global opt. of } f(\cdot) \\ ② \ \text{If } g(x) = \underbrace{f(x) + \langle w, x \rangle} \text{ then } \\ \quad \nabla g(x) = 0 \Rightarrow x \text{ is global opt. of } g(\cdot). \end{cases}$$

Lemma: $\boxed{\mathcal{F}} = \{ f : f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle \ \forall x, y \}$.

Proof: Any $f \in \mathcal{F}$ satisfies

Given x and y; $\underbrace{g(y)} = \underbrace{f(y)} + \underbrace{\langle -\nabla f(x), y - x \rangle}$

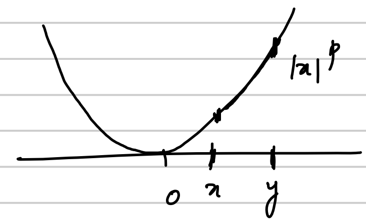
$\nabla g(y) = \nabla f(y) - \nabla f(x) \Rightarrow \nabla g(x) = 0 \Rightarrow x$ is global opt. of $g(\cdot)$.

$\Rightarrow g(x) \le g(y) \quad \forall y$

$\Rightarrow f(x) \le f(y) - \langle \nabla f(x), y - x \rangle$ $\square$

[Defn.] <u>CONVEX FNS</u>: $f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle \quad \forall\, x, y$

<u>Examples</u>: ① $f(x) = |x|^p \qquad p > 1$



$$\frac{df}{dx} = p\, x^{p-1}.$$

$$f(y) = f(x) + \underbrace{(y-x)\int_{\alpha=0}^{1} \frac{df}{dx}\bigg|_{(1-\alpha)x+\alpha y} \cdot d\alpha}$$

$$\frac{df}{dx}\bigg|_{(1-\alpha)x+\alpha y} = p\left[(1-\alpha)x + \alpha y\right]^{p-1} \geqslant p\, x^{p-1}$$

$$f(y) \geqslant f(x) + \underbrace{p\, x^{p-1}} \cdot \underbrace{(y-x)}.$$

② If $f_1(\cdot)$ is convex and $f_2(\cdot)$ is convex then

$f(x) = f_1(x) + f_2(x)$ is also convex.

$$f(y) = f_1(y) + f_2(y)$$

$$\geqslant f_1(x) + \langle \nabla f_1(x), y - x \rangle + f_2(x) + \langle \nabla f_2(x), y - x \rangle$$

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

③ $x = (x_1, x_2)$ $\quad f_1(x) = x_1^2$

$f_2(x) = x_2^2$

So, $f(x) = \|x\|^2 = x_1^2 + x_2^2$ is also convex.

$f(x) = \|x\|_p^p$ is also convex

④ If $f(\cdot)$ is convex then $f(Ax+b)$ is also convex.

Let $g(x) \triangleq f(Ax+b)$

$$g(y) = f(\underbrace{Ay+b}_{w_y}) \overset{f(\cdot) \text{ is convex}}{\geq} f(\underbrace{Ax+b}_{w_x}) + \langle \nabla f|_{Ax+b}, (Ay+b) - (Ax+b) \rangle$$

$$= g(x) + \langle \nabla f|_{Ax+b}, A(y-x) \rangle$$

$$= g(x) + \langle \underbrace{A^T \nabla f|_{Ax+b}}, y - x \rangle$$

$$= g(x) + \langle \nabla g(x), y - x \rangle.$$

$\min_x \frac{1}{2} \|Ax-b\|^2.$

$f(x) = \|Ax-b\|^2$ is convex.

Find x s.t. $a^T x \approx b$

Temp	Humidity	...		Measure Rainfall
a_1	a_2	. a_d		b