262B-Lecture 1

Date created: 2021.01.19
N. of Pages: 16

262B (262A: prerequisite)

Tue - Thu 12:30 - 2

office hours: Thu 2-3

small class: No GSI

Record lectures ⟶ Dropbox

Grading: 4-5 homework sets (15%)

Exam: March 30 (40%)

project: (45%)

project:
- pick a topic
- formulation / algorithm / results; formulate / simulations
- 10 min presentation in the last lecture
- Report at the end of RRR week

Optimization 2: Theory + algorithms

(non-convex)

$$\min_{x \in \mathbb{R}^n} f(x)$$

$$\text{s.t.} \quad h_i(x) = 0 \qquad i = 1, \dots, m$$

$$g_j(x) \leq 0 \qquad j = 1, \dots, k$$

optimality conditions

x_*

(local min / global min)

algorithms

hard $\xrightarrow[\text{approximate}]{}$ easy

1) Optimality Conditions: first- and second-order conditions, regularity, tangent plane, Lagrangian, duality, exact penalty, inexact penalty

2) Convexity: Convex set/function/optimization, Conic optimization

3) Convexification: relaxations, sum-of-squares, Low-rank optimization

4) Algorithms: Descent, Conjugate gradient, gradient projection, Conditional gradient, block coordinate, proximal, ADMM distributed computation, convergence

5) Applications: Machine Learning

Thm + Proofs

Bertsekas, Luenberger, Boyd, Yinyu ye

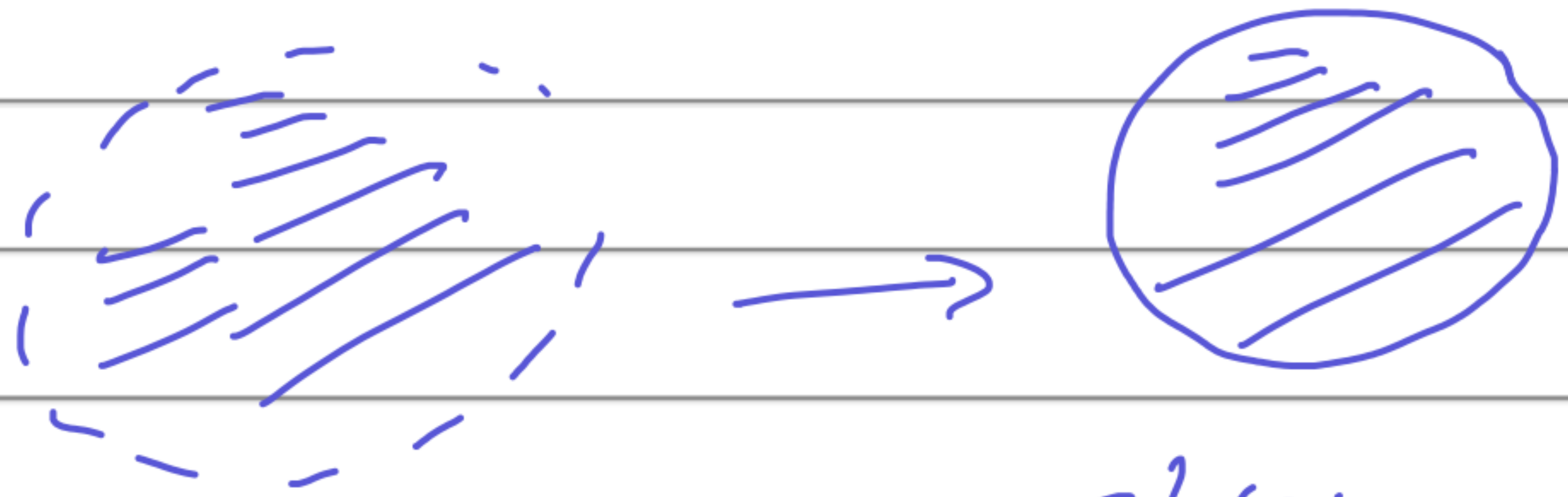
main textbook — optional — Convex — Low-rank

$$\min_{x \in \mathbb{R}^n} \quad f(x) \quad \text{s.t.} \quad x \in X$$

X : open, closed, bounded, compact ?

$X \longrightarrow$ closure of $X = cl(X)$

$X = cl(X)$



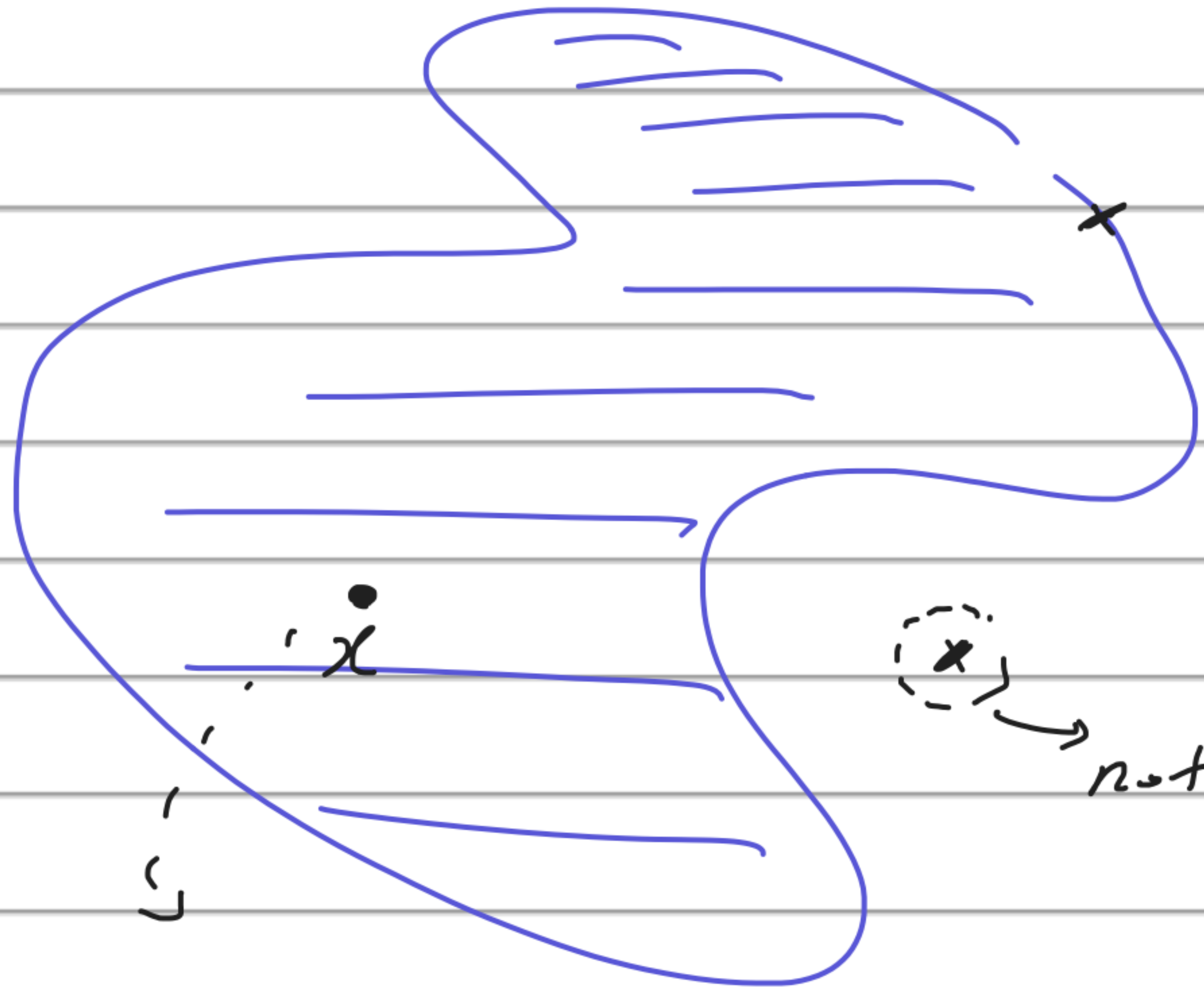
X $\qquad$ $cl(X)$

$x \in cl(X)$ if $\exists \{x_k\}_{k=1}^{\infty}$ s.t. $x_k \in X$

but $\lim_{k \to \infty} x_k = x$

x is in the closure if there is a

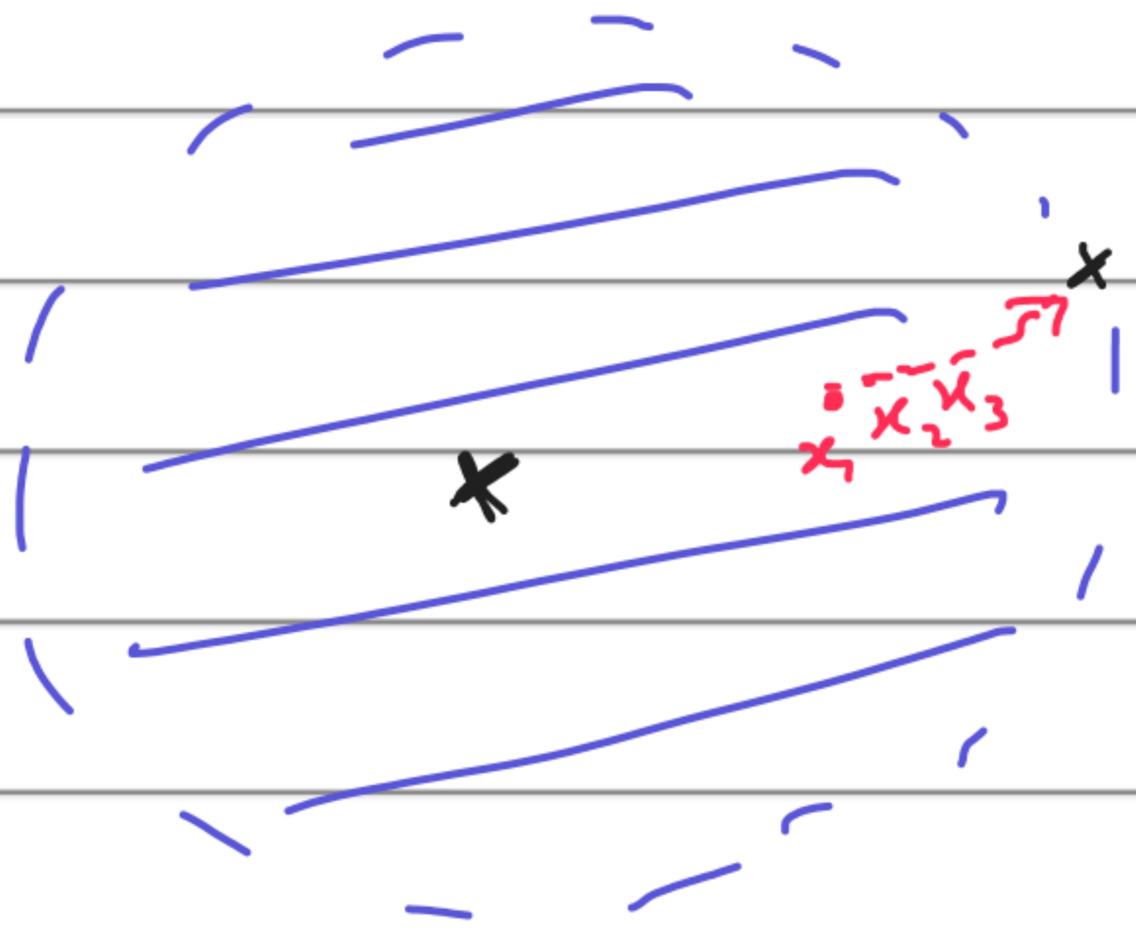
sequence of points in the set that converges to x.



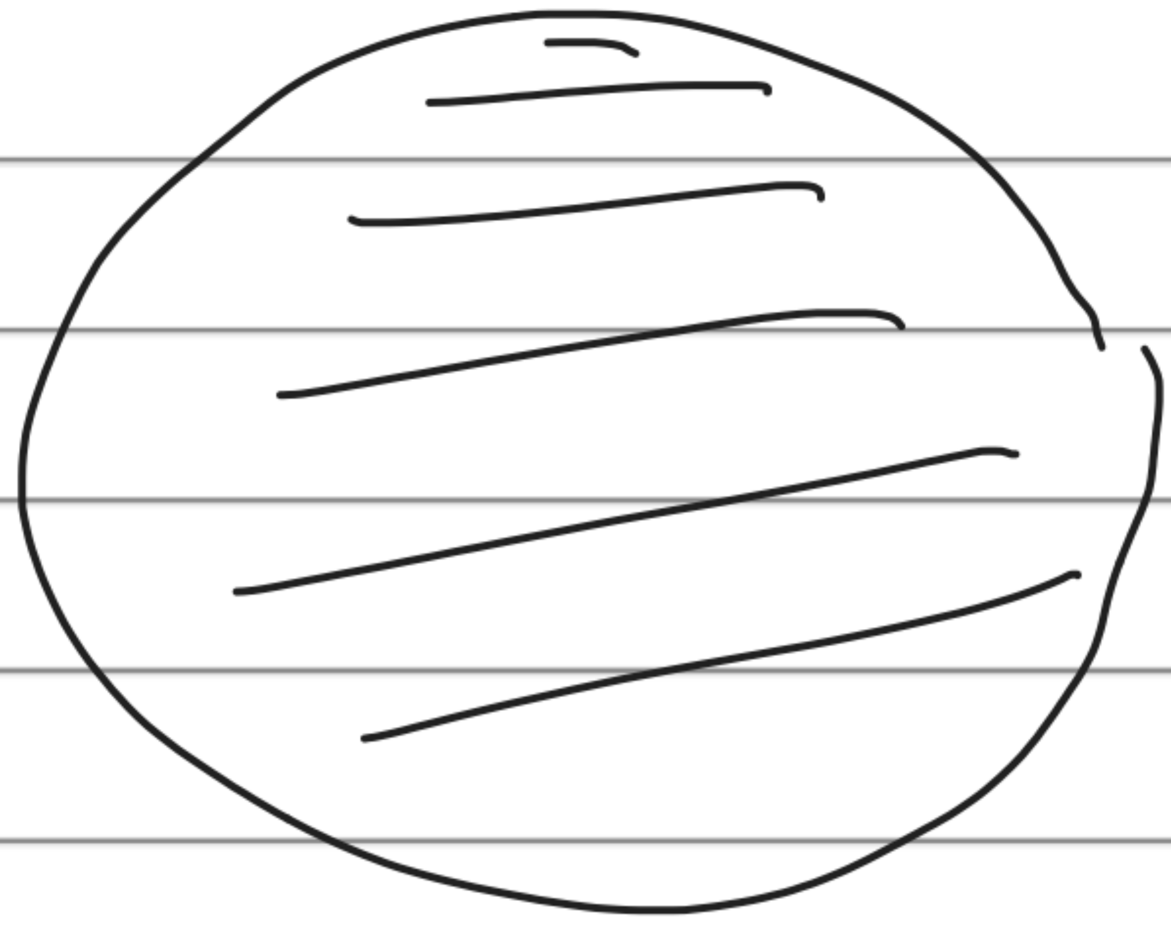
not in the closure

$\implies X = cl(X)$

$\underbrace{x_1}_{x}, \underbrace{x_2}_{x}, \ldots \longrightarrow x$



$\implies$

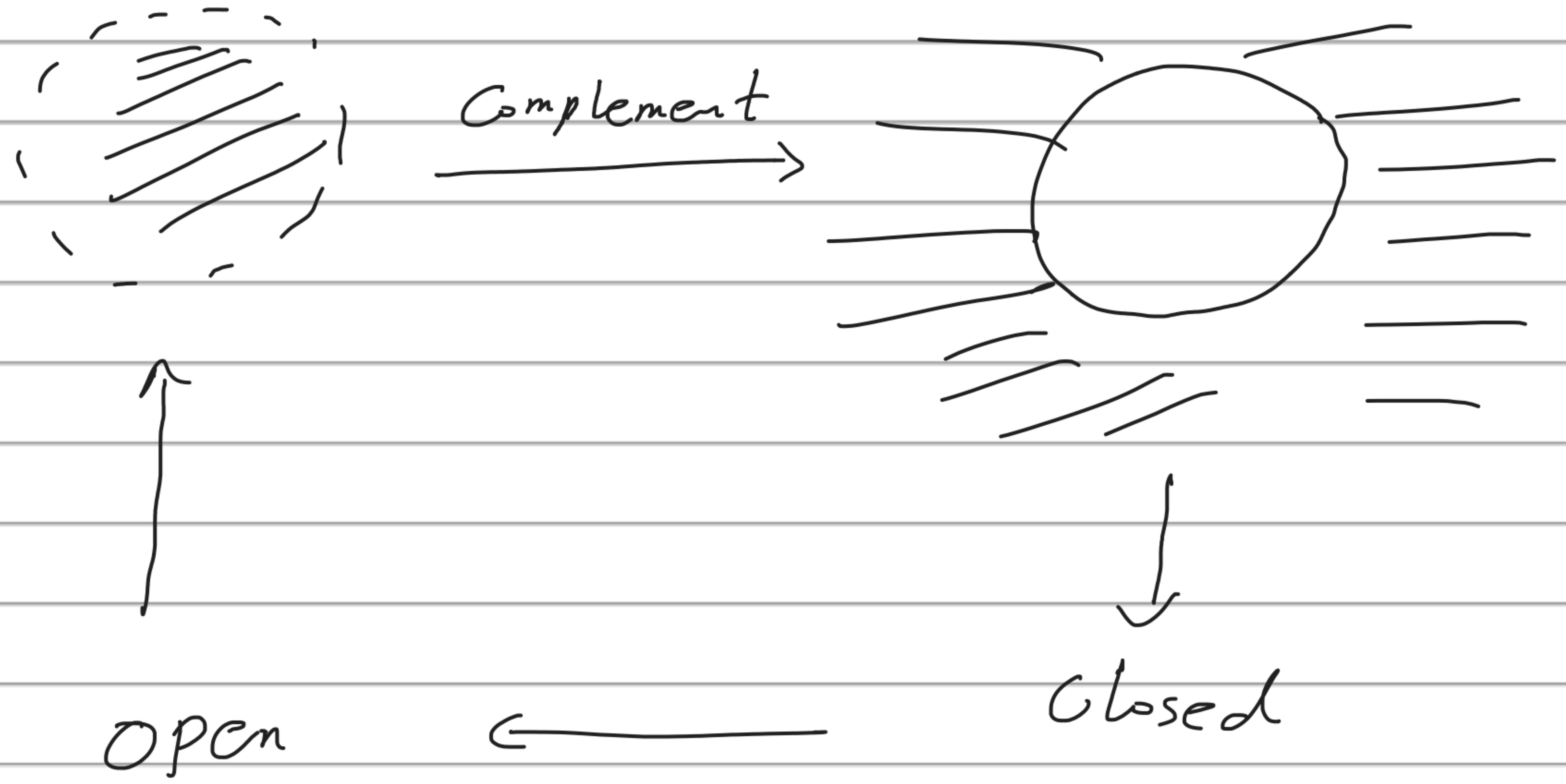


open ball

closed ball

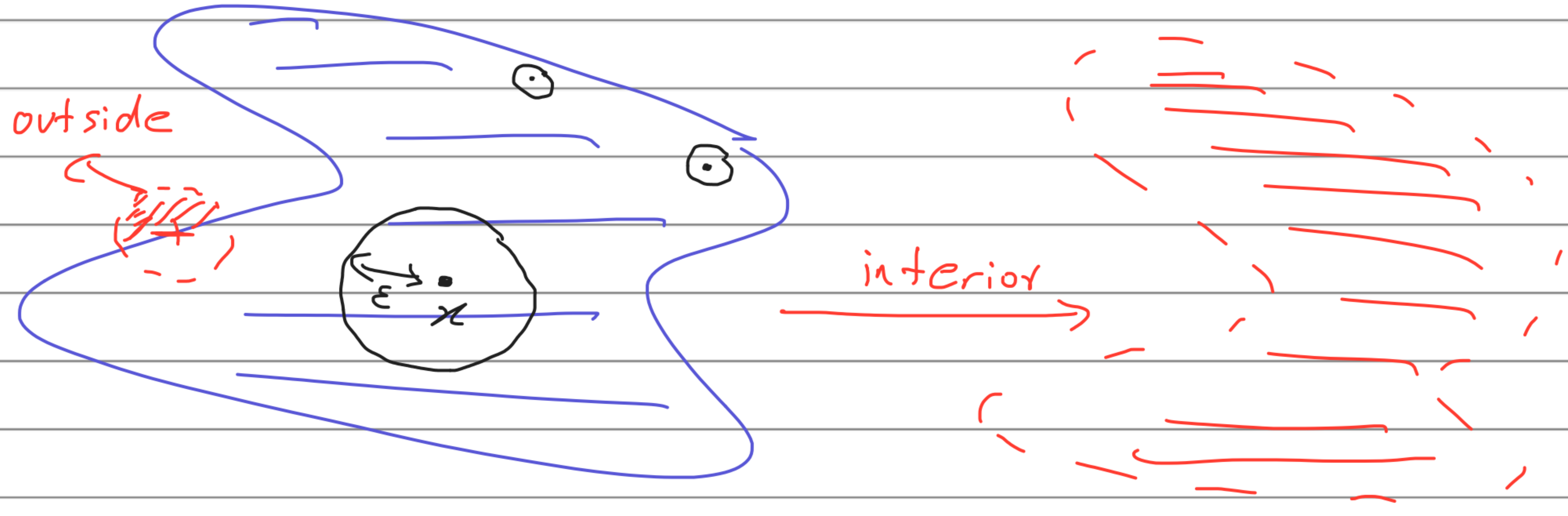
X : closed if $X = cl(X)$

X : open $\longrightarrow$ Complement of X = closed

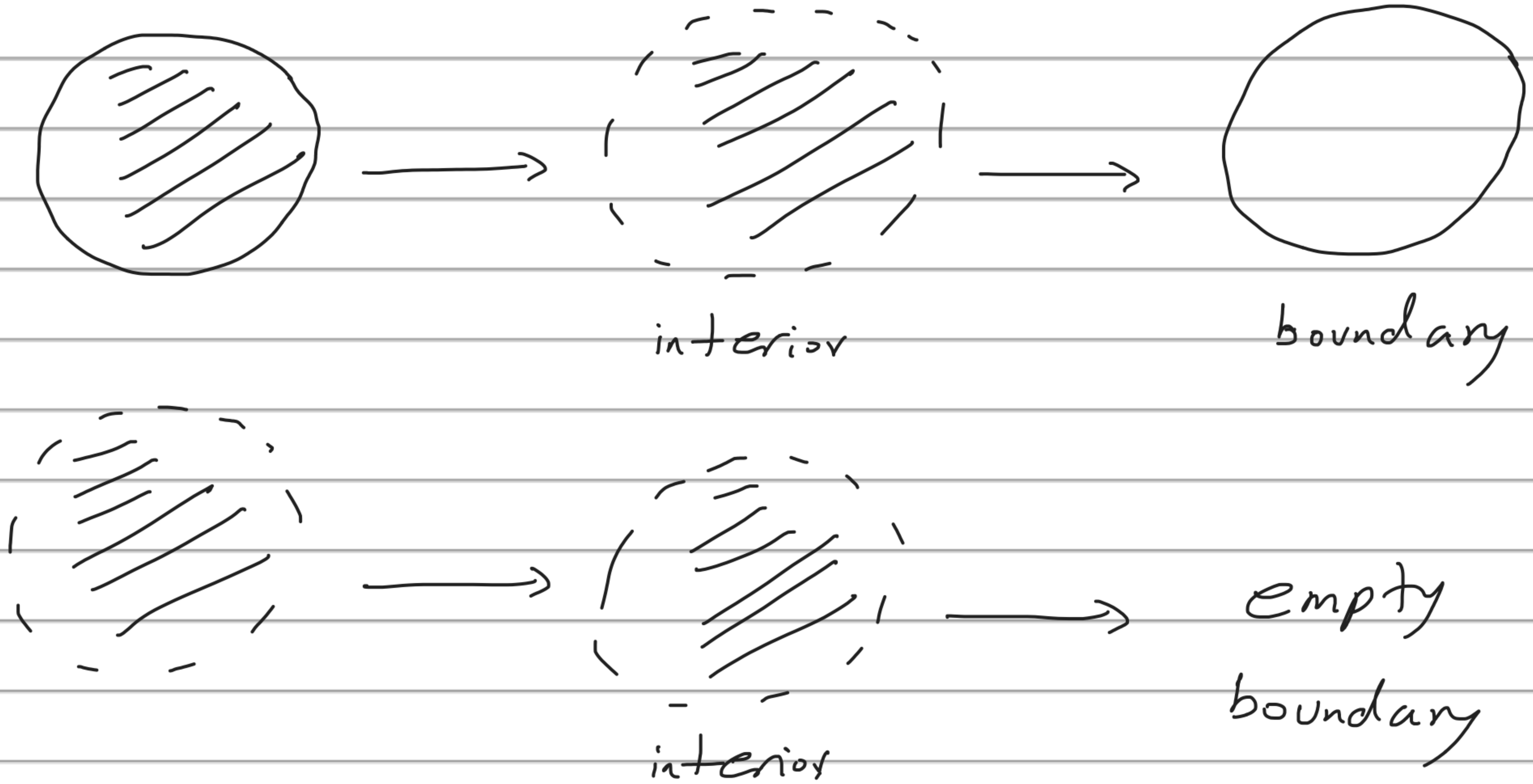


Interior point x : $x \in X$ and $\underbrace{\exists}_{\text{there exists}} \varepsilon > 0$

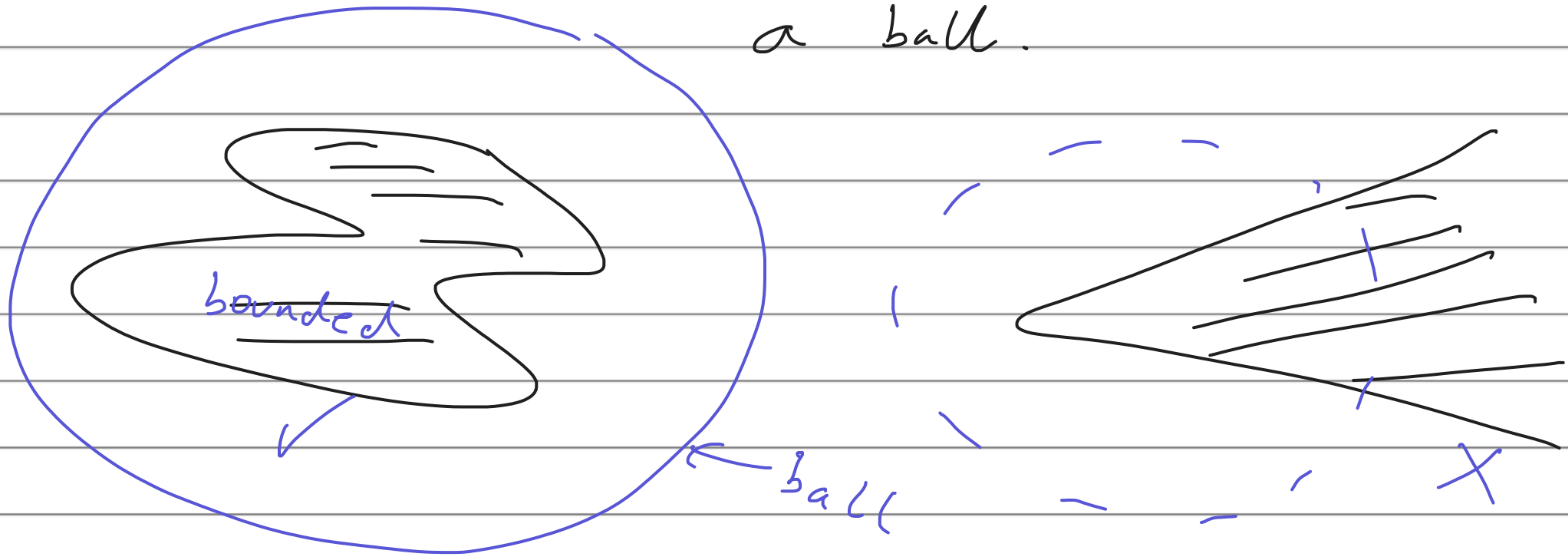
s.t. $\underbrace{\{ y \mid \|y - x\| < \varepsilon \}}_{\text{open ball}} \subseteq X$



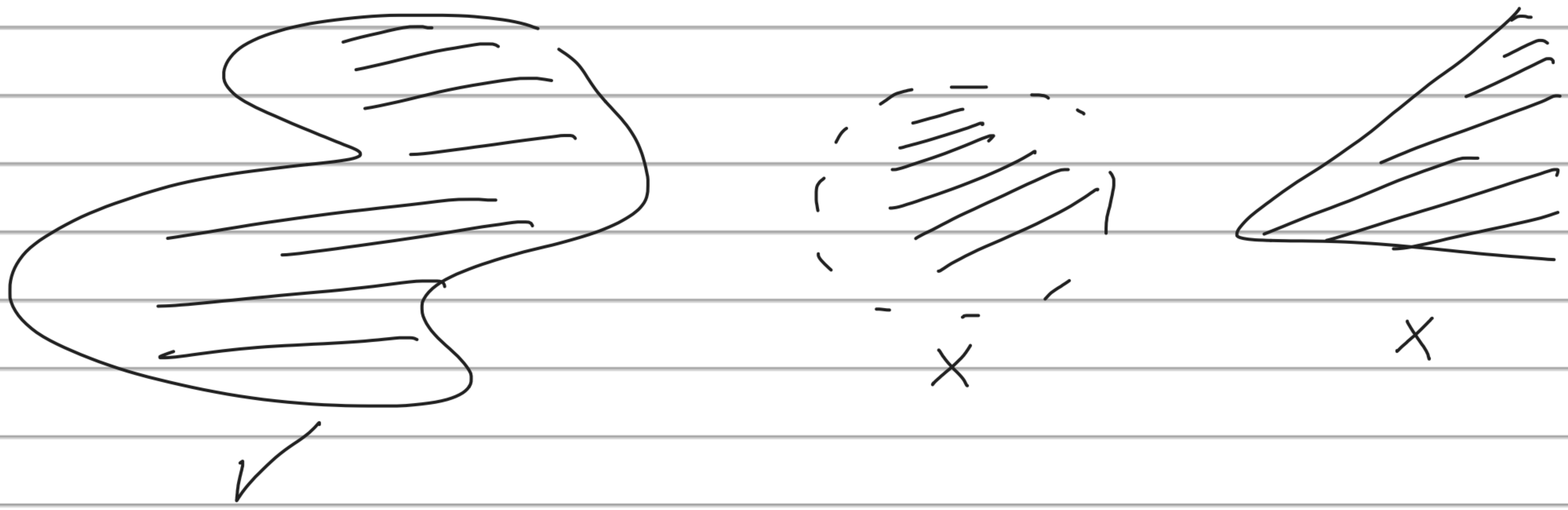
Boundary : X − interior



Bounded set : a set that fits in a ball.



Compact set : closed + bounded



$\min \; f(x) \quad \text{s.t.} \quad \cancel{x \in X}$

$\downarrow$

$\min \; f(x)$ unconstrained opt

solution : local min / strict local min / global min

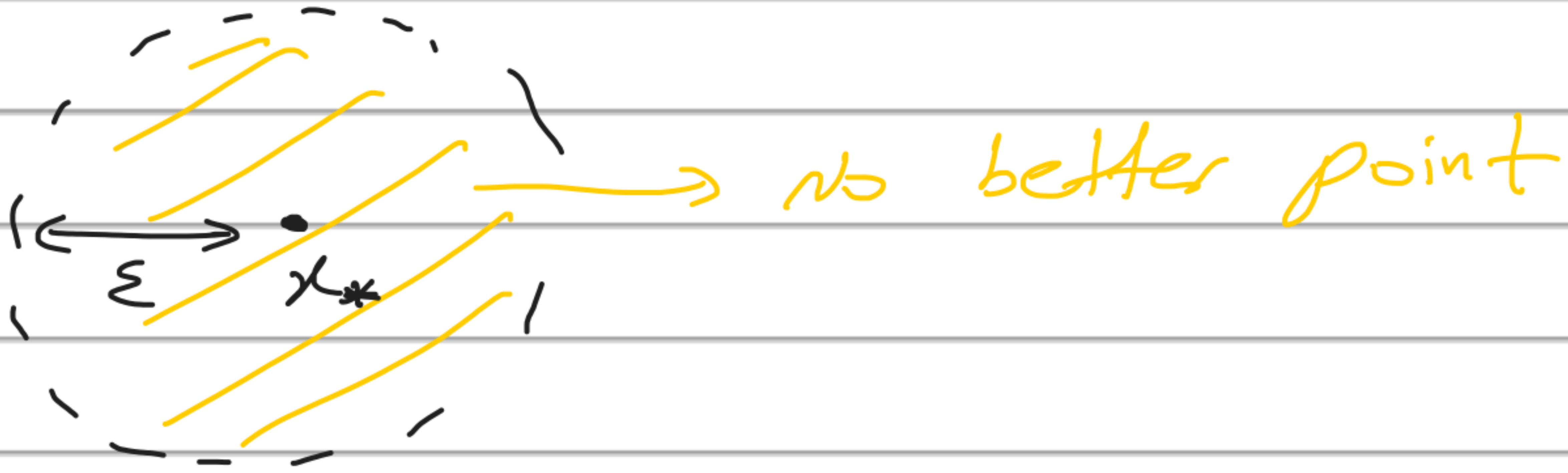
x_* : local min if $\exists \; \varepsilon > 0$

s.t. $f(x_*) \le f(x) \quad \forall x : \|x - x_*\| < \varepsilon$

open ball

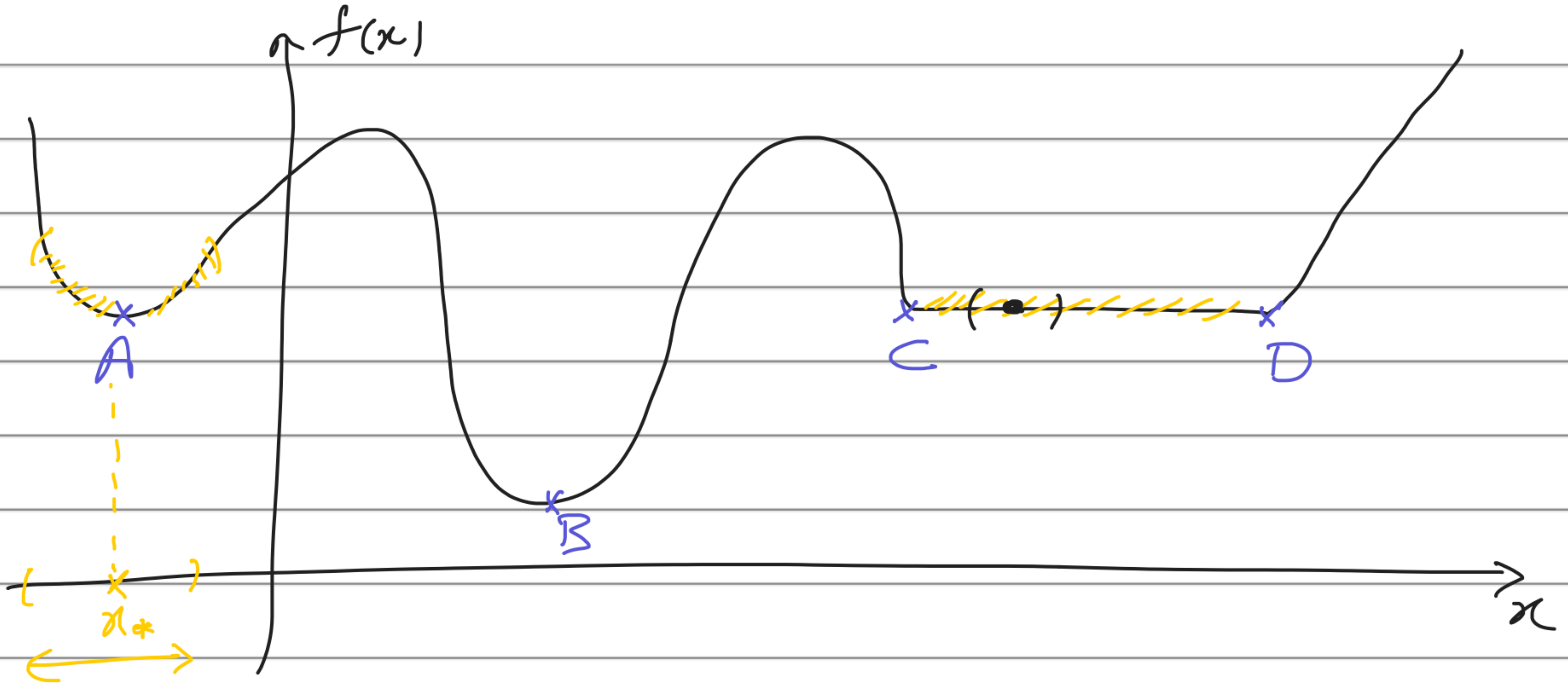
strict local min : $\exists\ \varepsilon > 0$ s.t

$$f(x_*) < f(x) \quad \forall x \neq x_* \ \&\ \|x - x_*\| < \varepsilon$$



global min : $f(x_*) \leq f(x) \quad \forall x$

($\leq \longrightarrow \geq$: min $\longrightarrow$ max)



A: strict local min

B: strict global min

, [C, D] = local min

(C, D) = local max

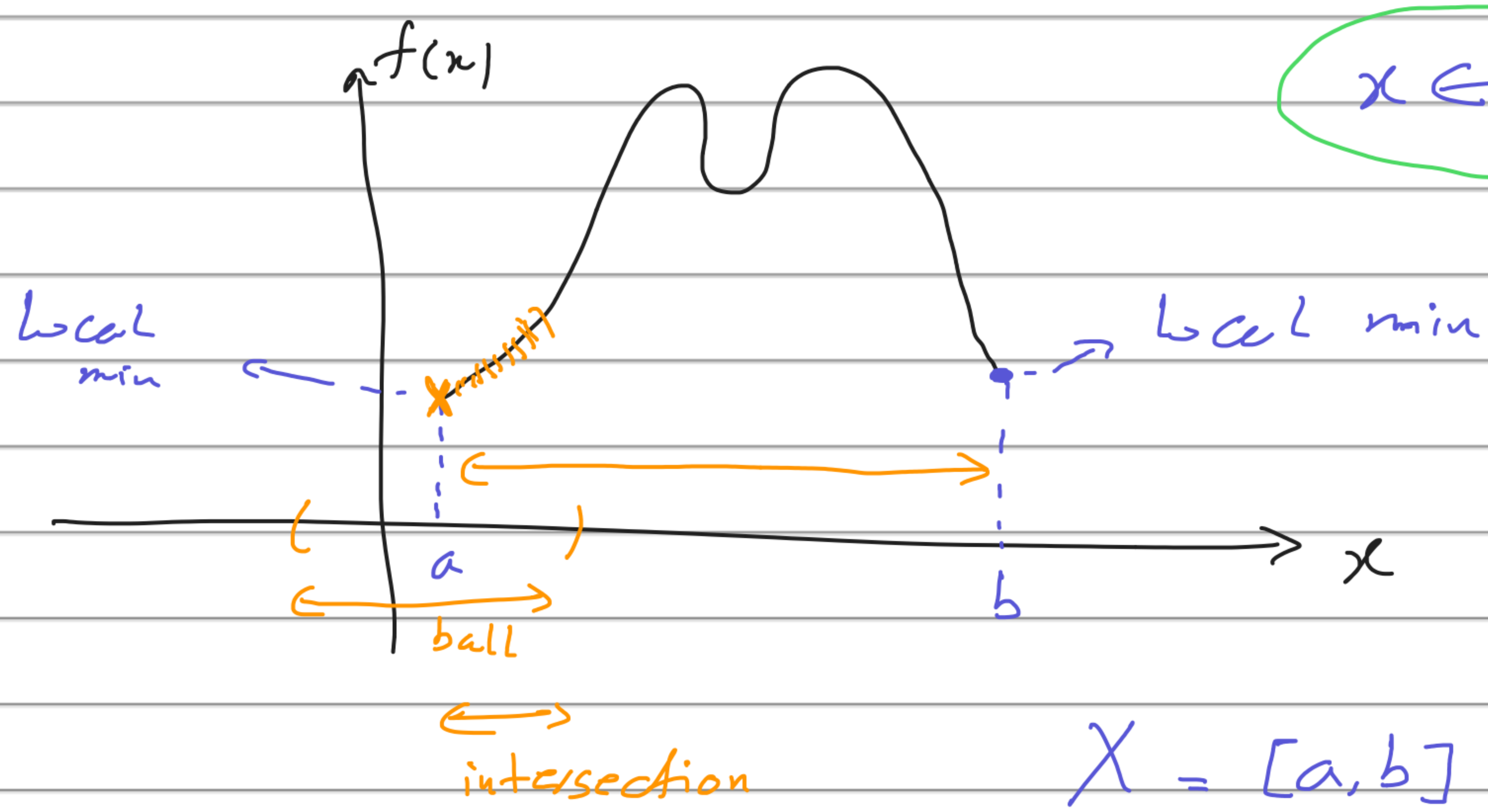
$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

x_* : local min if $\exists\ \varepsilon > 0$ s.t.

$f(x_*) \le f(x) \quad \forall x : \|x - x_*\| < \varepsilon$

and

$x \in X$



$X = [a, b]$

x_* : global min $\longrightarrow$ $f(x_*) \le f(x) \quad \forall x \in X$

for every

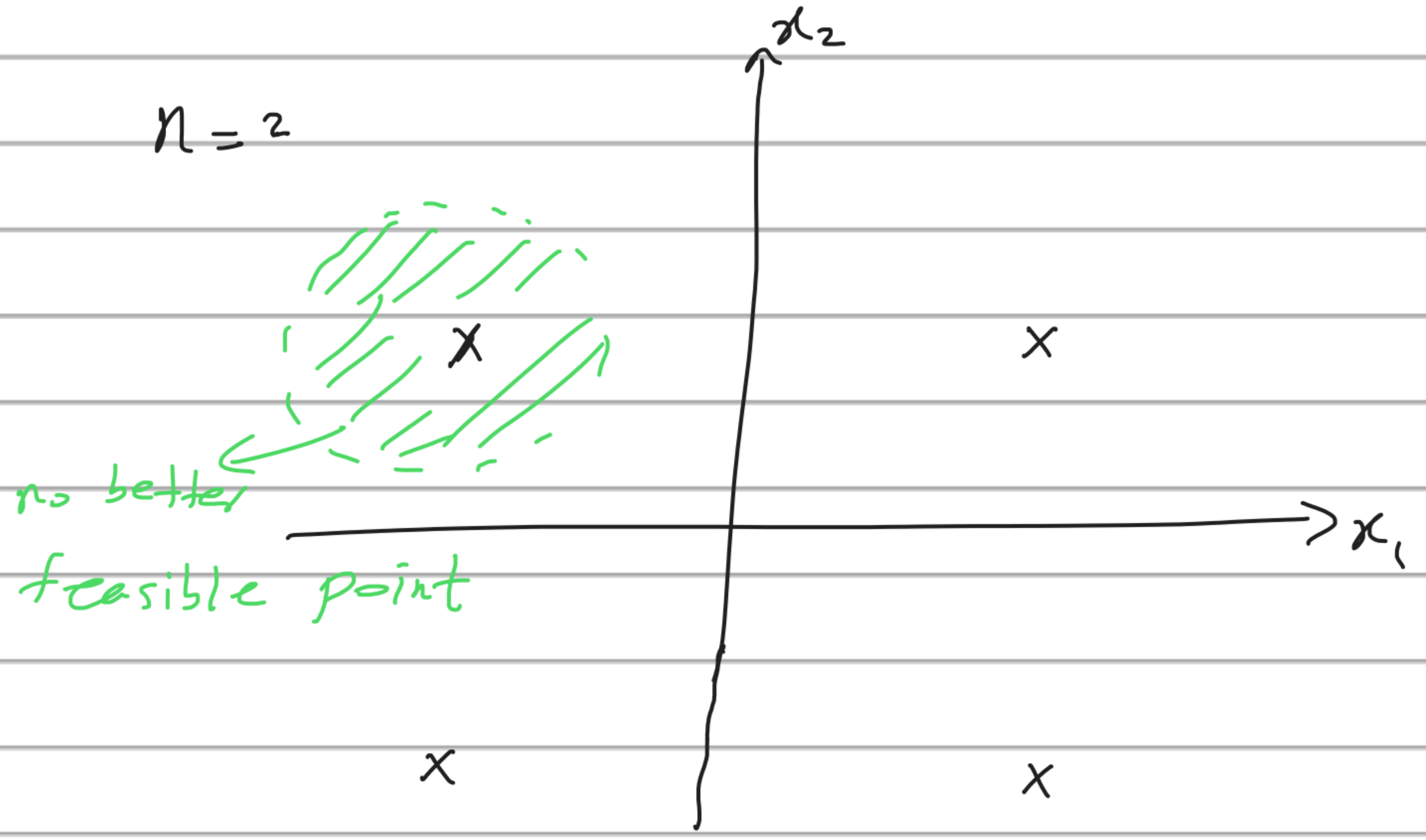
$$\min_{x \in \mathbb{R}^n} x^T Q x$$

s.t. $x_i \in \{-1, 1\}$ $\Longrightarrow$ $X = 2^n$ points

$i = 1, \dots, n$

$$\begin{bmatrix} \pm 1 \\ \pm 1 \\ \vdots \\ \pm 1 \end{bmatrix}$$

$n=2$



all feasible points = local minima

$$\min_{x \in \mathbb{R}^n} f(x) \quad : \text{ unconstrained opt}$$

- If x_* is a local min $\longrightarrow \nabla f(x_*) = 0$

FOC : first-order condition

$$\boxed{\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \end{bmatrix}}$$

(Three proofs)

- Taylor series approximation:

$$f(\underbrace{x_*}_{\text{nominal}} + \underbrace{\Delta x}_{\text{perturbation/direction}}) = f(x_*) + \nabla f(x_*)^T \Delta x + \underbrace{\cdots\cdots}_{\substack{\text{h.o.t} \\ O(\|\Delta x\|^2)}}$$

first-order approximation

$\nabla^2 f =$ i, j: $\frac{\partial^2 f}{\partial x_i \partial x_j}$ (symmetric)

$$f(x_* + \Delta x) = f(x_*) + \nabla f(x_*)^T \Delta x + \frac{1}{2} \Delta x^T \nabla^2 f(x_*) \Delta x + \underbrace{\cdots\cdots}_{O(\|\Delta x\|^3)}$$

second-order approximation

- Truncated Taylor:

(mean-value theorem)

$$f(x_* + \Delta x) = f(x_*) + \nabla f(\cancel{x_*})^T \Delta x + \cancel{\cdots\cdots}$$

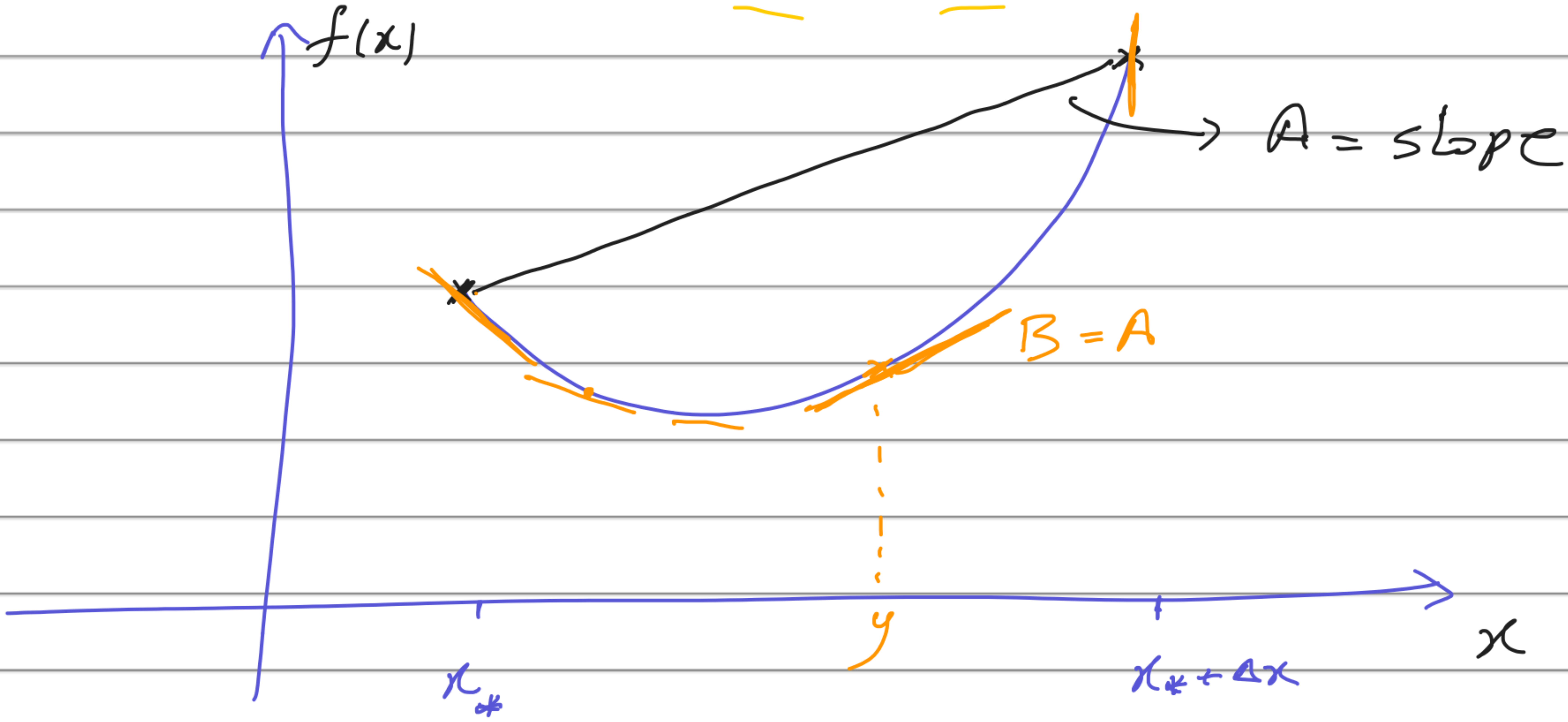
$y \in [x_*, x_* + \Delta x]$

$n = 1$:

$\exists\, y$ in the interval x_* & $x_* + \Delta x$ s.t.

$$\boxed{f(x_* + \Delta x) = f(x_*) + f'(y)\,\Delta x}$$

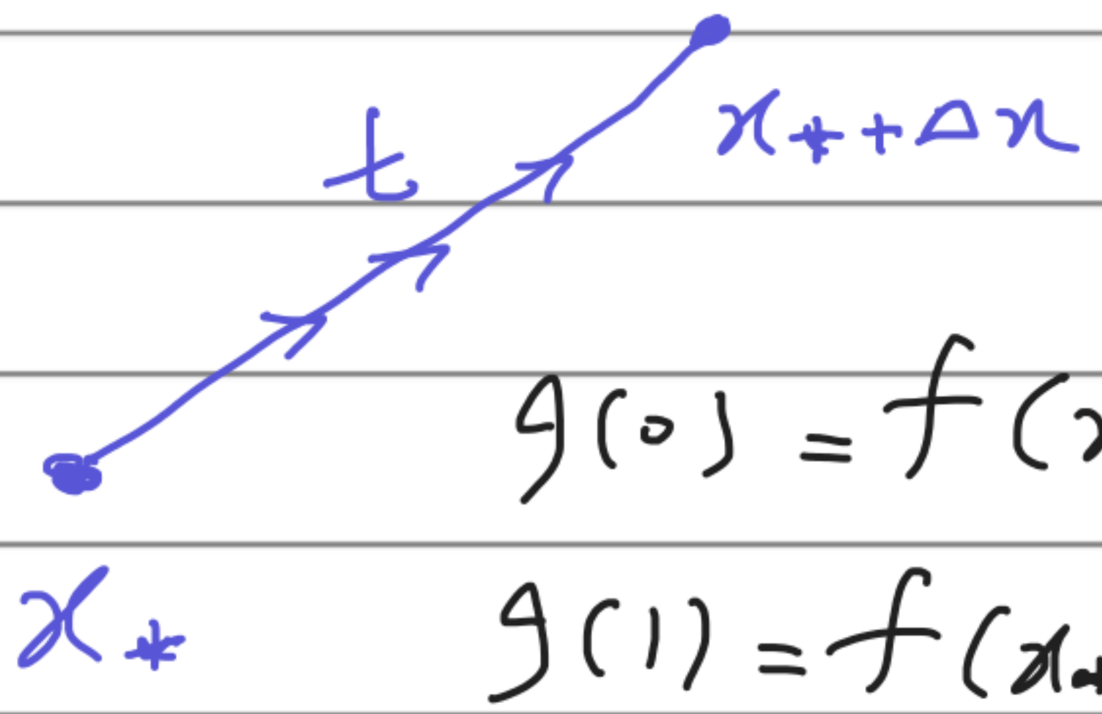
$$\Rightarrow \underbrace{f'(y)}_{B} = \underbrace{\frac{f(x_* + \Delta x) - f(x_*)}{\Delta x}}_{A}$$



Generalize to $\underline{\underline{n}}$ variables :

reduction / restriction

$$\underbrace{g(t)}_{\text{one variable}} = \underbrace{f(x_* + t\,\Delta x)}_{n \text{ variables}}$$



$g(0) = f(x_*)$

$g(1) = f(x_* + \Delta x)$

Mean-value Thm to $g(t)$:

$$\underbrace{g(1)}_{f(x_*+\Delta x)} = \underbrace{g(0)}_{f(x_*)} + g'(s)(1-0) \qquad s \in [0,1]$$

$$g(t) = f(\underbrace{x_* + t\Delta x}) \Longrightarrow g'(s) = \nabla f(x_* + s\Delta x)^T \times \Delta x$$

$$g'(s) = \nabla f(\underbrace{x_* + s\Delta x}_{y \in [x_*, x_*+\Delta x]})^T \Delta x$$

Extended mean-value :

$$f(x_* + \Delta x) = f(x_*) + \nabla f(x_*)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(\cancel{x_*})\Delta x + \cancel{\cdots}$$

$y \in [x_*, x_*+\Delta x]$

- If x_* is a local min for $\min f(x)$

  then $\nabla f(x_*) = 0$

proof $\underline{\underline{1}}$:

$$f(x_* + \Delta x) = f(x_*) + \nabla f(x_*)^T \Delta x + \underbrace{\cdots}_{O(\|\Delta x\|^2)}$$

Assume $\nabla f(x_*) \neq 0 \longrightarrow \exists\, i \in \{1, \dots, n\}$

s.t. $\dfrac{\partial f}{\partial x_i}(x_*) \neq 0$ $\quad \left(\nabla f(x_*) = \begin{bmatrix} 0 \\ 0 \\ |\cdot| \\ 0 \\ \vdots \end{bmatrix} \leftarrow i \right)$

$$\Delta x = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ -\varepsilon \times \dfrac{\partial f}{\partial x_i}(x_*) \\ \vdots \\ 0 \end{bmatrix} \rightarrow i^{th} \text{ coordinate}$$

$$\Longrightarrow f(x_* + \Delta x) = f(x_*) + \left(-\varepsilon \times \boxed{+}\right) + O(\varepsilon^2)$$

dominates if ε: small

$$\Longrightarrow f(x_* + \Delta x) < f(x_*)$$

if Δx: small & along a good direction

Contradiction!

proof $\Leftarrow$:

$$f(x_* + \Delta x) = f(x_*) + \nabla f(y)^T \Delta x$$

$y \in [x_*, x_* + \Delta x]$

Assume $\Delta x \to 0$

$\Longrightarrow y \to x_*$

$\Longrightarrow \nabla f(y) \to \nabla f(x_*)$

If $\nabla f(x_*) \neq 0 \Longrightarrow \exists \, \Delta x$ small s.t.

$$\nabla f(x_*)^T \Delta x < 0 \Longrightarrow \nabla f(y)^T \Delta x < 0$$

$$\Longrightarrow \underbrace{f(x_* + \Delta x)}_{\text{new point}} < \underbrace{f(x_*)}_{\text{old point}} \quad \text{(Contradiction)}$$