262B-Lecture 13

Date created: 2021.03.04
N. of Pages: 14

	$\min f(x)$	$\min f(x)$ s.t. $x \in X$
Convex	$O(\frac{1}{\varepsilon})$	$O(\frac{1}{\varepsilon}) \rightarrow$ Conditional gradient
strongly convex	$O(\log \frac{1}{\varepsilon})$	?

Can't be $O(\log \frac{1}{\varepsilon})$ in general.

Thm: There is a class of problems where

$f(x)$ = quadratic & strongly convex and

X = described by linear inequalities such

that the sequence generated by conditional gradient method satisfies:

$$f(x^{(k)}) - f_* \geq \frac{1}{k^{1+\varepsilon}}$$

$\forall \varepsilon > 0$, infinitely many values of k.

$\Rightarrow$ $O(\frac{1}{\varepsilon})$ can't be improved under strong convexity assumption on $f(x)$.

How about strong convexity for X ?

Thm: Assume - $f(x)$ is convex

- $\boxed{\min_{x \in X} \|\nabla f(x)\| > 0}$ $\Rightarrow$ ($f(\cdot)$ has no stationary point in X)

- $\|\nabla f(x) - \nabla f(y)\| \leq L \|x-y\| \quad \forall x,y \in X$

- X is a strongly convex set

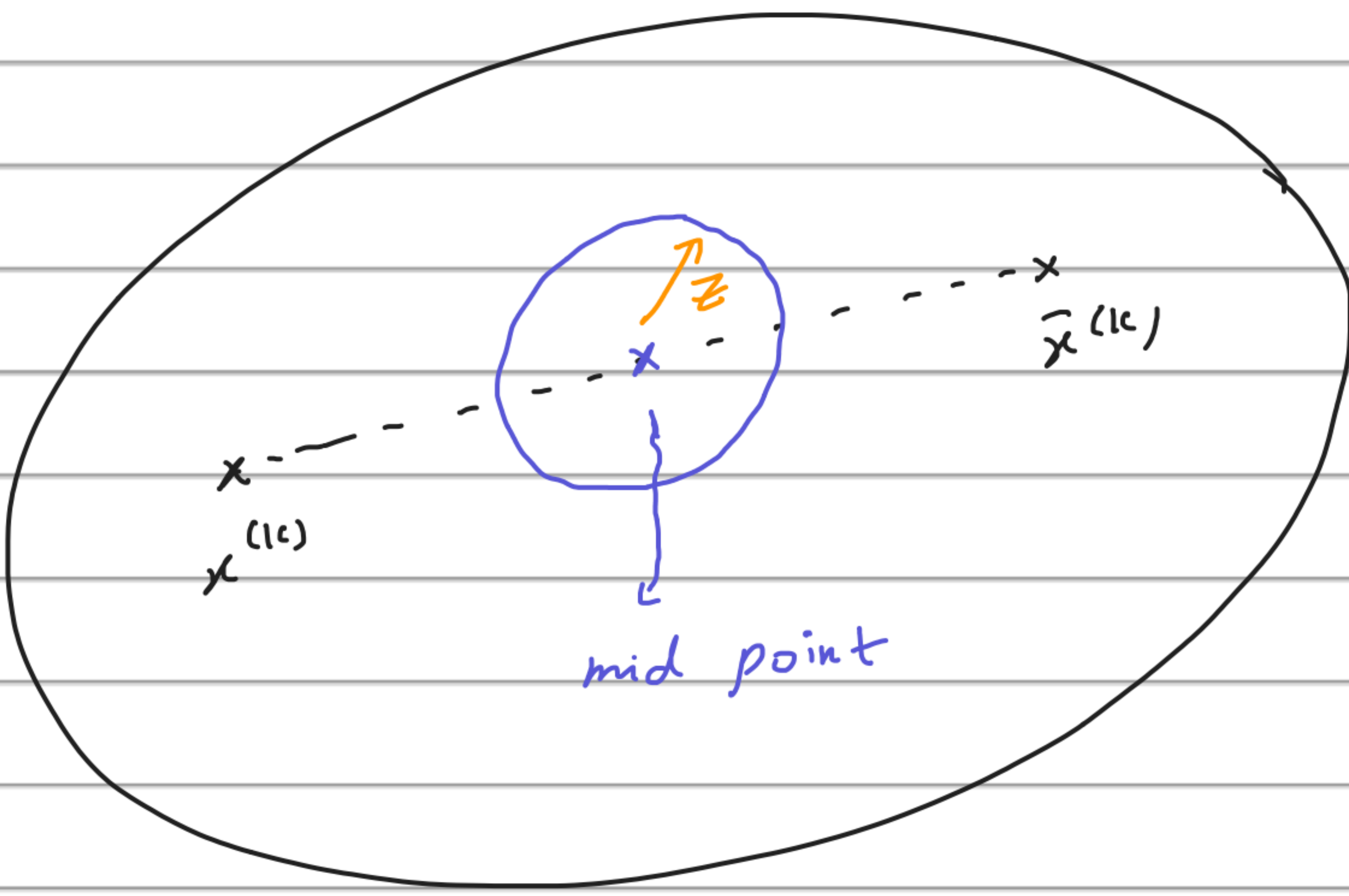
$\Rightarrow$ Linear convergence for conditional gradient method if stepsize is constant & small:

$O(\log \frac{1}{\varepsilon})$

Note: This doesn't require $f(x)$ to be strongly convex but an optimization sub-problem should be solved $O(\log \frac{1}{\varepsilon})$ times.

proof:

strong convexity of set X: with factor $\underline{\underline{m}}$



$$\underbrace{\frac{1}{2}x^{(k)} + \frac{1}{2}\bar{x}^{(k)} + \frac{1}{2}m\left(\frac{1}{2}\right)\left(1-\frac{1}{2}\right)\|x^{(k)} - \bar{x}^{(k)}\|^2 \times \underbrace{z}_{\|z\| \leq 1}}_{= y^{(k)}} \in X$$

Conditional gradient method:

$$\underbrace{\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})}_{\text{most descent direction}} \leq \nabla f(x^{(k)})^T(\underbrace{y^{(k)}}_{\in X} - x^{(k)})$$

$$= \frac{1}{2}\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}) + \frac{m}{8}\|x^{(k)} - \bar{x}^{(k)}\|^2 \times \nabla f(x^{(k)})^T z$$

Pick $z = -\dfrac{\nabla f(x^{(k)})}{\|\nabla f(x^{(k)})\|}$

Define: $\min_{x \in X} \|\nabla f(x^{(k)})\| = c > 0$

$$\Rightarrow \boxed{\nabla f(x^{(k)})^T (\bar{x}^{(k)} - x^{(k)}) \le \frac{1}{2} \nabla f(x^{(k)})^T (\bar{x}^{(k)} - x^{(k)}) - \frac{mc}{8} \|x^{(k)} - \bar{x}^{(k)}\|^2} \quad (1)$$

$$\boxed{\nabla f(x^{(k)})^T (\underbrace{\bar{x}^{(k)}}_{\text{conditional gradient method}} - x^{(k)}) \le \nabla f(x^{(k)})^T (\underbrace{x_*}_{\text{arbitrary solution}} - x^{(k)}) \underset{\text{Convexity}}{\le} f(x_*) - f(x^{(k)})} \quad (2)$$

$$(1), (2) \Rightarrow \boxed{\nabla f(x^{(k)})^T (\bar{x}^{(k)} - x^{(k)}) \le \frac{1}{2} (f(x_*) - f(x^{(k)})) - \frac{mc}{8} \|x^{(k)} - \bar{x}^{(k)}\|^2} \quad (3)$$

$$f(x^{(k+1)}) - f_* \le f(x^{(k)}) - f_*$$

$$+ \nabla f(x^{(k)})^T (\alpha^{(k)})(\bar{x}^{(k)} - x^{(k)}) + \frac{L}{2} \| \alpha^{(k)} (\bar{x}^{(k)} - x^{(k)}) \| \quad (4)$$

$\rightarrow$ quadratic over-estimator

$(3), (4) \Rightarrow \underbrace{f(x^{(k+1)}) - f_*}_{e^{(k+1)}} \le \underbrace{(f(x^{(k)}) - f_*)}_{e^{(k)}} \left(1 - \frac{\alpha^{(k)}}{2}\right)$

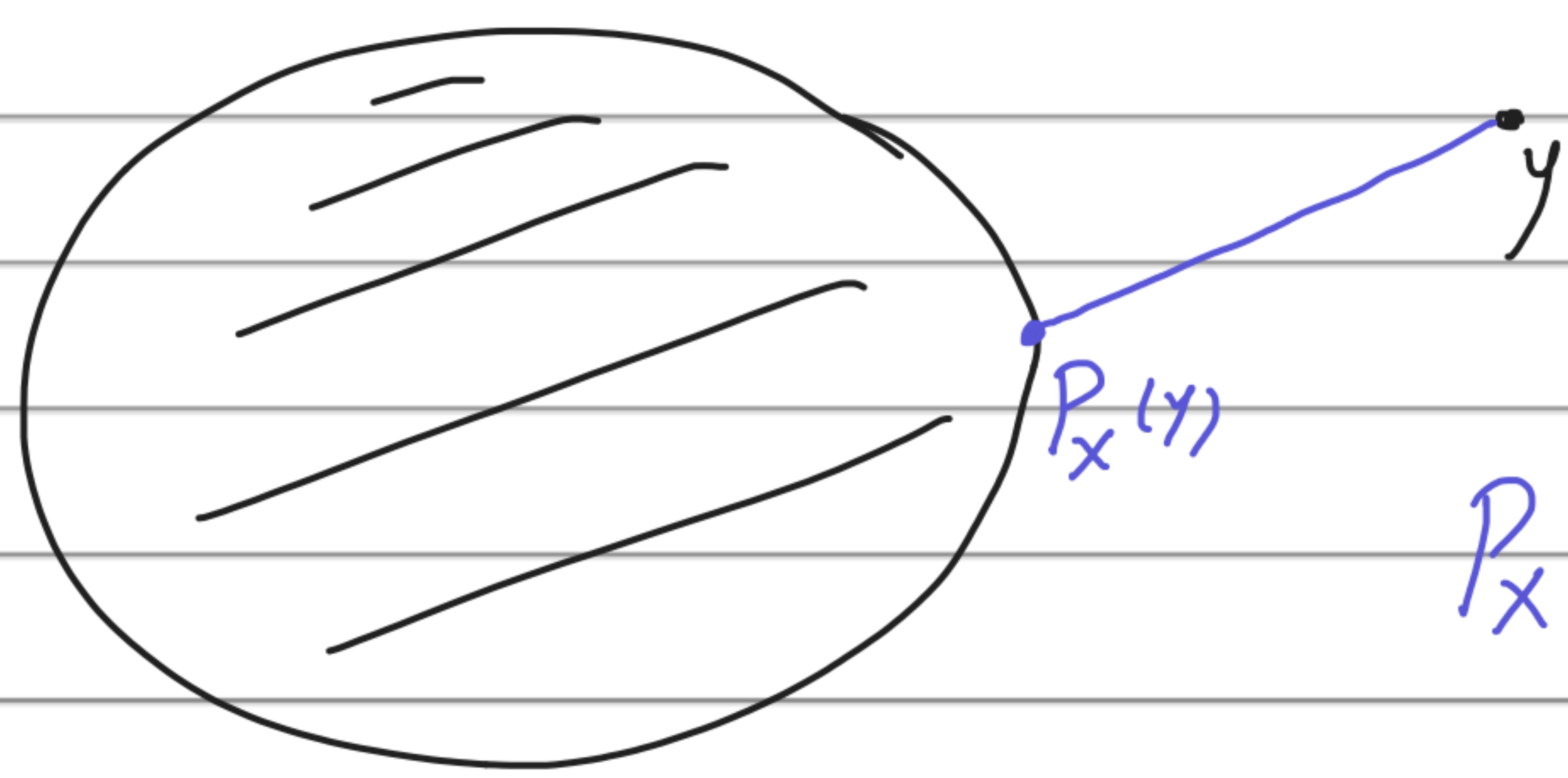
$$- \| x^{(k)} - \bar{x}^{(k)} \|^2 \times \alpha^{(k)} \times \underbrace{\left(\frac{-L\alpha^{(k)}}{2} + \frac{mc}{8} \right)}_{\text{is positive if } \alpha^{(k)} = \alpha = \text{small}}$$

So, α : small $\Rightarrow e^{(k+1)} \le e^{(k)} \underbrace{\left(1 - \frac{\alpha}{2}\right)}_{\in (0,1)}$

$\Rightarrow$ Linear convergence.

Gradient projection methods:

$$\begin{cases} x^{(k+1)} = x^{(k)} + \underbrace{\alpha^{(k)}}_{\text{stepsize}} \left(\bar{x}^{(k)} - x^{(k)}\right) \\ \bar{x}^{(k)} = \underbrace{P_X}_{\text{Projection on } X}\Big(\underbrace{x^{(k)} - \underbrace{s^{(k)}}_{\text{stepsize}} \nabla f(x^{(k)})}_{\text{regular gradient iteration}}\Big) \end{cases}$$



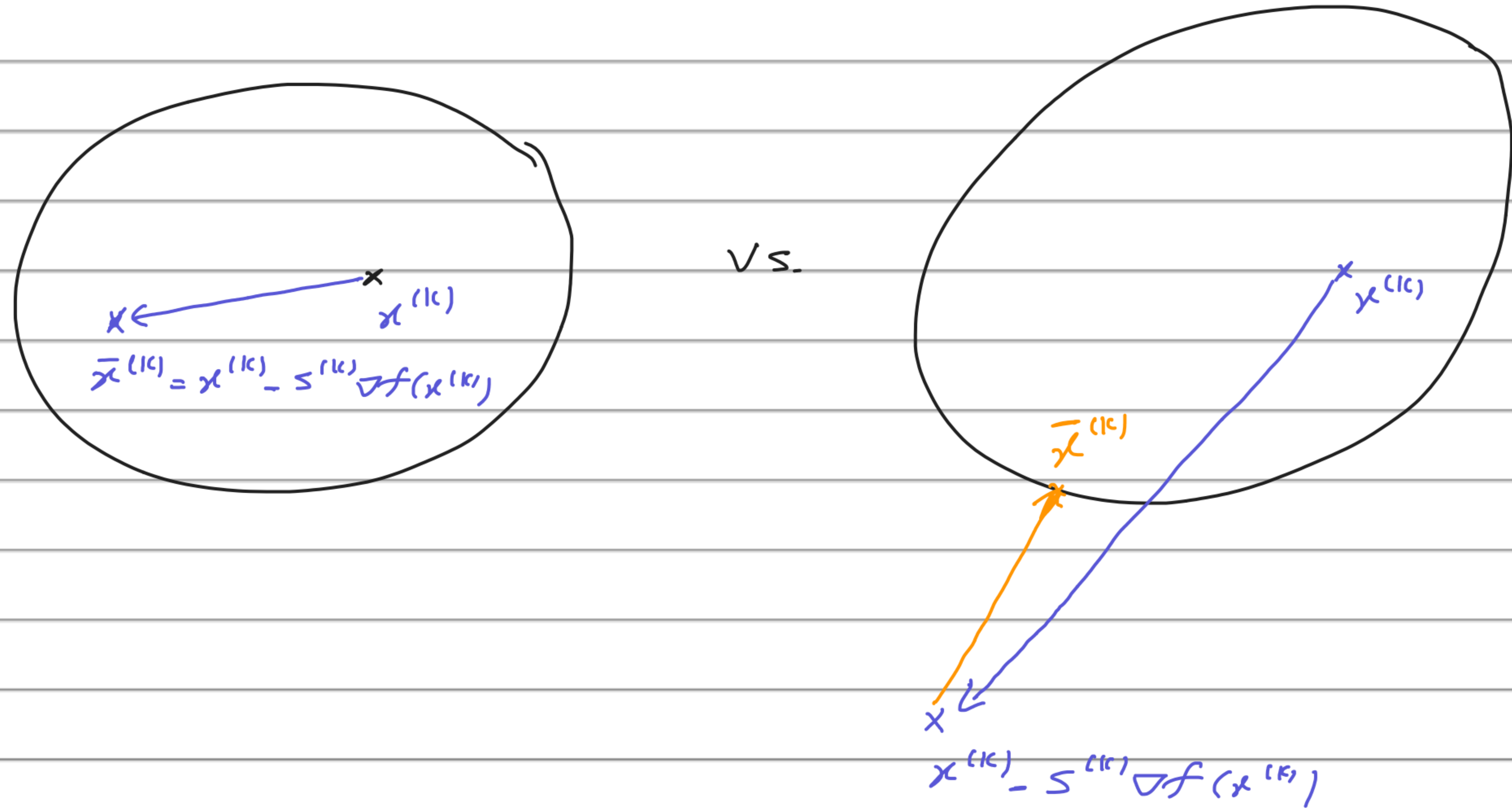
$$P_X(y) = \underset{x \in X}{\arg\min} \; \|x - y\|$$

$\|x-y\| \;\Rightarrow\; \|x-y\|^2 \rightarrow$ quadric (differentiable)

Projection: minimize a quadratic function over X

Idea: use gradient method on $x^{(k)}$ to find a better point with respect to $f(\cdot)$. It may not be feasible, so we project it onto X

to get a feasible point $\bar{x}^{(k)}$, and then get a feasible direction $\bar{x}^{(k)} - x^{(k)}$.



$$\bar{x}^{(k)} : \min_{x \in X} \| x - (x^{(k)} - s^{(k)} \nabla f(x^{(k)})) \|^2$$

$$= \min_{x \in X} \Big\{ \| x - x^{(k)} \|^2 + 2 s^{(k)} \nabla f(x^{(k)})^T (x - x^{(k)})$$

$$+ \underbrace{(s^{(k)})^2 \| \nabla f(x^{(k)}) \|^2}_{\text{no variable } x \text{ in it}} \Big\}$$

$$= \min_{x \in X} \Big\{ \nabla f(x^{(k)})^T (x - x^{(k)}) + \frac{1}{2 s^{(k)}} \| x - x^{(k)} \|^2 \Big\}$$

Conditional gradient :

$$\bar{x}^{(k)} : \underbrace{\min_{x \in X} \nabla f(x^{(k)})^T (x - x^{(k)})}_{\text{most descent direction}}$$

Foc ≥ 0
$\downarrow$
maximize violation of Foc

projected gradient :

$$\bar{x}^{(k)} : \min_{x \in X} \nabla f(x^{(k)})^T (x - x^{(k)}) + \underbrace{\frac{1}{2s^{(k)}} \| x - x^{(k)} \|^2}_{\text{regularizes } x \text{ with respect to } x^{(k)}}$$

At every iteration, we may need to perform a projection.

Assumption:

$\boxed{\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in X \end{array}}$ is much harder than $\boxed{\begin{array}{ll} \min & \text{quadratic function} \\ \text{s.t.} & x \in X \end{array}}$

Example: $X = \text{Box} = \{ x \mid a_i \leq x_i \leq b_i, \ i = 1, \dots, n \}$

$$P_X(y): \quad i\text{th entry} = \begin{cases} a & \text{if } y_i \le a \\ b & \text{if } y_i \ge b \\ y_i & \text{if } a \le y_i \le b \end{cases}$$

special case : projected gradient method

$$\alpha^{(k)} = 1 \implies$$

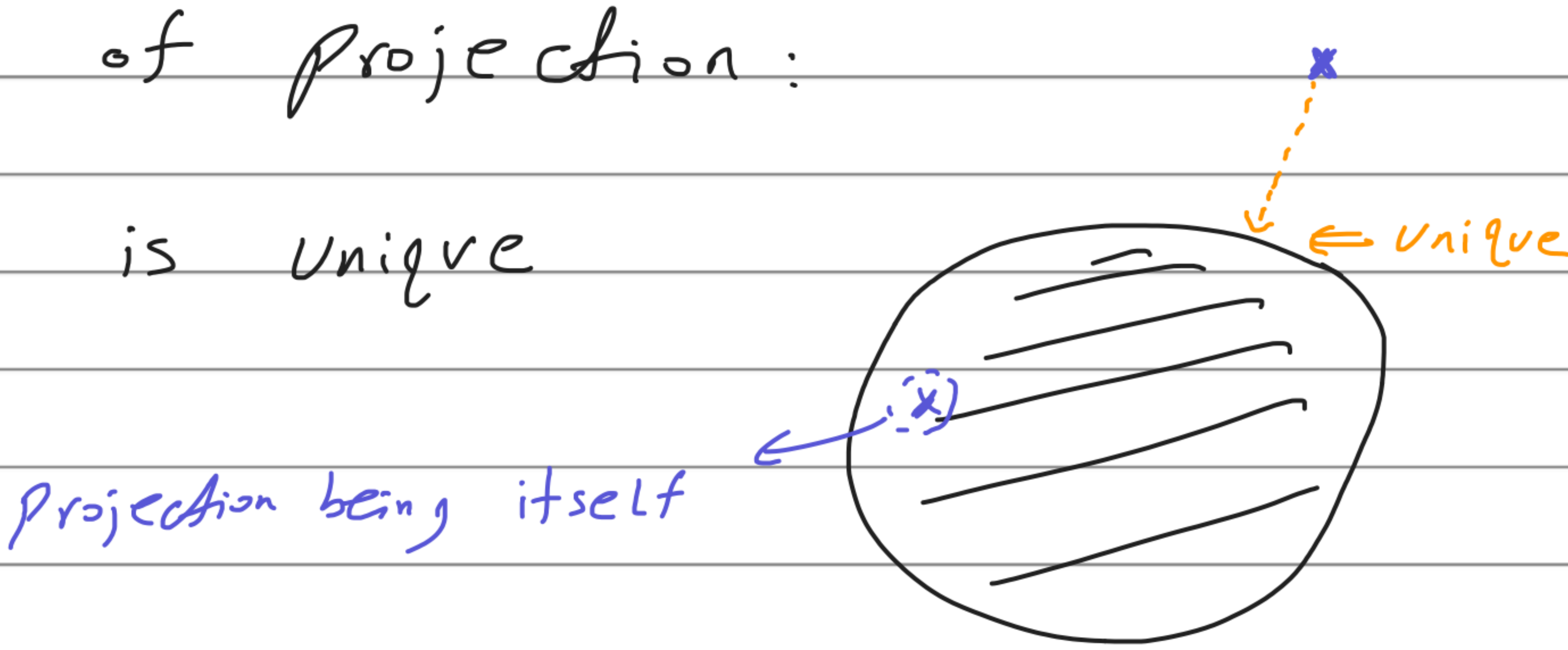
$$x^{(k+1)} = x^{(k)} + \alpha^{(k)}\left(\bar{x}^{(k)} - x^{(k)}\right) = \bar{x}^{(k)}$$

$$= P_X\left(x^{(k)} - s^{(k)} \nabla f(x^{(k)})\right)$$

Use regular gradient method, but whenever the point gets outside of the set, project it back into X.

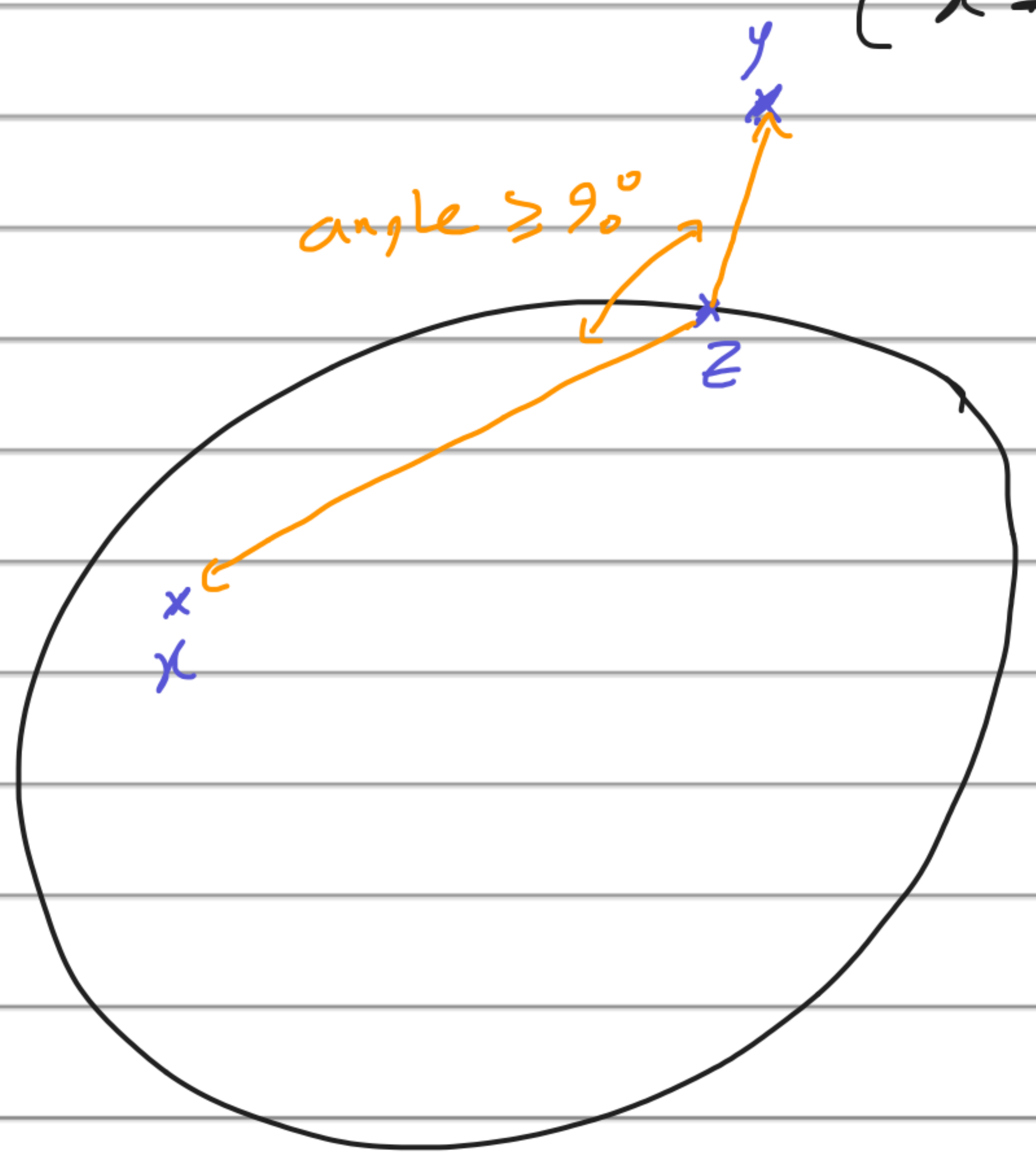
properties of projection:

1- $P_X(y)$ is unique



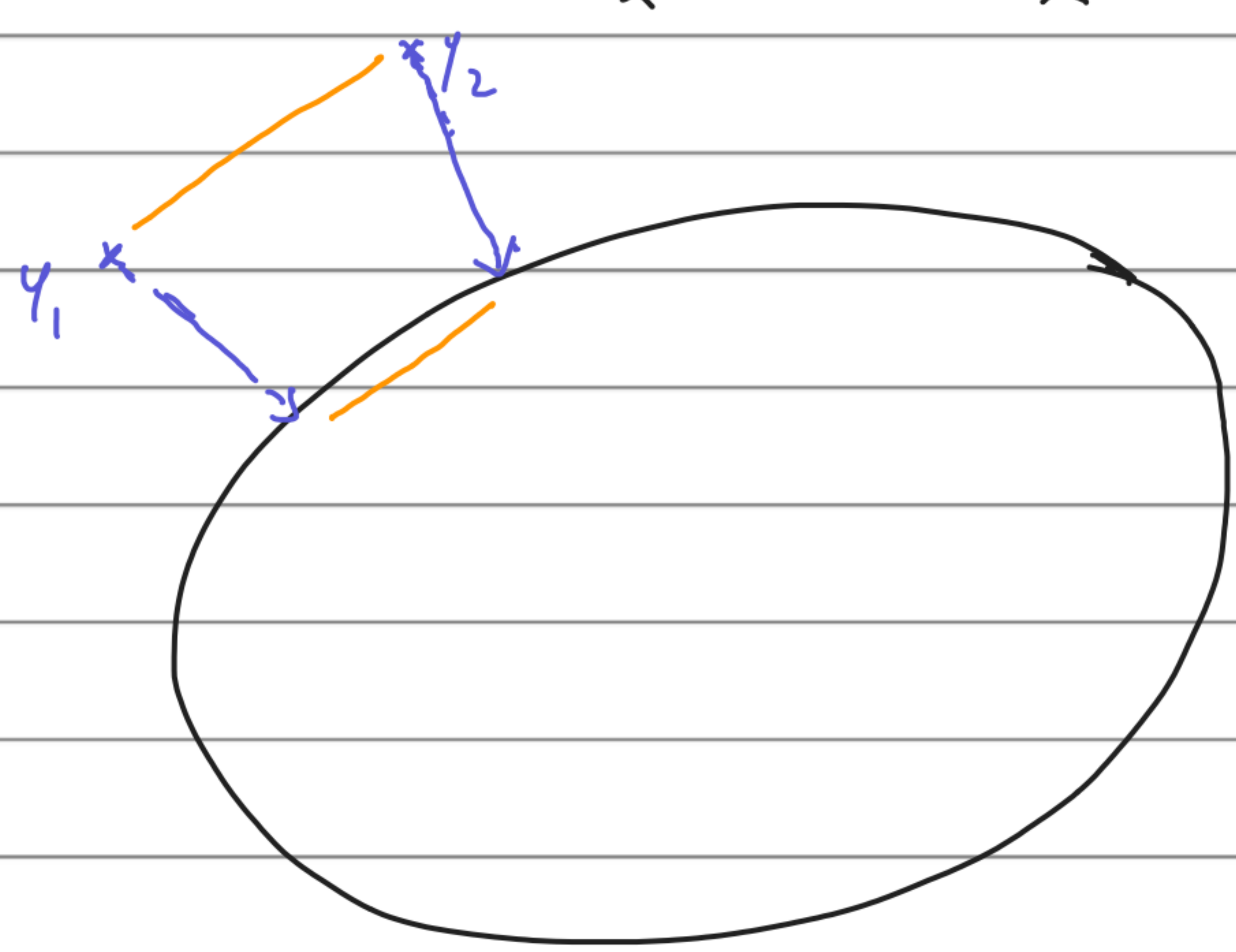
2 - z is projection of y on X if and only if :

$$(x-z)^T (y-z) \leq 0 \quad \forall x \in X$$



3 - $y_1, y_2 \in \mathbb{R}^n \Longrightarrow$

$$\| P_X(y_1) - P_X(y_2) \| \leq \| y_1 - y_2 \|$$



$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in X \end{array} \implies \text{FOC for a point } x_*:$$

$$\nabla f(x_*)^T (x - x_*) \geq 0 \quad \forall x \in X$$

$$\iff -s \nabla f(x_*)^T (x - x_*) \leq 0 \quad \forall x \in X$$

given any arbitrary $s > 0$

$$\iff (\underbrace{(x_* - s\nabla f(x_*))}_{y} - \underbrace{x_*}_{z})^T (x - \underbrace{x_*}_{z}) \leq 0 \quad \forall x \in X$$

Property 2

$$\iff P_X(x_* - s \nabla f(x_*)) = x_* \quad (\text{FOC})$$

By-product: gradient projection methods stop if and only if they find a stationary point

---

$$\begin{cases} x^{(k+1)} = x^{(k)} + \alpha^{(k)} (\bar{x}^{(k)} - x^{(k)}) \\ \bar{x}^{(k)} = P_X(x^{(k)} - s^{(k)} \nabla f(x^{(k)})) \end{cases}$$

How to select stepsizes?

1 - Limited line search:

$$s^{(k)} = s = \text{Constant}$$

$$\& \quad \alpha^{(k)}: \quad \min \; f(x^{(k)} + \alpha(\bar{x}^{(k)} - x^{(k)}))$$

$$\text{s.t.} \quad 0 \le \alpha \le 1$$

2 - Armijo rule along feasible direction:

$$s^{(k)} = s = \text{Const}$$

$$\& \quad \alpha^{(k)}: \quad 1, \; \beta, \; \beta^2, \; \vdots$$

$$\text{s.t.} \quad f(x^{(k+1)}) - f(x^{(k)}) \le \underbrace{\sigma}_{\in(0,1)} \times \alpha^{(k)} \times \nabla f(x^{(k)})^T (\bar{x}^{(k)} - x^{(k)})$$

3 - Armijo rule along projection arc:

$$\alpha^{(k)} = 1 = \text{Constant}$$

$$s^{(k)}: \quad s, \; s\beta, \; s\beta^2, \; \vdots \qquad s > 0, \quad 0 < \beta < 1, \quad 0 < \sigma < 1$$

s.t.

$$f(x^{(k+1)}) - f(x^{(k)}) \leq \sigma \nabla f(x^{(k)})^T \underbrace{(x^{(k+1)} - x^{(k)})}$$

$$P_X(x^{(k)} - s^{(k)} \nabla f(x^{(k)}))$$

note $x^{(k+1)} - x^{(k)} \neq -s^{(k)} \nabla f(x^{(k)})$ in general

How about Line search for $s^{(k)}$?

Then, we shoud minize a complex function over X at every iteration $\Longrightarrow$ not plausible

4 - Constant stepsize:

$$\alpha^{(k)} = 1 , \quad s^{(k)} = s$$

5 - Diminishing stepsize

$$\alpha^{(k)} = 1 , \quad s^{(k)} \longrightarrow 0 , \quad \sum s^{(k)} = \infty$$

special case:

$$X = \{ x \mid \underbrace{a_i}_{\text{vector}}^T x \le \underbrace{b_i}_{\text{scalar}}, \ i = 1, \dots, m \}$$

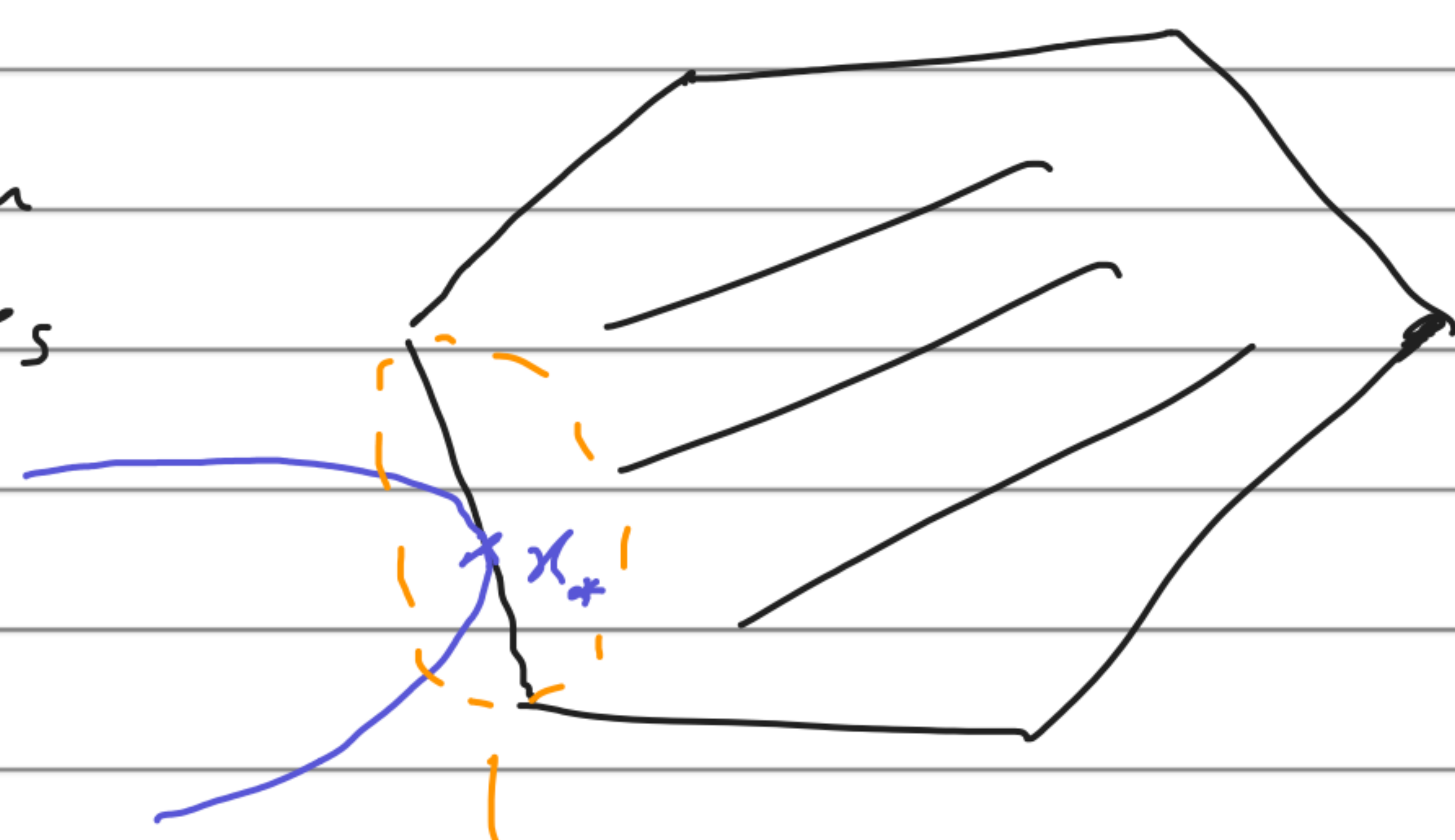
set of binding constraints at a point y:

$$\{ i \mid a_i^T y = b_i, \ i = 1, \dots, m \}$$

Thm: Use Armijo rule along projection arc.

If k is large enough, then the set of binding constraints at $x^{(k)}$ = set of binding constraints at x_*.

Ex: min quadratic function
s.t. Linear inequalities



Not true if we use Armijo rule along feasible direction

fast $\Leftarrow$ Convergence

after a finite number of iterations, we land on that face