Last class:

$$\begin{cases} f \text{ is } L\text{-Lip}, \\ \|x_0 - x^*\| \le R \\ f \text{ convex, diff.} \end{cases} \Rightarrow f\left(\frac{1}{T}\sum_{s=0}^{T-1} x_s\right) - f(x^*) \le \frac{RL}{\sqrt{T}}$$

when $\mu = \dfrac{R}{L\sqrt{T}}$

Def: A fctn $f : \mathbb{R}^n \to \mathbb{R}$ is L-smooth if gradient is L-Lip,

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$$

Thm: If f is L-smooth + twice diff, then

$$\underbrace{v^T \nabla^2 f(x)\, v}_{\|\nabla^2 f(x)\|_2} \leq L, \quad \forall x \in \Omega, \; v \in \mathbb{R}^n$$

$\|v\| = 1$

Thm: If $f: \mathbb{R}^n \to \mathbb{R}$ is L-smooth & convex, & $0 \leq \mu \leq \frac{1}{L}$, then GD satisfies

$$f(x^{(t)}) - f(x^*) \leq \frac{1}{2t\mu} \| x^{(0)} - x^* \|^2$$

Remark: $\|x^{(0)} - x^*\| \leq R^2$, $\mu = 1/L$

$$\text{RHS} = \frac{R^2 L}{2t}$$

Ex: $\|x^{(0)} - x^*\| \le 10$

$L = 2$ $\Rightarrow$ $t = 10{,}000$

$u = 1/2$

$f(x^t) - f(x^*) \le \frac{1}{1000} = .001$

Pf strategy comes from

$$f(y) \le f(x) + \nabla f(x)^T (y-x) + \underbrace{\frac{1}{2}(y-x)^T (L I)(y-x)}_{\frac{L}{2}\|y-x\|^2}$$

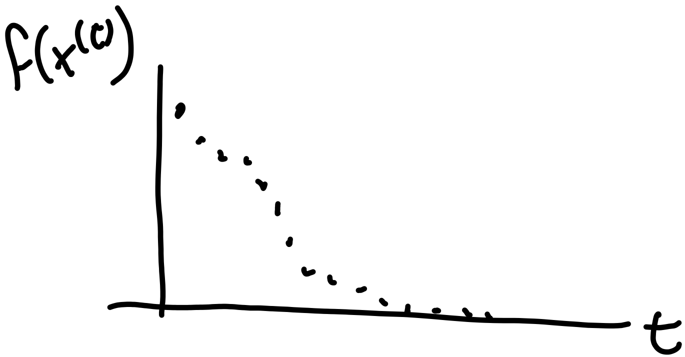
Thm: Let $f: \mathbb{R}^n \to \mathbb{R}$ L-smooth. For any $0 \le \mu \le \frac{1}{L}$, each step of GD gives

$$f(x^{(t+1)}) \le f(x^{(t)}) - \frac{\mu}{2} \| \nabla f(x^{(t)}) \|^2$$

Pf:
$$f(x^{(t+1)}) \le f(x^{(t)}) + \nabla f(x^{(t)})^T \underline{(x^{(t+1)} - x^{(t)})} + \frac{L}{2} \| x^{(t+1)} - x^{(t)} \|^2$$

GD eq: $X^{(t+1)} = x^{(t)} - \mu \nabla f(x^{(t)})$

$$
\begin{aligned}
f(x^{(t+1)}) &\leq f(x^{(t)}) - \mu \|\nabla f(x^{(t)})\|^2 \\
&\qquad + \frac{L}{2}\mu^2 \|\nabla f(x^{(t)})\|^2 \\
&= f(x^{(t)}) - \mu \underbrace{\left(1 - \frac{L}{2}\mu\right)}_{\mu \leq 1/L \;\Rightarrow\; \geq 1/2} \|\nabla f(x^{(t)})\|^2 \\
&\leq f(x^{(t)}) - \frac{\mu}{2} \|\nabla f(x^{(t)})\|^2 \qquad \square
\end{aligned}
$$



Thm: $f: \mathbb{R}^n \to \mathbb{R}$ L-smooth, $0 \le \mu \le 1/L$, GD for T iters. at least one x_t must satisfy

$$\|\nabla f(x^{(t)})\| \le \sqrt{\frac{2\left(f(x^{(0)}) - f(x^*)\right)}{\mu T}}$$

Pf: Assume f is lower bounded.

$$f(x^{(T)}) - f(x^{(0)}) = \sum_{t=0}^{T-1}\left(f(x^{(t+1)}) - f(x^{(t)})\right)$$

$$\leq -\sum_{t=0}^{T-1} \frac{\mu}{2} \|\nabla f(x^{(t)})\|^2$$

$$\sum_{t=0}^{T-1} \|\nabla f(x^{(t)})\|^2 \leq \frac{2}{\mu}\left(f(x^{(0)}) - f(x^*)\right)$$

$f(x^{(T)}) \geq f(x^*)$

Exists at least one $x^{(t)}$

s.t.

$$\|\nabla f(x^{(t)})\|^2 \leq \frac{2}{\mu T}\left(f(x^{(0)}) - f(x^*)\right)$$

(aka one term must be less than avg)

$$\|\nabla f(x^{(t)})\| \le \sqrt{2\,\frac{(f(x^{(0)}) - f(x^*))}{\mu T}} \qquad \square$$

Remark: Last two theorems, never assumed convexity!