DESCENT LEMMA

① $f: \mathbb{R}^n \to \mathbb{R}$ cont diff

② L-Lipschitz gradients, $\forall x, y \in \mathbb{R}^n$

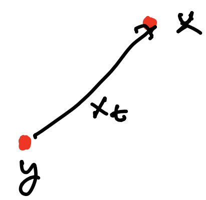
$$\|\nabla f(x) - \nabla f(y)\|_2 \le L\|x-y\|_2$$

WTS

$$f(x) \le f(y) + \nabla f(y)^T(x-y) + \frac{L}{2}\|x-y\|_2^2$$

Consider

$$x_t = tx + (1-t)y \quad \text{for} \quad t \in [0,1]$$



$$f(x) - f(y) = \int_0^1 \frac{d\, f(x_t)}{dt}$$

$$= \int_0^1 \nabla f(x_t)^T(x-y)$$

$$= \nabla f(y)^T(x-y) + \int_0^1 (\nabla f(x_t) - \nabla f(y))^T(x-y)$$

$$\overset{\text{C.S.}}{\le} \nabla f(y)^T(x-y) + \int_0^1 \|\nabla f(x_t) - \nabla f(y)\|\,\|x-y\|$$

$$\overset{\text{Lipschitz}}{\le} \nabla f(y)^T(x-y) + \int_0^1 Lt\|x-y\|^2$$

$$= \nabla f(y)^T(x-y) + L\|x-y\|^2 \int_0^1 t$$

$\square$