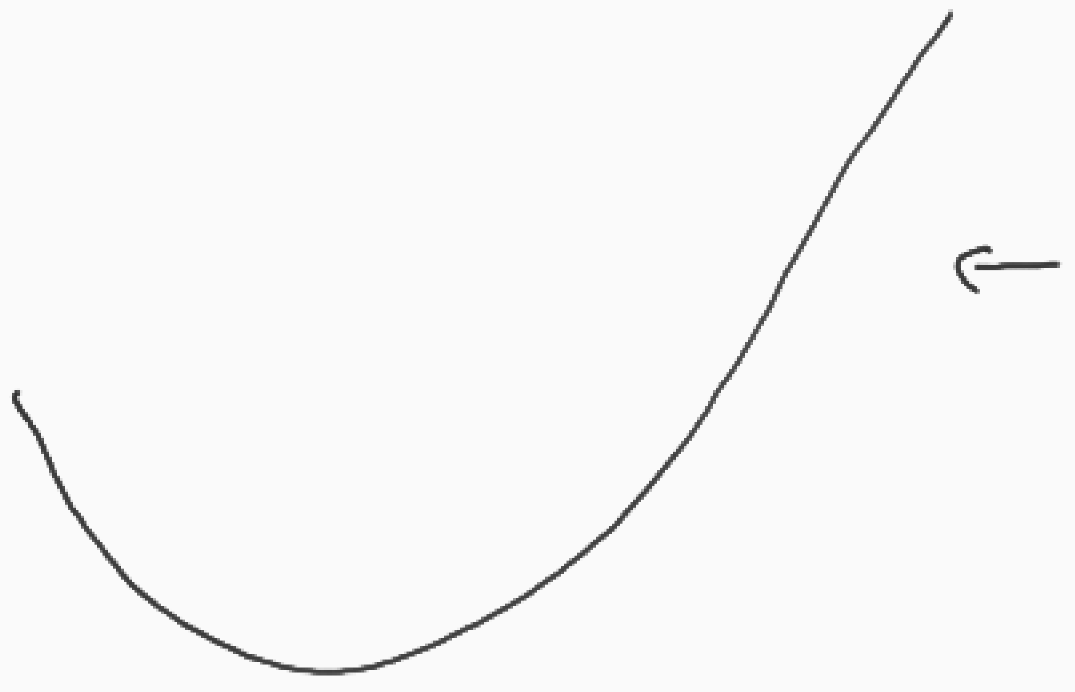
Lecture 18: Gradient Descent Framework

- Basics
- The algorithm and analysis
- Extension to online Convex Opt (OCO)
    - smoothness
    - well conditionedness



$f : \mathbb{R}^n \to \mathbb{R}$

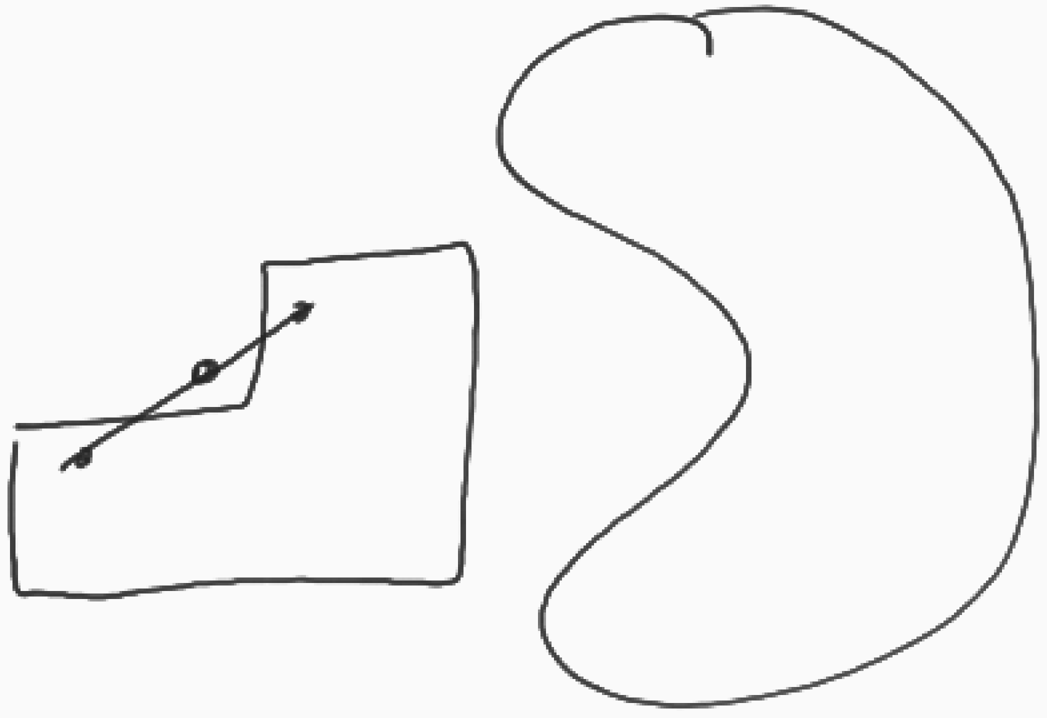
convex fn

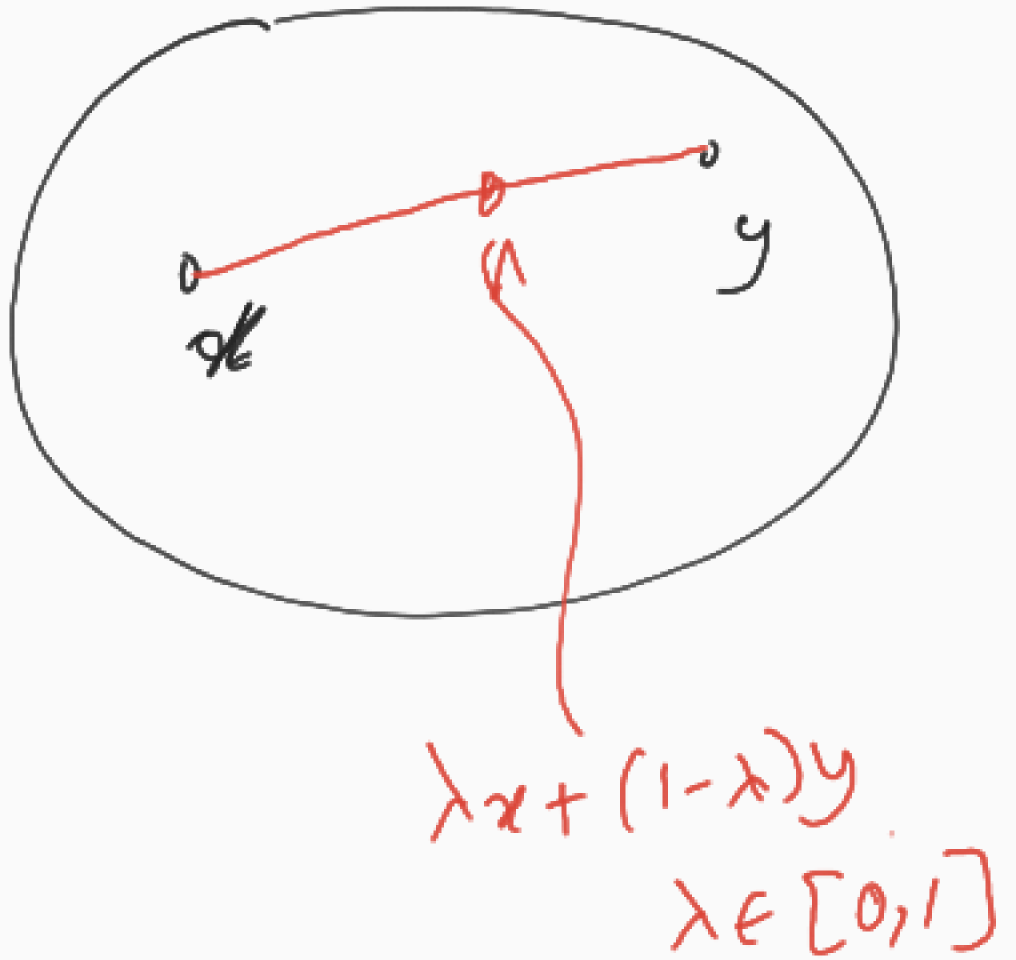
$\arg\min_{\cancel{x \in \mathbb{R}^n}} f(x)$ ← UCM

$x \in K$ ← CCM

Convexity

• K convex ~~conv~~ set





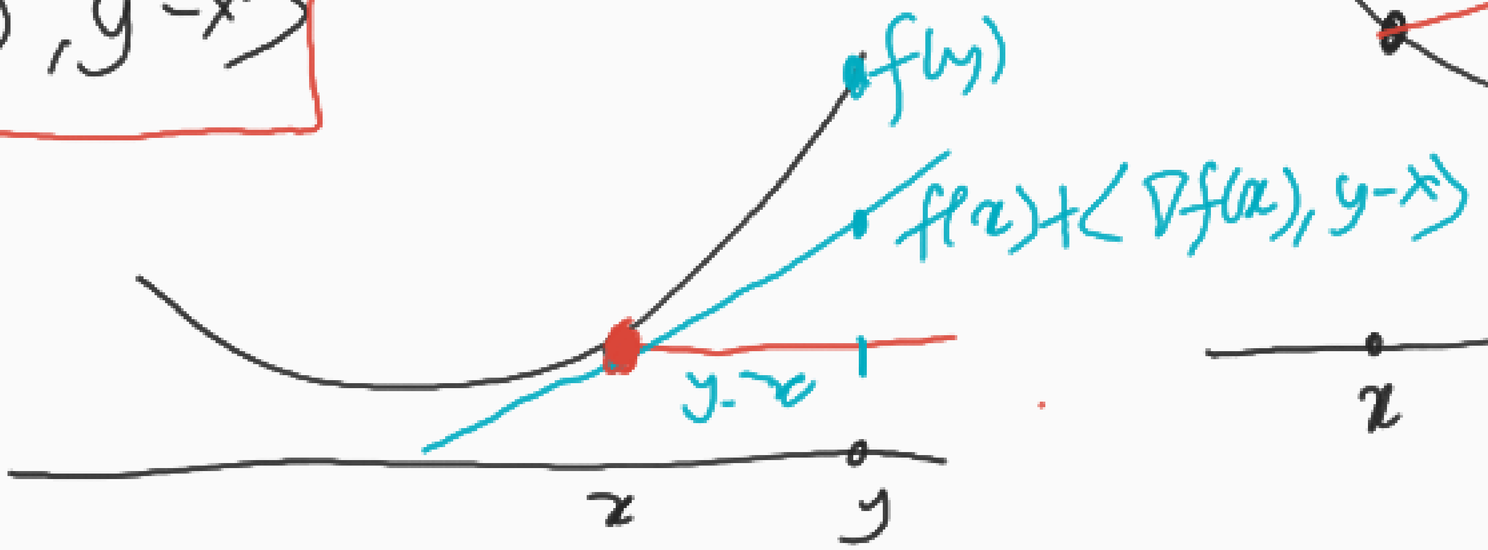
$$\forall x, y$$

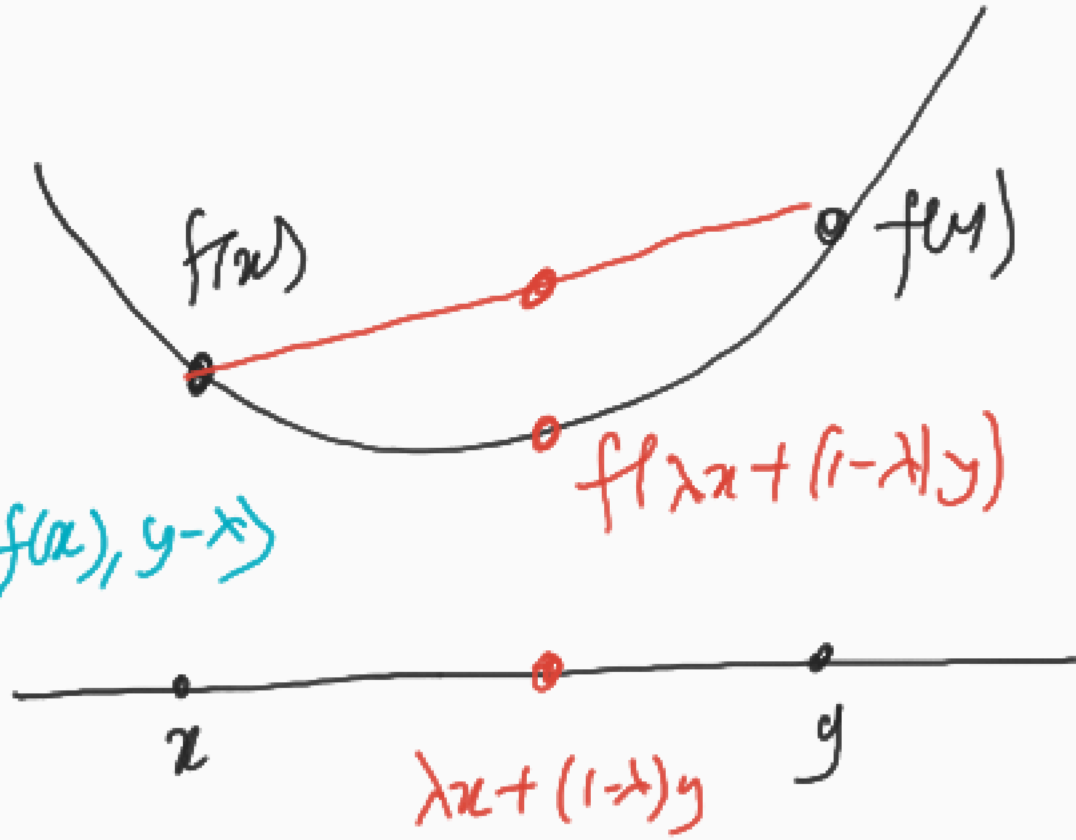
• $\lambda f(x) + (1-\lambda) f(y) \geq f(\lambda x + (1-\lambda) y)$

$$\boxed{\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f}{\partial x_2} \cdots \frac{\partial f(x)}{\partial x_n} \right)}$$

• $f(y) \geq \boxed{f(x) + \langle \nabla f(x), y - x \rangle}$

• $\boxed{H_f(x) \succcurlyeq 0}$





Unconstrained CM

$$x^* \leftarrow \arg\min_x f(x)$$

f convex fn.

$$\boxed{\min \mathcal{E}(f) = \sum_e f_e^2 r_e \quad \text{st } f \text{ unit flow from } s \to t, \quad f \in K}$$

Goal: satisfied with $\hat{x}$

$$f(\hat{x}) \le f(x^*) + \underbrace{\varepsilon}_{\uparrow \text{ error}}$$

Oracle access: **Gradient oracle**:

given x, return $\nabla f(x)$

(vs Value oracle: return $f(x)$)

solve $L\phi = (e_s - e_t)$

$$Ax = b$$

$$\boxed{\min f(x) = \|Ax - b\|^2}$$

"Thm" Can get goal in $O(1/\varepsilon^2)$ calls to oracle

↑ hiding "goodness of f"

[Next week: $O(\log 1/\varepsilon)$ calls] ← more baggage

Optimality Conditions

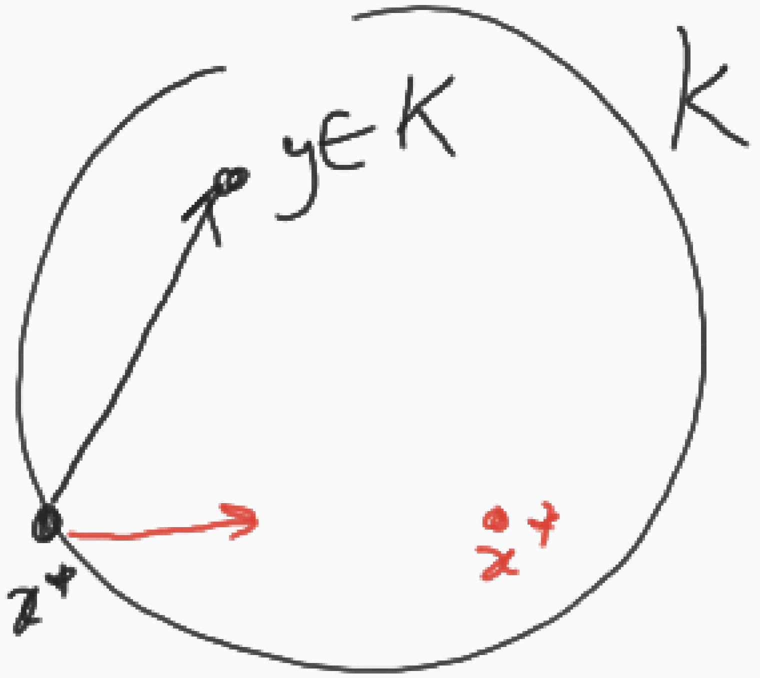
$$x^* = \arg\min_x f(x) \quad \underset{(f \text{ convex})}{\Longleftrightarrow} \quad \nabla f(x^*) = 0 \quad \Rightarrow \quad f(y) \geq f(x^*) + \underbrace{\langle \nabla f(x^*), y - x^* \rangle}_{=0}$$

$$x^* = \arg\min_{x \in K} f(x) \quad \Longleftrightarrow \quad \boxed{\langle \nabla f(x^*), y - x^* \rangle \geq 0 \quad \forall y \in K}$$

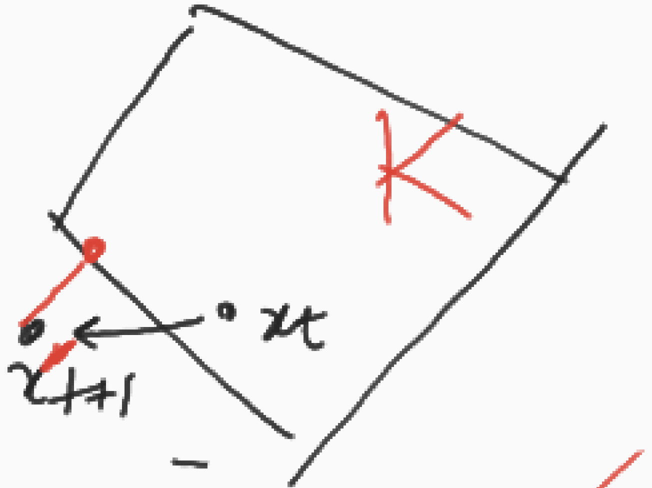
symmetric

$$f(x) = \frac{1}{2}\|Ax - b\|^2$$

$$\nabla f(x) = A^T A x - A^T b = 0$$



Gradient Descent

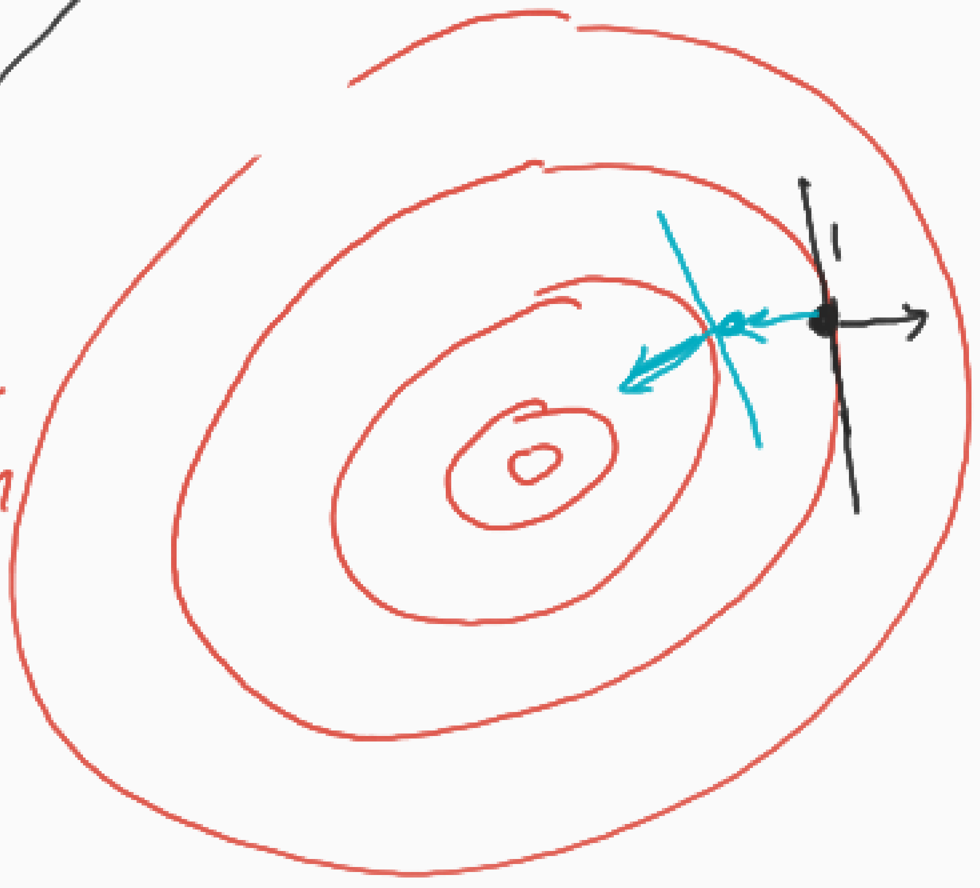


$x_1 \leftarrow$ same start pt

for $t = 1$ to T

step size $= \eta$

$$x'_{t+1} \leftarrow x_t - \eta_t \nabla f(x_t), \quad x_{t+1} = \Pi_K x'_{t+1}$$



return $\hat{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$ // or return $\underbrace{x_{T+1}}_{\text{last iterate}}$

Thm: $f(\hat{x}) \le f(x^*) + \epsilon$ for $\eta =$ ___

$T =$ ___

$$f(\hat{x}) \le f\left(\frac{1}{T}\sum x_t\right) \le \frac{1}{T}\sum_{t=1}^{T} f(x_t) \overset{!}{\le} \frac{1}{T}\sum_{t=1}^{T} f(x^*) + \varepsilon$$

Pf (cont'd) want $\frac{1}{T}\sum_1^T f(x_t) \overset{!}{\le} \frac{1}{T}\sum_1^T f(x^*) + \varepsilon$

$$\Phi_t = \frac{1}{2\eta}\|x_t - x^*\|^2$$

$$\boxed{f(x_t) + \underbrace{(\Phi_{t+1} - \Phi_t)}\le f(x^*) + \underbrace{\text{junk}}}$$

$$\Rightarrow \frac{1}{T}\sum f(x_t) \le \frac{1}{T}\sum f(x^*) + \underbrace{\boxed{\frac{1}{T}\sum \text{junk} + \frac{1}{T}\Phi_1}}_{\le \varepsilon} \cancel{+\Phi_{T+1}}^{\ge 0}$$

Pf $\le f(x_t) + \frac{1}{2\eta}\left[\|\underbrace{x_{t+1} - x^*}_{a+b}\|^2 - \|\underbrace{x_t - x^*}_{a}\|^2\right] \le$

$$\|a+b\|^2 = \|a\|^2 + \|b\|^2 + 2\langle a, b\rangle$$

$$= f(x_t) + \frac{1}{2\eta}\left[\|\underbrace{x_{t+1} - x_t}\|^2 + 2\langle \underbrace{x_{t+1} - x_t}, x_t - x^*\rangle\right]$$

$$x_{t+1} = x_t - \eta\underbrace{\nabla f(x_t)}_{\nabla}$$

$$= f(x_t) + \frac{1}{2\eta}\left[\|\eta^2\nabla^2\| + 2\langle -\eta\nabla, x_t - x^*\rangle\right]$$

$$= \underline{f(x_t) + \langle \nabla f(x_t), x^* - x_t\rangle} + \eta/2\,\|\nabla f(x_t)\|^2$$

$$\le f(x^*) + \eta/2\,\|\nabla f(x_t)\|^2$$

$$\frac{1}{T}\sum_t f(x_t) \le \frac{1}{T}\sum f(x^*) + \underbrace{\left[\frac{1}{T}\sum_t \underbrace{\|\nabla f(x_t)\|^2}_{\le G^2}\cdot\frac{\eta}{2} + \frac{\overbrace{\|x^* - x_1\|}^{D}{}^2}{2\eta T}\right]}_{\text{want } \le \varepsilon}$$

$$\cancel{\tfrac{1}{T}}\,\cancel{T}\cdot G^2\cdot \eta/2 + D^2/2\eta T$$

$$\eta = \frac{D}{G\sqrt{T}}$$

$$\Rightarrow \#\text{ iters} = O(1/\varepsilon^2)$$

$$\boxed{T = \left(\frac{GD}{\varepsilon}\right)^2}$$

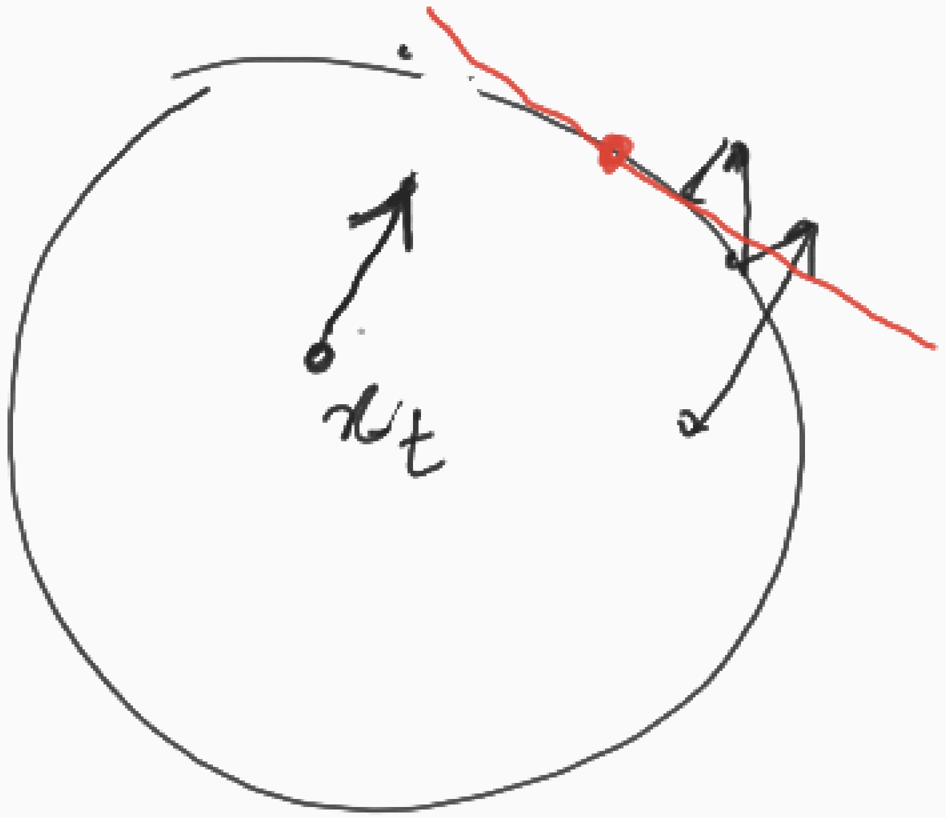
Constrained Convex Opt: $\quad K \qquad \min_{x\in K} f(x)$

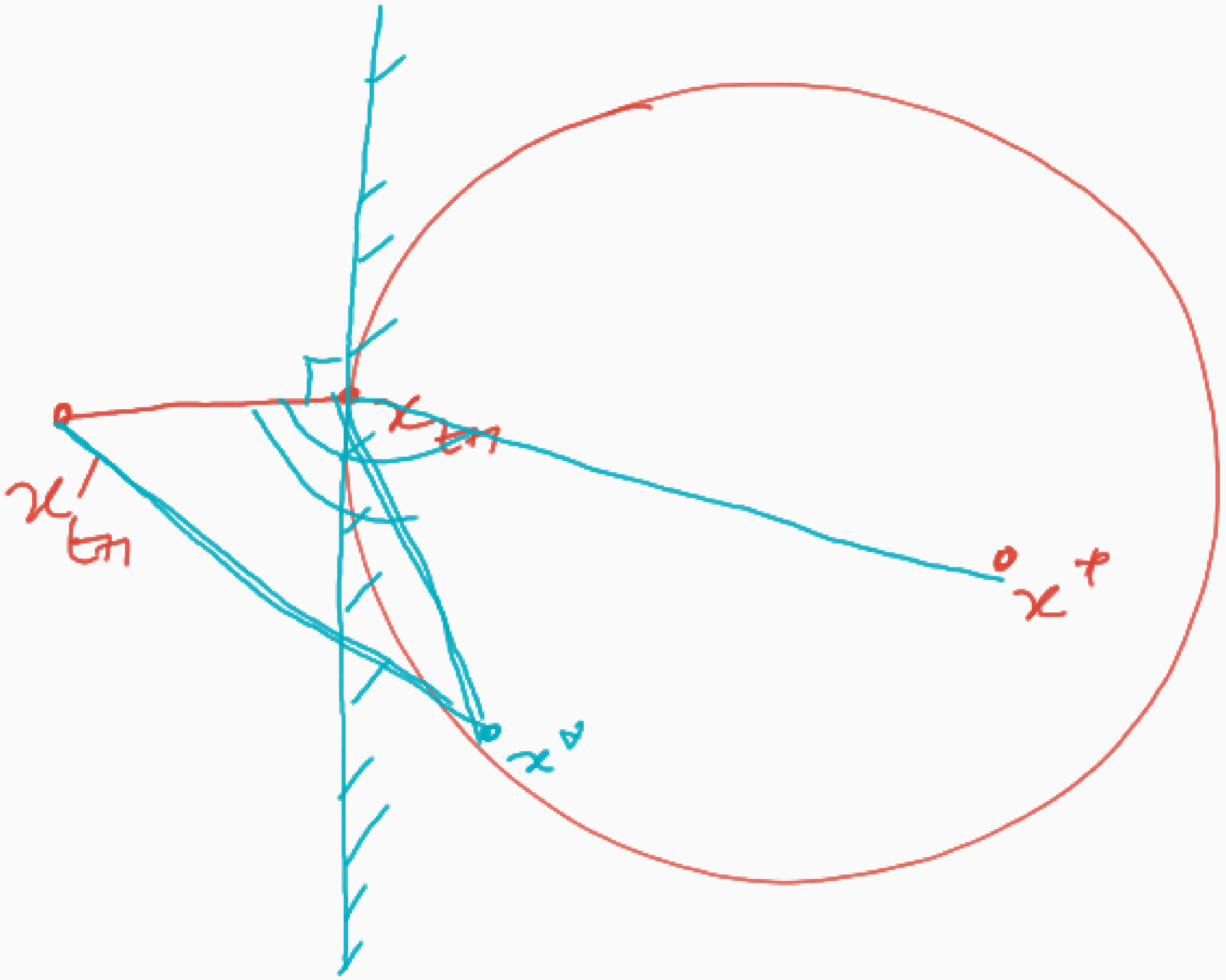
$$G = \max_{x\in K}\|\nabla f(x)\|$$

$$D = \|x_1 - x^*\| \le \operatorname{diam}(K)$$

Projected GD

$$\boxed{\eta_t = \frac{D}{G\sqrt{t}}}$$

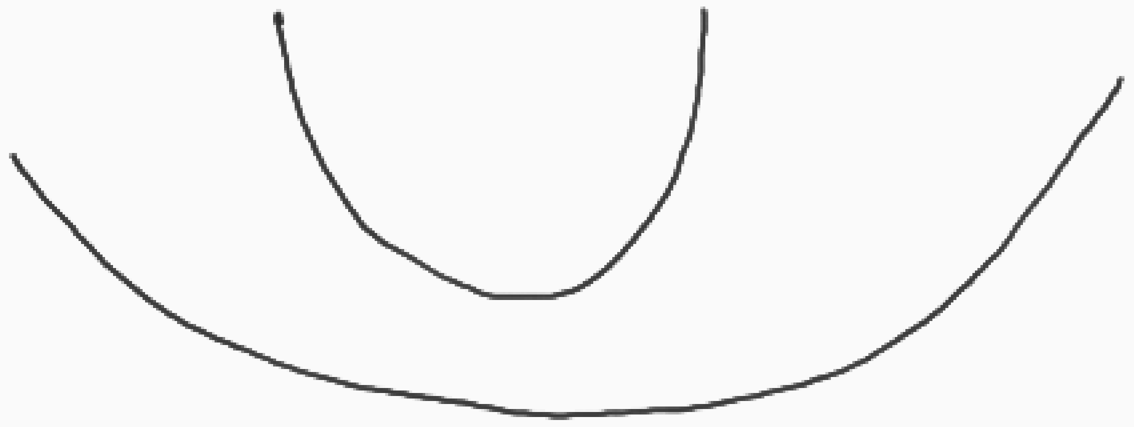


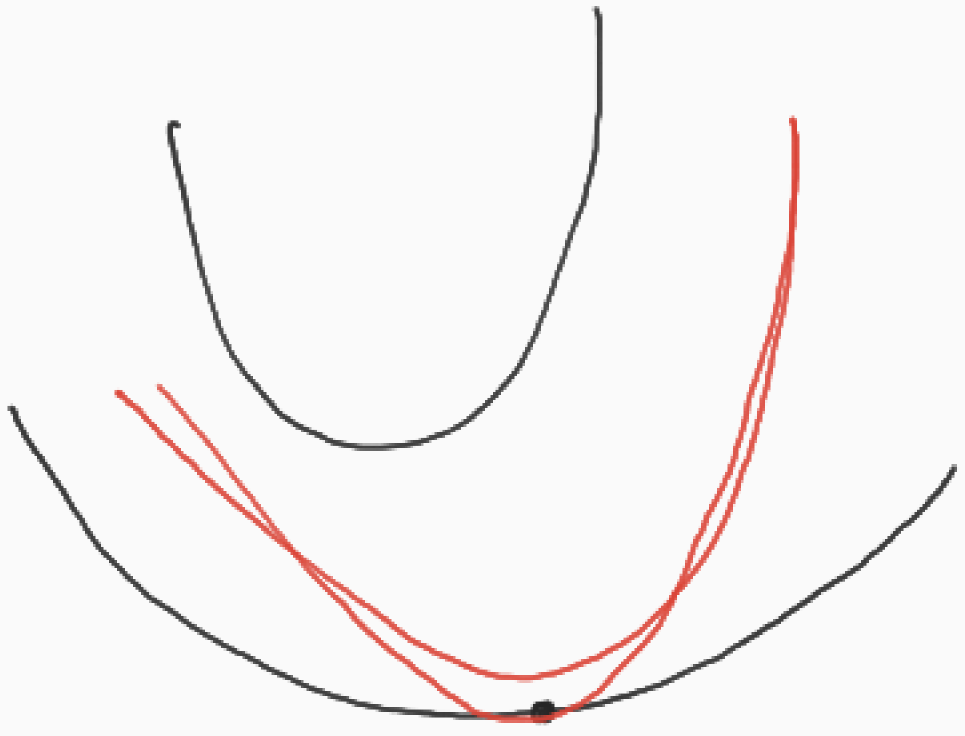


Frank Wolfe

"Nice"

"Nice"





Smooth

$$f(y) \le f(x) + \langle \nabla f(x), y - x \rangle + \frac{\beta}{2} \|y - x\|^2$$

$1/\varepsilon$

Strongly Convex

$$f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2} \|y - x\|^2$$

$1/\varepsilon$

Well-Conditioned Smooth + SC

$$O(\lg 1/\varepsilon)$$

K

$K = \frac{\beta}{\alpha}$

↑ condition number

$\underbrace{Ax = b}$ $\quad \alpha \le$ eigenvalue of $A \le \beta$