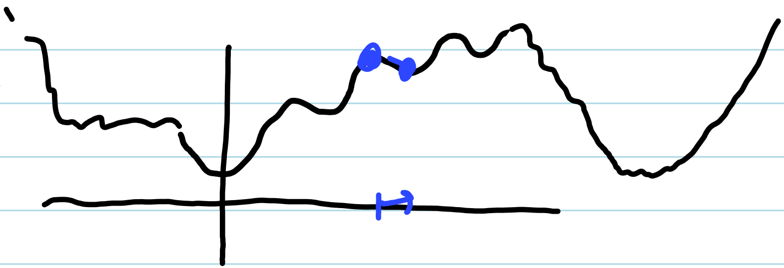
Non convex cost functions

$\min_x f(x)$ ← need not be convex fctn

Ex:



Changes: - need not be one x^*

- descent direction not necessarily good
- stuck at local mins.

Can't prove nonconvex opt works in general

Assumption: Polyak-Lojasiewicz Condition

PL-condition

Def: $f: \mathbb{R}^n \to \mathbb{R}$ satisfies μ-PL-cond. if $\forall x \in \mathbb{R}^n$

$$\frac{1}{2} \|\nabla f(x)\|_2^2 \geq \mu \left(f(x) - f(x^*) \right)$$

where x^* is global min.

Interp: when far from x^*, gradient is big.

Ex. $f(x) = x^2 + 3\sin^2 x$

$f'(x) = 2x + 6\sin(x)\cos(x)$

$$\left[\begin{array}{l} x^* = 0 \\ f(x^*) = 0 \end{array}\right.$$

LHS: $\frac{1}{2}\|\nabla f(x)\|_2^2 = \frac{1}{2}\left(2x + 6\sin(x)\cos(x)\right)^2$

$= 2x^2 + 12x\sin(x)\cos(x)$

$+ 18\sin^2(x)\cos^2(x)$

RHS: $\mu\left(f(x) - f(x^*)\right) = \mu\left(x^2 + 3\sin^2 x\right)$

Choice of $\mu = \frac{1}{32}$

$$g(x) = LHS - RHS = 16\left(2x + 6\sin(x)\cos(x)\right)^2 - \left(x^2 + 3\sin^2(x)\right)$$

Fact: $\sin^2(x) \leq x^2$

$$g(x) \geq 16\left(2x + 6\sin(x)\cos(x)\right)^2 - 4x^2$$

$$= 64x^2\Big(1 + 3\underbrace{\frac{\sin(x)\cos(x)}{x}}_{\frac{\sin(2x)}{x}}\Big)^2 - 4x^2$$

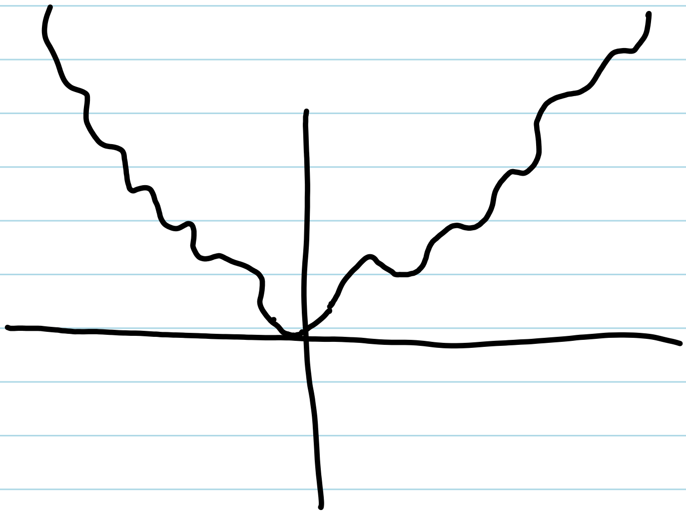
Cases: 1) $|x| \in \frac{\pi}{2}$: $\frac{\sin(2x)}{x} \geq 0 \Rightarrow$

$$g(x) \geq 64x^2 - 4x^2 \geq 0$$

2) $|x| \geq \frac{\pi}{2}$: $\frac{\sin(2x)}{x} \geq -1/4$

$g(x) \geq 64\left(1 - \frac{3}{4}\right)^2 - 4x^2 \geq 0$

$\square$



assume A rank m.

Ex: $A \in \mathbb{R}^{m \times n}$ $\qquad m < n$

$\boxed{f(x^*) = 0}$

$$f(x) = \|Ax - b\|_2^2$$

$$= x^T \underbrace{(A^T A)}_{\substack{n \times n \text{ but rank at } m.}} x - 2b^T A x + b^T b$$

In this setting, f is convex but not strongly convex

$(\lambda_{min} = 0)$

Show PL:

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

LHS: $\frac{1}{2} \|\nabla f(x)\|^2 = \frac{1}{2} \| A^T A x - 2 A^T b \|^2$

RHS: $\mu (f(x) - f(x^*)) = \mu \| Ax - b \|^2$

Notice $g(x) = \|x\|^2$ is strongly convex

$$g(y) \geq g(x) + \nabla g(x)^T (y - x) + \frac{1}{2} \| y - x \|^2$$

$$g(Ay) \geq g(Ax) + \underbrace{\nabla g(Ax)^T (Ay - Ax)}_{\substack{\langle A^T \nabla g(Ax),\, y-x \rangle \\ \langle \nabla f(x),\, y-x \rangle}} + \frac{1}{2} \| A(y - x) \|^2$$

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{1}{2} \| A(y-x) \|_2^2$$

Define x_p · orth. proj. of x onto the soln set of X^*

$$f(x_p) \geq f(x) + \nabla f(x)^T (x_p - x) + \frac{1}{2} \| A(x_p - x) \|^2$$

$$\geq f(x) + \nabla f(x)^T (x_p - x) + \frac{1}{2} \sigma_{min} \| x_p - x \|^2$$

$$\geq f(x) + \min_y \left(\nabla f(x)^T (y-x) + \frac{1}{2} \sigma_{min} \| y - x \|^2 \right)$$

$$\geq f(x) - \frac{1}{2\sigma_{min}} \| \nabla f(x) \|^2$$

$\square$

Thm: If f is L-smooth & μ-PL, then

GD w/ $\eta = 1/L$ converges at rate

step size

$$f(x^{(k)}) - f(x^*) \leq \left(1 - \frac{\mu}{L}\right)^k \left(f(x^{(0)}) - f(x^*)\right)$$

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

$$\leq f(x^{(t)}) - \frac{\mu}{L}\left(f(x^{(t)}) - f(x^*)\right)$$

$$f(x^{(t+1)}) - f(x^*) \leq \left(1 - \frac{\mu}{L}\right)\left(f(x^{(t)}) - f(x^*)\right)$$

$\square$

Takeaways: 1) PL-cond can hold for non convex fctns & for non strongly convex fctn.

- showed two examples (third next class)

2) If PL & L-smooth, get conv rate

$$\left(1 - \frac{\mu}{L}\right)^K$$