Unconstrained minimization:

$$\min_x f(x)$$

f continuous & diff. (for now)

Note: $\rightarrow$ max $\Rightarrow$ take $g(x) = -f(x)$

$\rightarrow$ constraints $\Rightarrow$ take

$$g(x) = f(x) + \psi(x)$$

$$\psi(x) = \begin{cases} 0 & x \in S \\ +\infty & x \notin S \end{cases}$$

How to solve it?

$x^0, x^1, x^2, \dots$ $\qquad \forall_t \quad f(x^{t+1}) < f(x^t)$

x^t

Taylor approx.:

$y = x^t + \delta$

$f: \mathbb{R}^m \rightarrow \mathbb{R}$

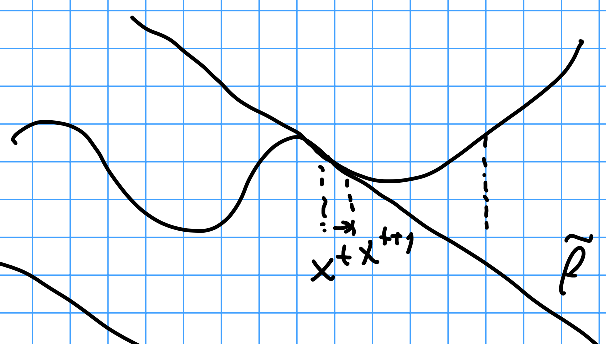
$$f(y) = \underbrace{f(x^t) + \nabla f(x^t)^T \delta}_{\tilde{f}(y)} + \underbrace{\delta^T \nabla^2 f(x^t) \delta + \dots}$$

$\overset{??}{\delta} \quad \leftrightarrow \quad \overset{??}{\delta^2}$

Best move to minimize $\tilde{f}$

$$x^{t+1} \Leftarrow x^t - \eta \nabla f(x^t)$$

η ← step size



β-smoothness

$$\forall_{x,y} \quad \|\nabla f(x) - \nabla f(y)\| \leq \beta \|x - y\|$$

$$\Leftrightarrow \quad \forall_{\delta, x} \quad \delta^T \nabla^2 f(x) \delta \leq \beta \|\delta\|^2 \quad \Leftrightarrow$$

Largest eigenvalue of $\nabla^2 f(x) \leq \beta$

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \qquad \nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \end{bmatrix} \qquad \nabla^2 f = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & & \\ & \frac{\partial^2 f(x)}{\partial x_2^2} & \\ & & \ddots \end{bmatrix}$$

Taylor approx. (II)

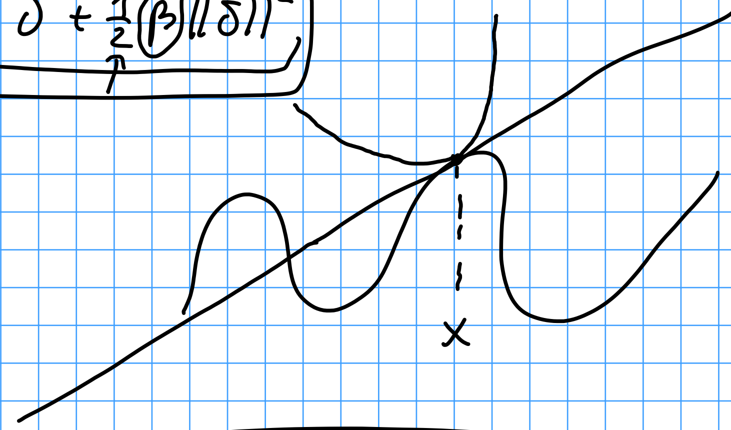
$$\delta + x = y$$

$$f(y) = f(x) + \nabla f(x)^T \delta + \tfrac{1}{2}\delta^T \nabla^2 f(z)\, \delta \qquad z \in [x, y]$$

$$\nabla^2 f(z) \preceq \beta \|\delta\|^2$$

β-smoothness

$$\Rightarrow \boxed{f(y) \le \underbrace{f(x) + \nabla f(x)^T \delta + \tfrac{1}{2}\beta \|\delta\|^2}_{\hat{f}(y)}}$$



$$\rightarrow \quad f(x) = \hat{f}(x)$$

$$\rightarrow \quad \forall_y \; f(y) \le \hat{f}(y)$$

$$x^{t+1} \leftarrow x^t - \eta \nabla f(x^t)$$

$$f(x^{t+1}) \le \hat{f}(x^{t+1}) = f(x^t) - \eta \underbrace{\nabla f(x)^T (\nabla f(x))}_{\|\nabla f(x)\|^2} + \eta^2 \frac{\beta}{2} \underbrace{\nabla f(x)^T \nabla f(x)}_{\|\nabla f(x)\|^2}$$

$$= f(x^t) - \eta \|\nabla f(x)\|^2 \left(1 - \eta \frac{\beta}{2}\right) \quad \Rightarrow \boxed{\eta = \frac{1}{\beta}}$$

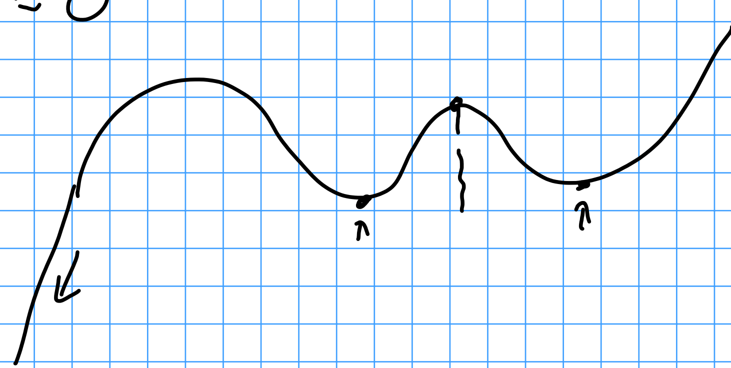
$$= f(x^t) - \underbrace{\frac{1}{2\beta} \|\nabla f(x)\|^2} < f(x^t)$$

Critical point: $\|\nabla f(x)\| \approx 0$

$\Rightarrow$ local maximum

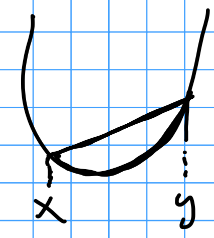
local minimum

saddle point

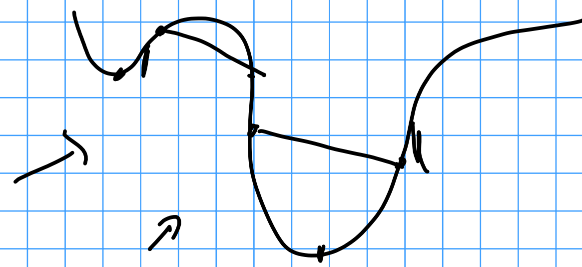


Convexity:

f is convex iff



$$\forall_{x,y} \ \forall_{0 \le \lambda \le 1} \quad \lambda f(x) + (1-\lambda) f(y) \geq$$
$$\geq f(\lambda x + (1-\lambda) y)$$



If f is convex

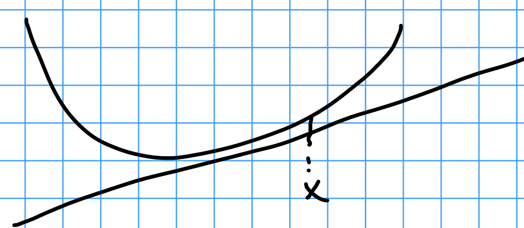
$\Rightarrow$ if $\|\nabla f(x)\| = 0$

$\Rightarrow$ x is a global min

f is convex

$$\Rightarrow \boxed{\forall_{y} \quad \delta^T \nabla f(y) \delta \geq 0}$$

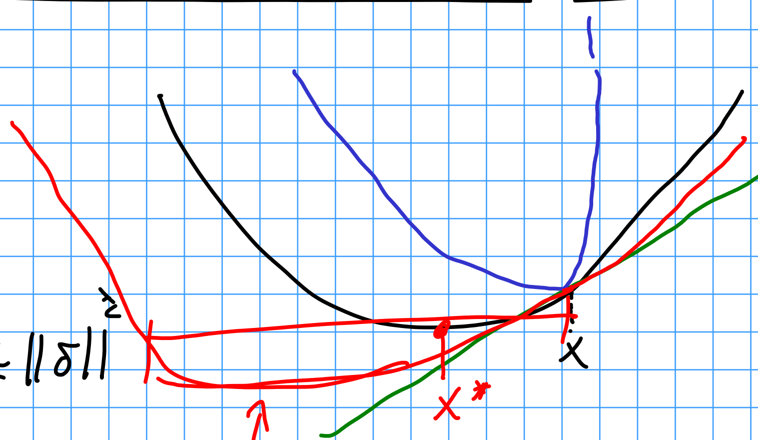
$$\forall_{y,x} \quad \underbrace{f(y) \geq f(x) + \nabla f(x)^T \delta}_{y = x+\delta} \ \underbrace{\cancel{+ \tfrac{1}{2} \delta^T \nabla^2 f(z) \delta}}_{\geq 0}$$



f is α-strong convex

$$\forall_{y} \quad \delta^T \nabla^2 f(y) \delta \geq \alpha \|\delta\|^2$$

$$\Rightarrow \forall_{x,y} \quad f(y) \geq f(x) + \nabla f(x)^T \delta + \frac{\alpha}{2} \|\delta\|^2$$



Let $x^* = \arg\min_x f(x)$

$\|x^t - x^*\|_2^2 \Leftarrow$ potential

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2x^T y$$

$$\|x^{t+1} - x^*\|^2 = \|x^t - \eta \nabla f(x^t) - x^*\|^2 =$$

$$= \|x^t - x^*\|^2 + \|\eta \nabla f(x^t)\|^2 - 2\eta \nabla f(x^t)^T (x^t - x^*)$$

$$\Delta \text{ potential} = 2\eta \underbrace{\nabla f()^T (x^t - x^*)} - \underbrace{\|\eta \nabla f\|^2}$$

$$\Rightarrow \quad \|x^{t+1} - x^*\|^2 \le \left(1 - \frac{\alpha}{\beta}\right) \|x^t - x^*\|^2 \le \underbrace{\left(1 - \frac{\alpha}{\beta}\right)^t}_{\substack{\| \\ K^{-1}}} \|x^0 - x^*\|_2^2$$

After $O\left(K \left(\log \frac{\|x^0 - x^*\|^2}{\varepsilon}\right)\right)$

$\|x^t - x^*\| \le \varepsilon$

$$\boxed{K = \frac{\beta}{\alpha} \geq 1}$$

Condition number

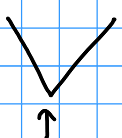
Max flow:

$$\min_{x \in F_{s,t}} \|x\|_\infty$$

Not diff

$F_{s,t}$ = space of all the unit s-t flows

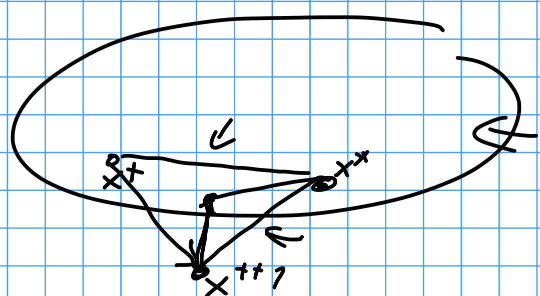
$= \{x : Bx = \chi_{s,t}\}$



① Constraints:

Add projection step

$\Rightarrow$ everything works

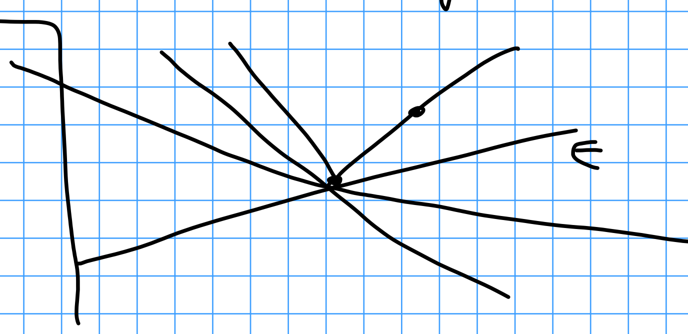


② $f(x) = \|x\|_\infty$

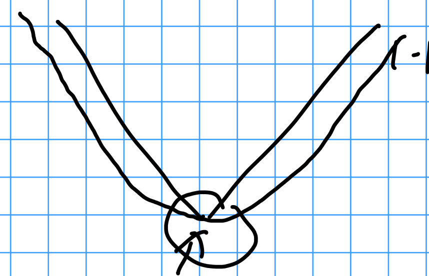
- Not diff
- α-strongly convex

Instead of gradients $\rightarrow$ use subgradients

Analisis that proves (much) weaker convergence



Smoothening $\quad \|x\|_\infty$



$$\text{smax}_\delta(x) := \ln\left(\frac{\sum_i e^{\frac{x_i}{\delta}} + e^{-\frac{x_i}{\delta}}}{2m}\right)$$

1. $\text{smax}_\delta(x) \leq \|x\|_\infty$
2. $\text{smax}_\delta(x) \geq \|x\|_\infty - \underbrace{\delta \ln(2m)}$
3. $\text{smax}_\delta(x)$ is β-smooth with $\boxed{\beta \approx \frac{1}{\delta}}$

---

3. Not α-strongly convex

$\Rightarrow \quad f(x^t) - f(x^*) \leq \varepsilon \qquad t \geq \Omega\left(\frac{\beta \|x^0 - x^*\|^2}{\varepsilon}\right)$

$\Rightarrow \quad$ Max flow: $\quad \beta = \frac{1}{\varepsilon} \qquad \delta \approx \varepsilon$

$$\|x^0 - x^*\|_2^2 \qquad x^0 = 0$$

$$\approx \Theta(m) \overset{\text{sparse graphs}}{=} \Theta(n)$$

$\Rightarrow \quad$ Total runtime:

ε-approx solution in

time $\quad O\left(\frac{m}{\varepsilon^2}\right) \cdot \tilde{O}(m) \overset{\text{sparse}}{=} \tilde{O}\left(n^2/\varepsilon^2\right)$

Regularization: $\quad g(x) = f(x) + \alpha \underbrace{\|x - x_0\|_2^2}_{\text{regularization}}$

$$\alpha = \frac{\varepsilon}{\|x - x^0\|^2}$$

$\min_{x \in F_{st}} |x|_1 \leftarrow \ell_2$ $\leftarrow$ shortest path problem

$\uparrow$ ∞ met