Lecture 3 3/10/2022

Definition: $g : \mathbb{R}^N \to \mathbb{R}$ is said to be L-Lipshitz if $\exists L$ s.t.

$\forall x, y \in \mathbb{R}^N$:

$$\| g(x) - g(y) \| \leq L \|x - y\|$$

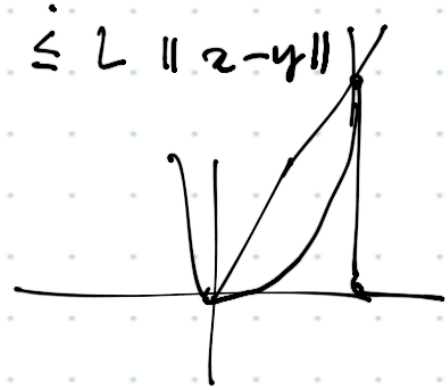
N=1

$g(x) = c$ is constant is Lipshitz with $L = 0$

$g(x) = x^2$ No!

$$\| x^2 - y^2 \| \overset{?}{\leq} L \| x - y \|$$

$y = 0$ $\quad x^2 \leq L |x|$



$g(x) = x$ is Lipshitz

⋮

Convergence of GD

Start from $x^0 \in \mathbb{R}^N$ and do

$$x^{k+1} = x^k - \tau \nabla f(x^k) , \qquad \tau > 0$$

$$f(x^{k+1}) = f(\underset{x}{\underset{\uparrow}{x^k}} - \tau \underset{y}{\underset{\uparrow}{\nabla f(x^k)}})$$

Lemma 1 $\quad f \in C^1(\mathbb{R}^N, \mathbb{R})$, then $\forall\, x, y \in \mathbb{R}^N$

$$f(x - \tau y) = f(x) - \int_0^\tau \nabla f(x - sy) \cdot y \, ds$$

Proof

$$\frac{d}{ds} f(x - sy) = \nabla f(x - sy) \cdot (-y)$$

$$\int_0^\tau \underbrace{\frac{d}{ds} f(x-sy)}\, ds = -\int_0^\tau \nabla f(x-sy) \cdot y \, ds$$

$$f(x - \tau y) - f(x) = -\int_0^\tau \nabla f(x - sy) \cdot y \, ds \qquad \blacksquare$$

Now use Lemma 1 with $x = x^k$ $\quad y = \nabla f(x^k)$

$$f(x^{k+1}) = f(x^k) - \int_0^\tau \nabla f(x^k - s\nabla f(x^k)) \cdot \nabla f(x^k) \, ds$$

$$= f(x^k) - \int_0^\tau \left(\nabla f(x^k - s\nabla f(x^k)) - \nabla f(x^k) + \nabla f(x^k) \right) \cdot \nabla f(x^k) \, ds$$

$$= f(x^k) - \tau \|\nabla f(x^k)\|^2 - \int_0^\tau \left(\nabla f(x^k - s\nabla f(x^k)) - \nabla f(x^k) \right) \cdot \nabla f(x^k) \, ds$$

Lemma 2 $\quad$ If $g : \mathbb{R}^N \to \mathbb{R}^N$ is L-Lipschitz then

$\forall\, x, y \in \mathbb{R}^N \qquad \underline{(g(x) - g(y))} \cdot (x - y) \leq L \|x - y\|^2$

Proof $\quad (g(x)-g(y))\cdot(x-y) \le |(g(x)-g(y))\cdot(x-y)|$

Chauchy - Schwartz $\quad \forall\, a,b \in \mathbb{R}^N \quad |a\cdot b| \le \|a\|\,\|b\|$

$$\le \|g(x)-g(y)\|\,\|x-y\| \le L\,\|x-y\|\,\|x-y\| \qquad \blacksquare$$

Now I want to use Lemma 2 with $x = x^k$

$$y = x^k - s\nabla f(x^k)$$

$$x - y = s\nabla f(x^k)$$

$$f(x^k) - \tau\|\nabla f(x^k)\|^2 + \int_0^\tau \left(-\nabla f(x^k - s\nabla f(x^k)) + \nabla f(x^k)\right) \nabla f(x^k)\,ds$$

$$f(x^k) - \tau\|\nabla f(x^k)\|^2 + \int_0^\tau \frac{1}{s}\underbrace{\left(\nabla f(x^k) - \nabla f(x^k - s\nabla f(x^k))\right)}_{\nabla f(x) - \nabla f(y)} \cdot \underbrace{\left(s\nabla f(x^k)\right)}_{(x-y)} ds$$

$$\int_0^\tau s\,ds$$

$$\le f(x^k) - \tau\|\nabla f(x^k)\|^2 + \int_0^\tau \frac{1}{\cancel{s}} L\, s^{\cancel{2}}\|\nabla f(x^k)\|^2\,ds$$

$$= f(x^k) - \tau\|\nabla f(x^k)\|^2 + \frac{L\tau^2}{2}\|\nabla f(x^k)\|^2$$

$$\frac{\tau^2}{2} = \int_0^\tau s\,ds$$

$$= f(x^k) - \tau\left(1 - \frac{L\tau}{2}\right)\|\nabla f(x^k)\|^2$$

We found that

$$f(x^{k+1}) \leq f(x^k) - \tau \underbrace{\left(1 - \frac{L\tau}{2}\right)}_{C>0} \|\nabla f(x^k)\|^2$$

We choose $1 - \frac{L\tau}{2} > 0$ $\quad \boxed{\tau < \frac{2}{L}}$

$$f(x^{k+1}) \leq f(x^k) - \tau C_1 \|\nabla f(x^k)\|^2$$

$$\leq f(x^{k-1}) - \tau C_1 \|\nabla f(x^{k-1})\|^2 - \tau C \|\nabla f(x^k)\|^2$$

$$\leq \cdots \leq f(x^0) - \tau C \sum_{j=0}^{k} \|\nabla f(x^j)\|^2$$

$$f(x^n) \leq f(x^0) - \tau C \sum_{k=0}^{n-1} \|\nabla f(x^k)\|^2$$

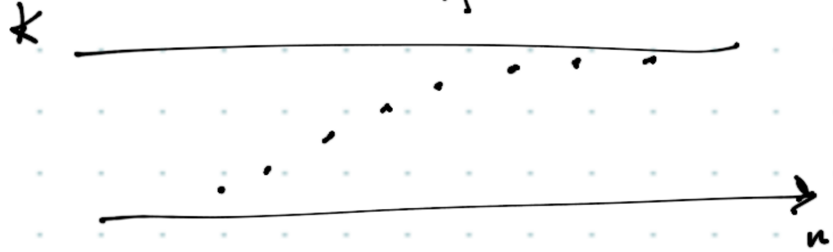
$\boxed{\inf f > -\infty}$ $\quad$ f should be bounded from below!

$$\inf f \leq f(x^n) \leq f(x^0) - \tau C \sum_{k=0}^{n-1} \|\nabla f(x^k)\|^2$$

$$\underbrace{\sum_{k=0}^{n-1} \|\nabla f(x^k)\|^2}_{/\!/} \leq \frac{f(x^0) - \inf f}{\tau C_1} = K$$

$$\begin{cases} S_{n-1} \leq K & \forall n \geq 1 \\ S_{n-1} \leq S_n \end{cases}$$

So $(S_n)_{n \geq 1}$ is an increasing sequence bounded from above



Bolzano Weirstrass theorem $\Rightarrow$ $S_n \longrightarrow S$

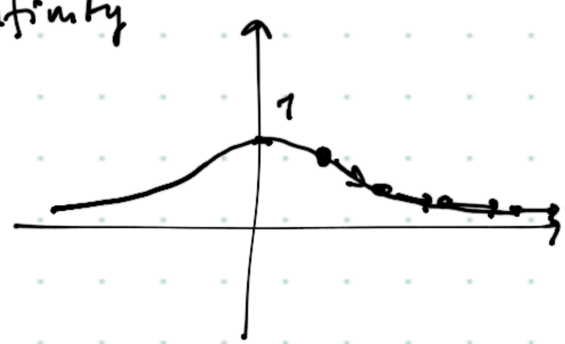
Which in our case means that

$$\sum_{k=0}^{+\infty} \|\nabla f(x^k)\|^2 \quad \text{is convergent.}$$

Then $\|\nabla f(x^k)\|^2 \longrightarrow 0$ as $k \to +\infty$ $\square$

$$x^k \to x^* \in \operatorname{argmin} f$$

Coercivity = "ifinity at infinity"



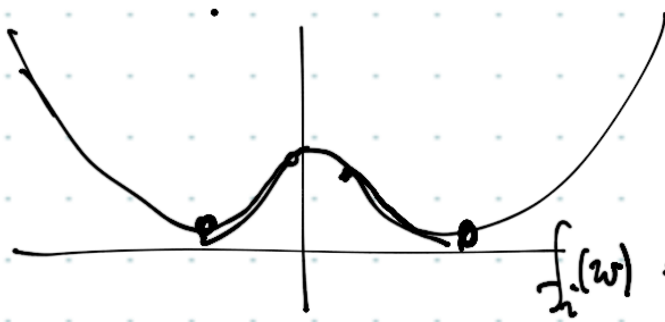
Example

$$f(x) = \frac{1}{1+x^2}$$

$\|\nabla f(x^k)\| \longrightarrow 0$

If you want to stop to local minima $\rightarrow$ convexify f

$$f(x) \longrightarrow f(x) + \frac{x^2}{2}$$



$$f_i(w) = \ell\left(F(x_i, w), y_i\right)$$

Convergence of SGD

$$X^{k+1} = X^k - \tau_k \nabla f_{i_k}(X^k) \qquad \tau_k > 0 \quad \forall k \geq 0$$

$$f_j(X^{k+1}) = f_j\left(X^k - \tau_k \nabla f_{i_k}(X^k)\right)$$

Now I want to use Lemma 1

$$= f_j(X^k) - \int_0^{\tau_k} \nabla f_j\left(X^k - s\nabla f_{i_k}(X^k)\right) \cdot \nabla f_{i_k}(X^k)\, ds$$

$$= f_j(X^k) - \int_0^{\tau_k} \left[\nabla f_j\left(X^k - s\nabla f_{i_k}(X^k)\right) - \nabla f_j(X^k) + \nabla f_j(X^k)\right] \cdot \nabla f_{i_k}(X^k)\, ds$$

$$= f_j(X^k) - \tau_k \nabla f_j(X^k) \cdot \nabla f_{i_k}(X^k) + \int_0^{\tau_k} \frac{1}{s}\left(-\nabla f_j\left(X^k - s\nabla f_{i_k}(X^k)\right) + \nabla f_j(X^k)\right) \cdot \left(s\nabla f_{i_k}\right)$$

$$\leq f_j(x^k) - \tau_k \underbrace{\nabla f_j(x^k) \cdot \nabla f_{i_k}(x^k) + \frac{L_j \tau_k^2}{2} \|\nabla f_{i_k}(x^k)\|^2}$$

$$g = \nabla f_j \qquad x = x^k$$

$$y = x^k - s\nabla f_{i_k}(x^k)$$

$$x - y = s\nabla f_{i_k}(x^k)$$

$$f_j(x^{k+1}) \leq f_j(x^k) - \tau_k \nabla f_j(x^k) \cdot \nabla f_{i_k}(x^k) + \frac{L_j \tau_k^2}{2} \|\nabla f_{i_k}(x^k)\|^2$$

$$\frac{1}{n}\sum_{j=1}^{n} \left(\quad \right) \leq \frac{1}{n}\sum_{j=1}^{n} \left(\quad \right)$$

$$f(x^{k+1}) \leq f(x^k) - \tau_k \nabla f(x^k) \cdot \nabla f_{i_k}(x^k) + \frac{\bar{L}\tau_k^2}{2} \|\nabla f_{i_k}(x^k)\|^2$$

where $\bar{L} = \frac{1}{n}\sum_{j=1}^{n} L_j$

$$E\left(f(x^{k+1}) \mid x^k\right) \leq f(x^k) - \tau_k \nabla f(x^k) \cdot \frac{1}{n}\sum_{i=1}^{n} \nabla f_i(x^k) + \frac{\bar{L}\tau_k^2}{2}\left(\frac{1}{n}\sum_{i=1}^{n} \|\nabla f_i(x^k)\|^2\right)$$

$$= f(x^k) - \tau_k \|\nabla f(x^k)\|^2 + \frac{\bar{L}\tau_k^2}{2}\left(\underline{\frac{1}{n}\sum_{i=1}^{n}\|\nabla f_i(x^k)\|^2}\right)$$

$$= f(x^k) - \tau_k \|\nabla f(x^k)\|^2 + \frac{\bar{L}\tau_k^2}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\right.$$

$$\frac{1}{n}\sum_{i=1}^{n}\|\nabla f_i(x^k) - \nabla f(x^k)\|^2 = \frac{1}{n}\sum_{i=1}^{n}\Big(\|\nabla f_i(x^k)\|^2 - 2\nabla f_i(x^k)\cdot\nabla f(x^k) + \|\nabla f(x^k)\|^2\Big)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\|\nabla f_i(x^k)\|^2 - 2\underbrace{\frac{1}{n}\sum_{i=1}^{n}\nabla f_i(x^k)}_{=\nabla f(x^k)}\cdot\nabla f(x^k) + \|\nabla f(x^k)\|^2$$

$$= \boxed{\frac{1}{n}\sum_{i=1}^{n}\|\nabla f_i(x^k)\|^2} - \|\nabla f(x^k)\|^2 = \frac{1}{n}\sum_{i=1}^{n}\|\nabla f_i(x^k) - \nabla f(x^k)\|^2$$

$$= f(x^k) - \tau_k \|\nabla f(x^k)\|^2 + \frac{\bar{L}\tau_k^2}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\|\nabla f_i(x^k) - \nabla f(x^k)\|^2\right) + \frac{\tau_k^2\bar{L}}{2}\|\nabla f(x^k)\|^2$$

$$E\big(f(x^{k+1}) \mid x^k\big) \le f(x^k) - \tau_k\left(1 - \frac{\tau_k\bar{L}}{2}\right)\|\nabla f(x^k)\|^2$$

$$+ \frac{\bar{L}\tau_k^2}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\|\nabla f_i(x^k) - \nabla f(x^k)\|^2\right)$$

$$E\left(E\left(f(x^{k+1}) \mid x^k\right)\right) = E f(x^{k+1})$$

$$E\left(E(X \mid Y)\right) = E X$$

$$E f(x^{k+1}) \le E f(x^k) - \tau_k\left(1 - \frac{\tau_k \bar{L}}{2}\right) E\|\nabla f(x^k)\|^2$$

$$+ \frac{\bar{L}\tau_k^2}{2} E\left(\frac{1}{n}\sum_{i=1}^{n}\|\nabla f_i(x^k) - \nabla f(x^k)\|^2\right)$$

Assumption:

$$\frac{1}{n}\sum_{i=1}^{n}\|\nabla f_i(x) - \nabla f(x)\|^2 \le \sigma^2 \quad \forall x \in \mathbb{R}^N$$

$$\boxed{1 - \frac{\tau_k \bar{L}}{2} > 0 \qquad \tau_k < \frac{2}{\bar{L}}}$$

$$E f(x^{k+1}) \leq E f(x^k) - \tau_k \left(1 - \frac{\tau_k \bar{L}}{2}\right) E \|\nabla f(x^k)\|^2$$

$$+ \frac{\bar{L} \tau_k^2}{2} \sigma^2$$

$$E f(x^n) \leq f(x^0) \underset{\uparrow}{-} \sum_{k=1}^{n-1} \tau_k \left(1 - \frac{\tau_k \bar{L}}{2}\right) E \|\nabla f(x^k)\|^2 + \frac{\bar{L} \tau_k^2}{2} \sigma^2$$

If I chose

$$1 - \frac{\tau_k \bar{L}}{2} \leq \frac{1}{2}$$

$$E f(x^m) \leq f(x^0) - \frac{1}{2} \sum_{k=0}^{m-1} \tau_k \, E \|\nabla f(x^k)\|^2 + \sum_{k=0}^{m-1} \frac{\bar{L} \tau_k^2}{2} \sigma^2$$

$$-\left(1 - \frac{\tau_k \bar{L}}{2}\right) \leq -\frac{1}{2}$$

$$1 - \frac{\tau_k \bar{L}}{2} \geq \frac{1}{2} \qquad \tau_k < \frac{2}{L}$$

$$-\frac{\tau_k \bar{L}}{2} \geq -\frac{1}{2} \qquad \tau_k \leq \frac{1}{L}$$

Choose $\sum_{k=0}^{+\infty} \tau_k^2 < +\infty$

$$\inf f \leq \mathbb{E} f(x^n) \leq f(x^0) - \frac{1}{2} \sum_{k=0}^{m-1} \tau_k \|\nabla f(x^k)\|^2 + \frac{L\sigma^2}{2} \sum_{k=0}^{+\infty} \tau_k^2$$

$$\frac{1}{2} \sum_{k=0}^{m-1} \tau_k \|\nabla f(x^k)\|^2 \leq L\sigma^2/2 \sum_{k=0}^{+\infty} \tau_k^2 + f(x^0) - \inf f$$

$\Rightarrow \sum_{k=0}^{+\infty} \tau_k \|\nabla f(x^k)\|^2$ converges

If $\sum_{k=0}^{+\infty} \tau_k$ converge $\|\nabla f(x^k)\|^2 \to$ costant it ok

Choose $\sum_{k=0}^{+\infty} \tau_k = +\infty$ $\qquad \tau_k = \frac{1}{k}$

$\Rightarrow \exists$ subsequence k_n such that

$$\|\nabla f(x^{k_n})\|^2 \to 0 \quad n \to +\infty$$

$$\|\nabla f(x^k)\| \to 0 \quad \text{a.s.}$$

$\blacksquare$

$$x^{k+1} = x^k - \tau g^k$$

$$x^{k+1} \in \underset{s \in \mathbb{R}^N}{\operatorname{argmin}} \; \underbrace{f(x^k) + g^k \cdot (s - x^k)}_{\text{I order approx of } f \text{ around } x^k} + \frac{1}{2\tau}\|s - x^k\|^2$$

$$\nabla_s \left(f(x^k) + g^k \cdot (s - x^k) + \frac{1}{2\tau}\|s - x^k\|^2 \right)\Big|_{s = x^{k+1}} = 0$$

$$g^k + \frac{1}{\tau}\left(x^{k+1} - x^k\right) = 0$$

$$\boxed{x^{k+1} = x^k - \tau g^k}$$

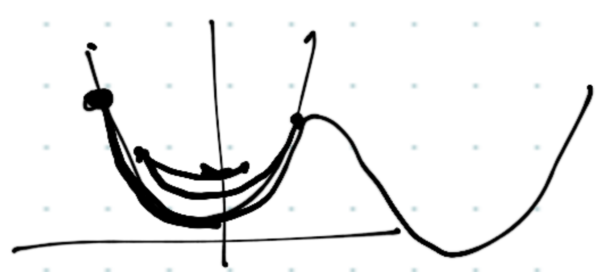
$\alpha = \frac{1}{10}$ $\quad a_0, a_1, a_2, a_3$ $\quad 1$

$$\langle a \rangle_1^{1/10} = \left(1 - \frac{1}{10}\right) \sum_{k=0}^{1} \left(\frac{1}{10}\right)^{1-k} a_k$$

$$= \frac{9}{10}\left(\frac{1}{10} a_0 + a_1 \right) =$$

$$g^k + \frac{1}{t}\left(x^{k+1} - x^k\right) + \frac{1}{\mu}\left(x^{k+1} - 2x^k + x^{k-1}\right) = 0$$

$$m\ddot{x} + \gamma \dot{x} + x = 0$$



$\beta = 0.9$