262B-Lecture 3

Date created: 2021.01.28
N. of Pages: 14

$\min f(x) \longrightarrow$ FOC $\nabla f(x_*) = 0$

$\downarrow$

Algorithms

$x^{(0)} \longrightarrow x^{(1)} \longrightarrow x^{(2)} \longrightarrow \cdots$

$x^{(k+1)} = x^{(k)} + \underbrace{\alpha^{(k)}} \underbrace{\Delta x^{(k)}}$

exact line search

Limited line search

Backtracking

descent if $\nabla f(x^{(k)})^T \Delta x^{(k)} < 0$

$\alpha > 0 \;, \quad \beta \in (0,1) \longrightarrow$

α

$\alpha\beta$

$\alpha\beta^2$

$\vdots$

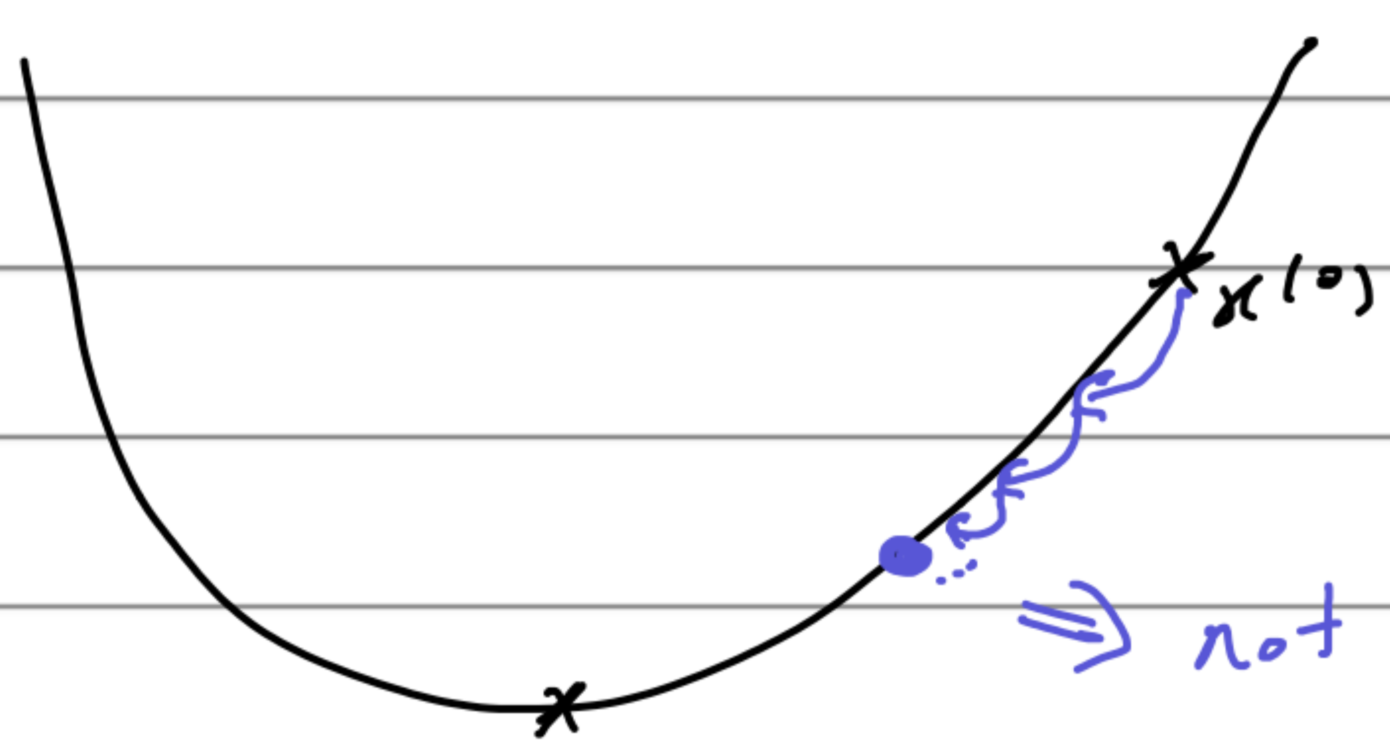
$\alpha\beta^t$

$\vdots$

$\downarrow$ pick $\alpha^{(k)}$

previously : $f(x^{(k+1)}) < f(x^{(k)})$

improvement : $\underbrace{f(x^{(k+1)}) - f(x^{(k)})}_{\text{Considerable}} < 0$



$\Rightarrow$ not enough improvement

x_*

Armijo rule: pre-set parameter $0 < \sigma < 1$

$\alpha^{(k)}$: α, $\alpha\beta^1$, $\alpha\beta^2$, $\vdots$, $\alpha\beta^t$, $\vdots$

$$\Longrightarrow \quad f(x^{(k+1)}) - f(x^{(k)}) < \underbrace{\sigma}_{\text{constant}} \times \underbrace{\nabla f(x^{(k)})^T \Delta x^{(k)}}_{\text{inner product of gradient \& direction}} \times \alpha^{(k)} \qquad (*)$$

perturbation

improvement: 10^{-5}, 1, 10^{3} big / small

Compare against the inner product

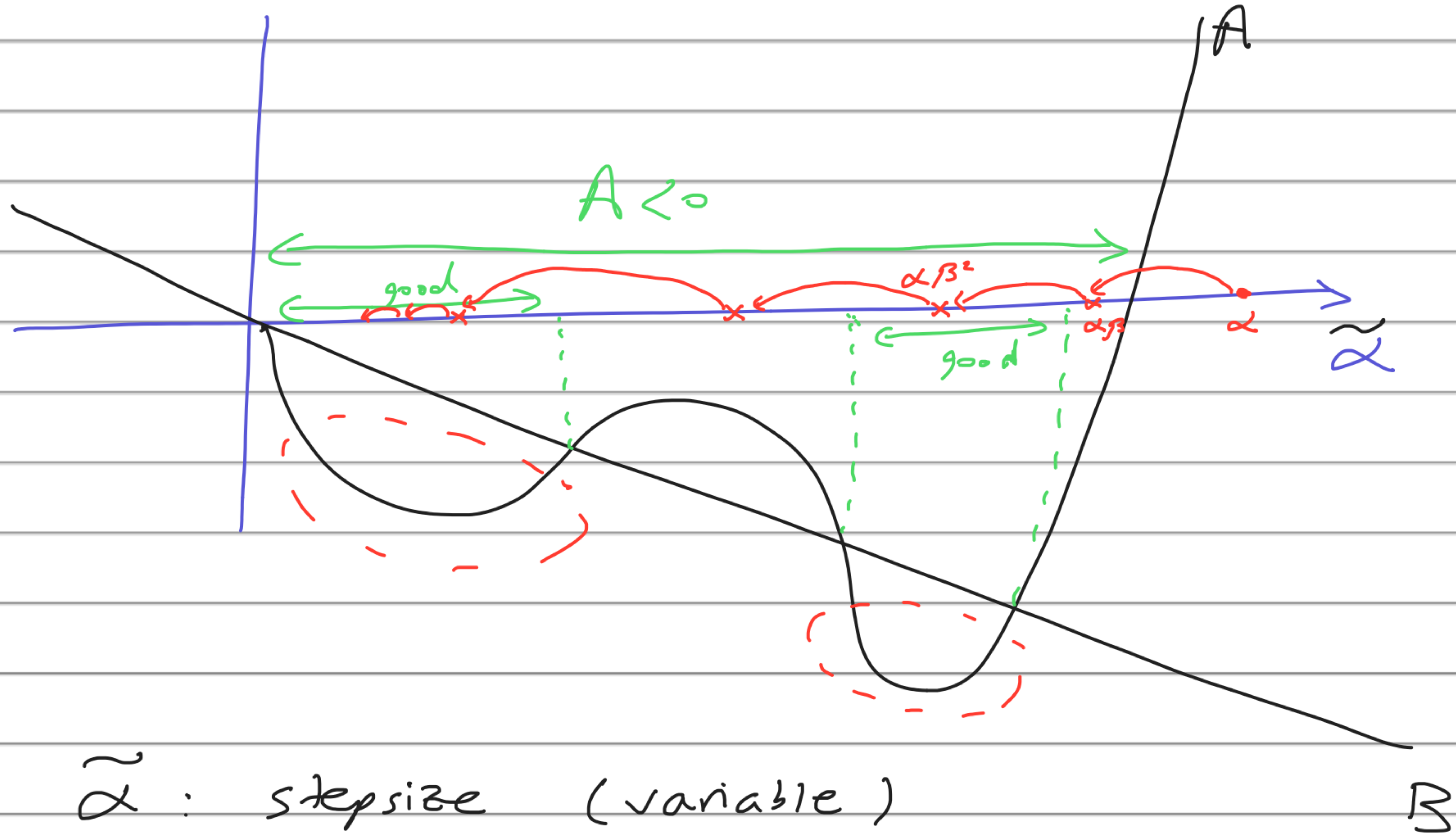
$$\underbrace{f(\underbrace{x^{(k+1)}}_{\text{new point}}) - f(\underbrace{x^{(k)}}_{\text{old point}})}_{\text{improvement}} = \underbrace{\nabla f(x^{(k)})^T \Delta x^{(k)} \times \alpha^{(k)} + \cdots}_{\text{same order as first term}}$$

Armijo: pick smallest $\underline{\underline{t}}$ s.t. $\alpha^{(k)} = \alpha\beta^t$

satisfies the inequality $(*)$.

Thm: If $t >$ threshold, then $(*)$ is always satisfied. $t \uparrow +\infty \Rightarrow \alpha^{(k)} \downarrow 0$

$$\underbrace{f(x^{(k)} + \underbrace{\tilde{\alpha}}_{\text{stepsize}} \Delta x^{(k)}) - f(x^{(k)})}_{A} \quad \text{versus} \quad \underbrace{\sigma \times \nabla f(x^{(k)})^T \times (\underbrace{\tilde{\alpha} \Delta x^{(k)}}_{\text{linear}})}_{B}$$



$\tilde{\alpha}$: stepsize (variable)

$\alpha, \alpha\beta, \alpha\beta^2, \alpha\beta^3, \ldots$

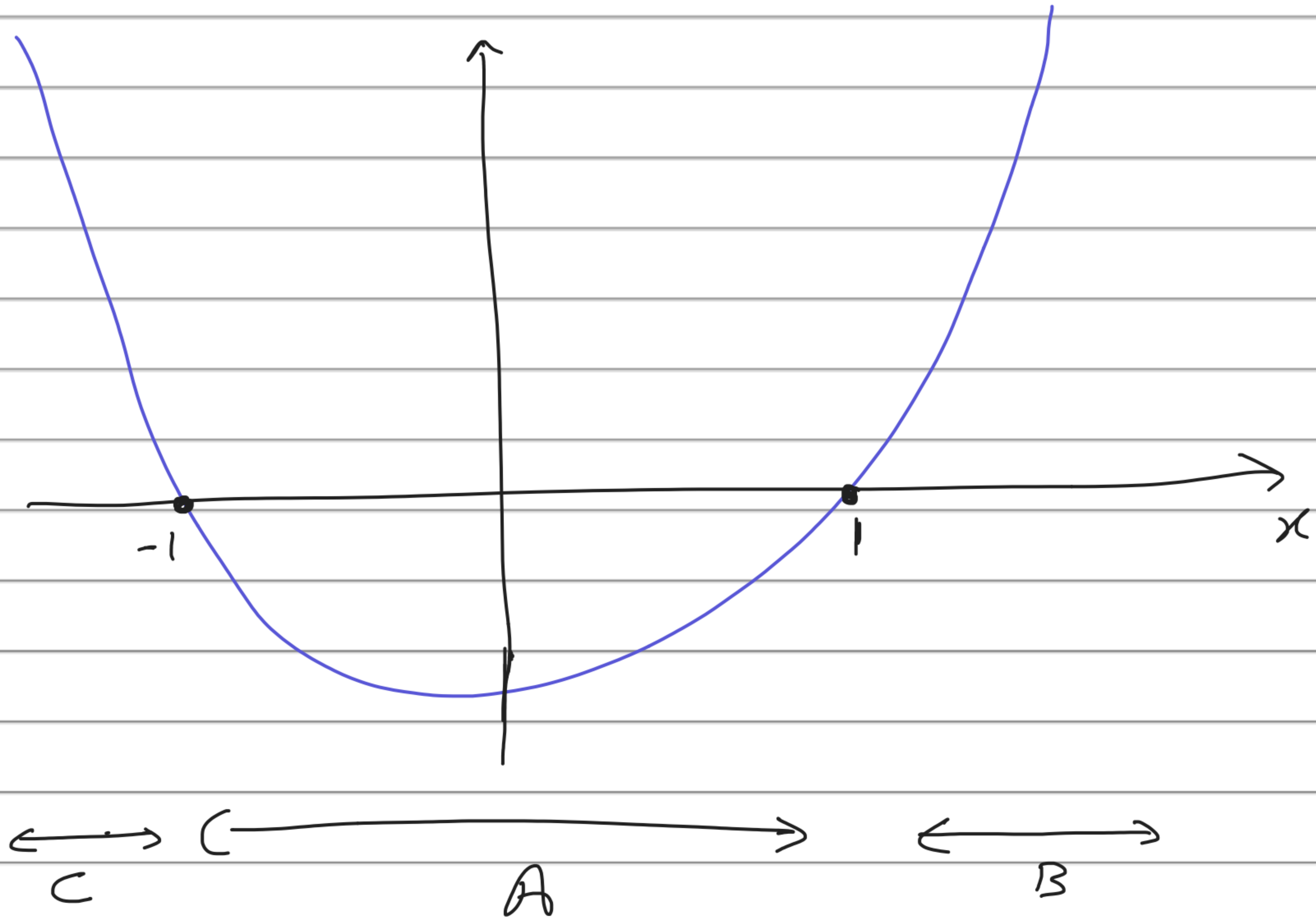
$A < B$

armijo

So far, $\alpha^{(k)}$: varying

- what if $\alpha^{(k)} = \alpha$ = fixed

Gradient algorithm: $x^{(k+1)} = x^{(k)} - \alpha \nabla f(x^{(k)})$



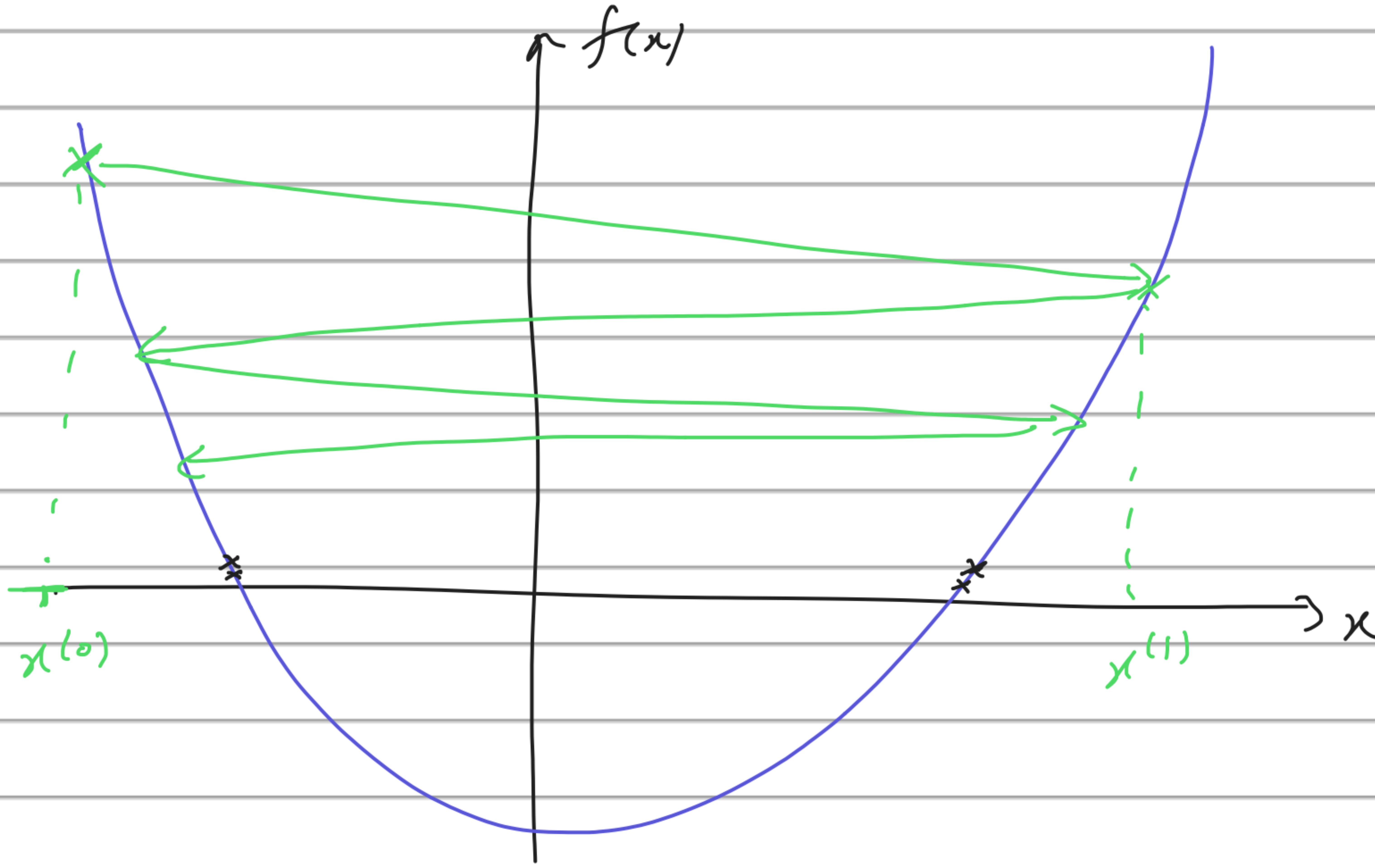
A: $(-1 \leq x \leq 1)$: $f(x) = x^2 - 1$

B: $x \geq 1$: $f(x) = \frac{3(1-x)^2}{4} - 2(1-x)$

C: $x \leq -1$: $f(x) = \frac{3(1+x)^2}{4} - 2(1+x)$

$\boxed{\alpha^{(k)} = \alpha = 1}$ $\rightarrow$ Gradient algorithm

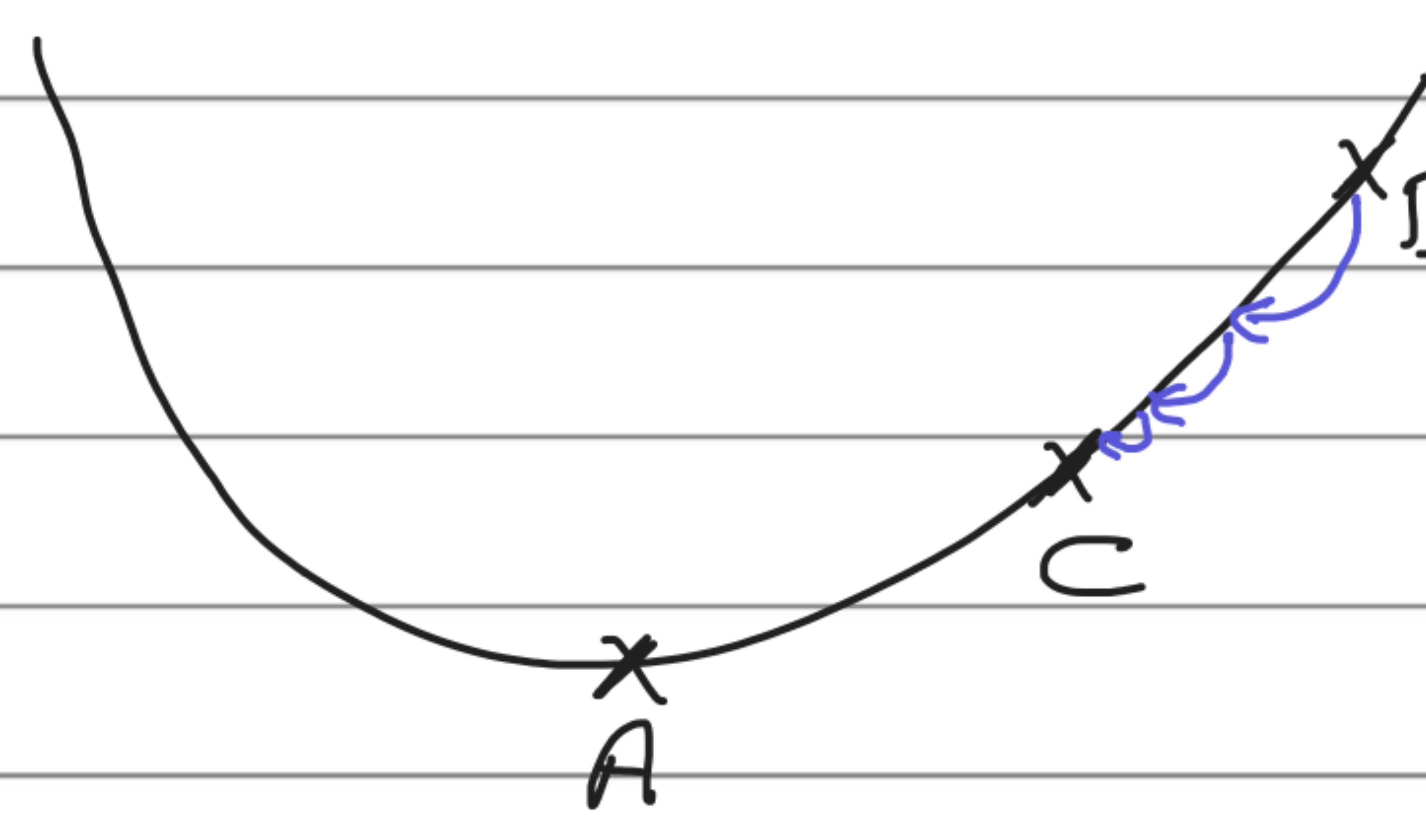
$|x^{(0)}| > 1$



$x^{(0)} \rightarrow x^{(1)} \rightarrow x^{(2)} \rightarrow \cdots\cdots \rightarrow$ get close to points $+1, -1$ in an oscillatory way

$\Rightarrow$ sequence won't converge, has two limit points.

Diminishing stepsize :



$\alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}, \ldots \rightarrow 0$

If $\alpha^{(k)} \to 0$ but $\sum_{k=1}^{\infty} \alpha^{(k)} = +\infty$

$\Rightarrow$ It's going to work.

$\alpha^{(k)} = \frac{1}{k}$ ✓ $\qquad \alpha^{(k)} = \frac{1}{k^2}$ ✗

Proof: Assume we have convergence:

$\lim_{k\to\infty} x^{(k)} = \bar{x} \overset{?}{=}$ stationary point (FOC)

pick m, l (integers) s.t. $m > l \to \infty$

$x^{(l)} \simeq x^{(l+1)} \simeq \dots \simeq x^{(m)} \simeq \bar{x}$

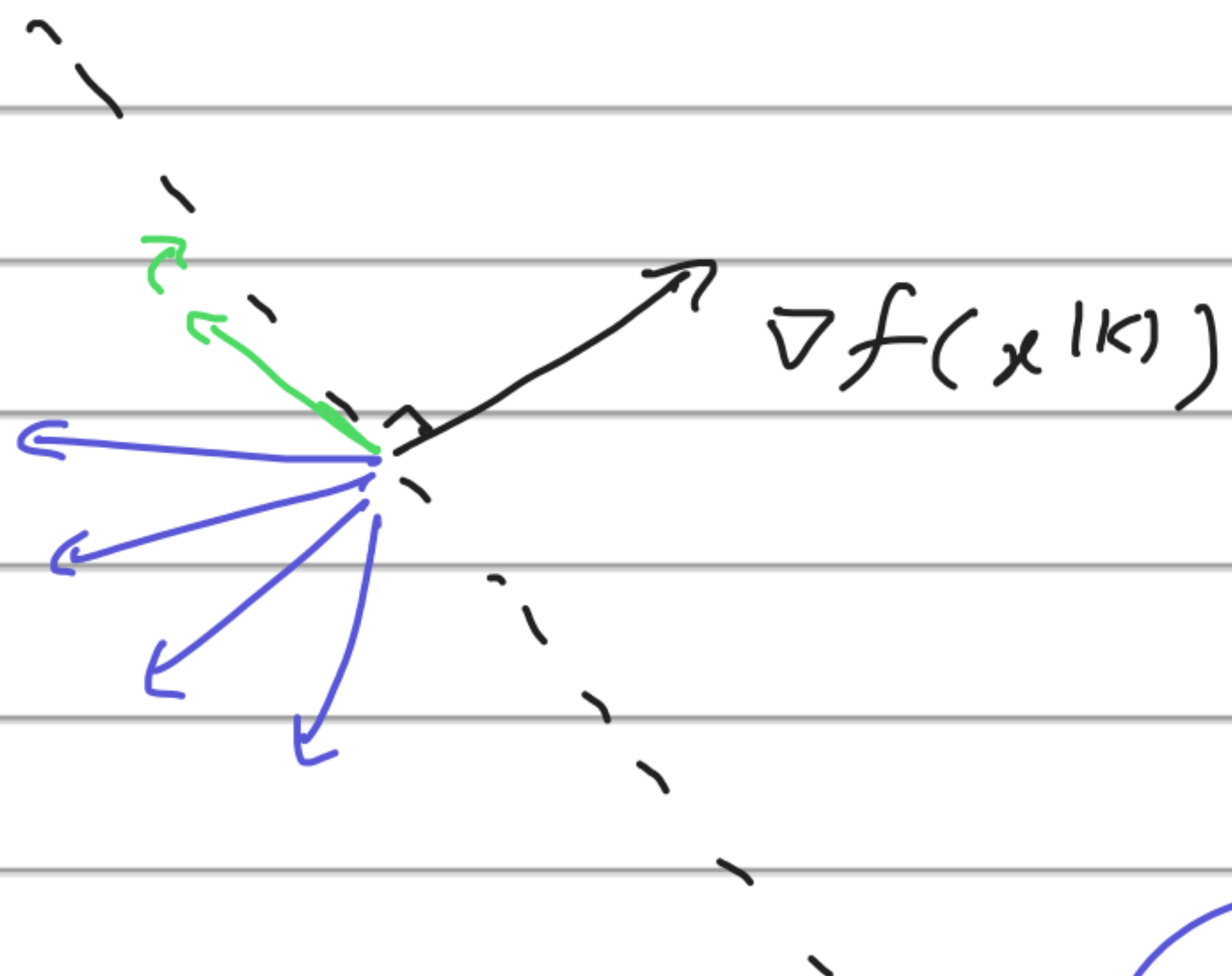
$x^{(0)} \to x^{(1)} \to \dots \to \underbrace{x^{(l)} \to \dots \to x^{(m)}} \to \dots \to \bar{x}$

$\cancel{x^{(m)}} \simeq \underbrace{\cancel{x^{(l)}} - \underbrace{\sum_{k=l}^{m-1} \alpha^{(k)}}_{\to\infty} \nabla f(\bar{x})}_{\to 0}$

$\Rightarrow \nabla f(\bar{x}) = 0$

Convergence issue:

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} \underbrace{\Delta x^{(k)}}_{\text{descent}}$$



(*)

$$\underbrace{\nabla f(x^{(k)})^T \Delta x^{(k)}}_{\text{sequence}} < 0$$

$-1, -0.5, -0.1, \ldots, -10^{-5}, \ldots \to 0$

direction is becoming nearly orthogonal to gradient, that would be a problem.

$$\lim_{k \to \infty} \text{angle between direction and gradient} \neq 90^\circ$$

$$\lim_{k \to \infty} \frac{\nabla f(x^{(k)})^T \Delta x^{(k)}}{\|\nabla f(x^{(k)})\| \, \|\Delta x^{(k)}\|} < 0 \qquad (**)$$

(**) : has a limit

Gradient related : $\{\Delta x^{(k)}\}_{k=1}^{\infty}$ is

gradient related if for any sub-sequence

$\{x^{(k)}\}_{k\in K}$ s.t. $\{x^{(k)}\}_{k\in K}$ converges to

a non-stationary point , then $\{\Delta x^{(k)}\}_{1}^{\infty}$ is bounded and

$$\lim_{k\to\infty,\ k\in K} \sup \underbrace{\nabla f(x^{(k)})^T \Delta x^{(k)}}_{} < 0$$

$$x^{(0)} \longrightarrow x^{(1)} \longrightarrow x^{(2)} \longrightarrow \cdots\cdots$$

$\Longrightarrow K$

$+1, -1, +1, -1, +1, -1$: Example

$$\Delta x^{(k)} = -\underbrace{D^{(k)}}_{>0} \nabla f(x^{(k)})$$

scaling : related to gradient

$$\underbrace{c_1 \|\nabla f(x^{(k)})\|^{p_1} \leq \text{eigs of } D^{(k)} \leq c_2 \|\nabla f(x^{(k)})\|^{p_2}}$$

If $\exists\ \underbrace{c_1}_{>0}, \underbrace{p_1}_{\geq 0}, \underbrace{c_2}_{>0}, \underbrace{p_2}_{\geq 0} \implies$ gradient related

For example: $\Delta x^{(k)} = -\nabla f(x^{(k)})$

Gradient algorithm

$\implies D^{(k)} = I$ ✓ $\quad (p_1 = p_2 = 0)$

$x^{(0)} \to x^{(1)} \to x^{(2)} \to \cdots\cdots \to$ stop ?

$$\Big(x^{(k+1)} = x^{(k)} + \underbrace{\alpha^{(k)}}_{\substack{\text{line}\\\text{search}}} \underbrace{\Delta x^{(k)}}_{\substack{\text{gradient}\\\text{related}}} \Big)$$

or backtracking (armijo)

Stopping criterion ?

stop if $\| \nabla f(x^{(k)}) \| \leq \varepsilon$ ① ✓

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

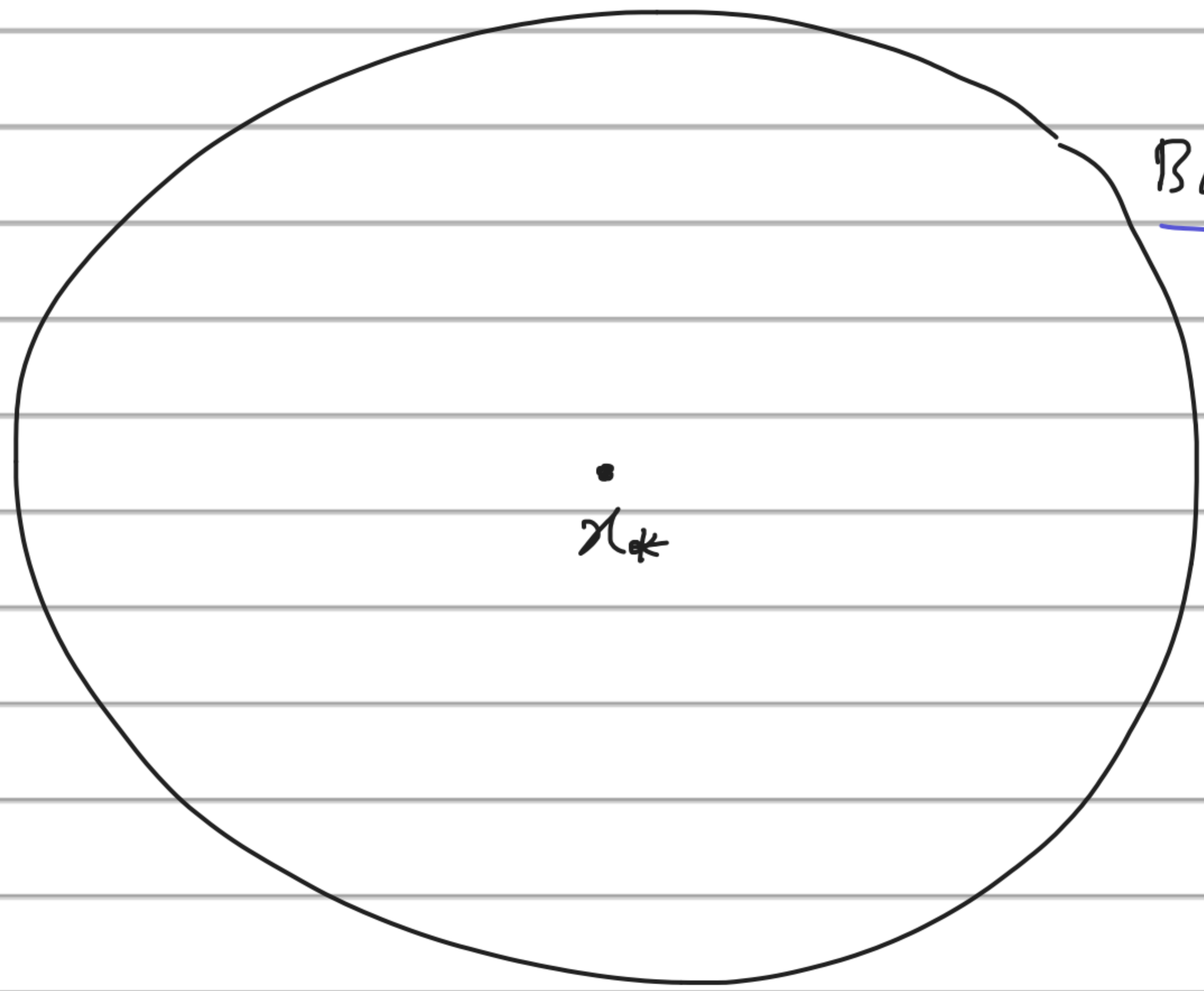
$\min f(x)$ v.s. $\min 100 f(x)$

$x^{(k)}$: $\underbrace{\nabla f(x^{(k)})}_{10^{-2}}$ v.s. $\underbrace{100 \nabla f(x^{(k)})}_{1}$

→ no direct relationship between gradient & point.

easy fix : $\dfrac{\| \nabla f(x^{(k)}) \|}{\underbrace{\| \nabla f(x^{(0)}) \|}} < \varepsilon$ ②

$\| x^{(k+1)} - x^{(k)} \| \leq \varepsilon$ ③



$$\min \operatorname{eig} \nabla^2 f(x) \geq m > 0$$
$$\forall x \in \text{Ball}$$

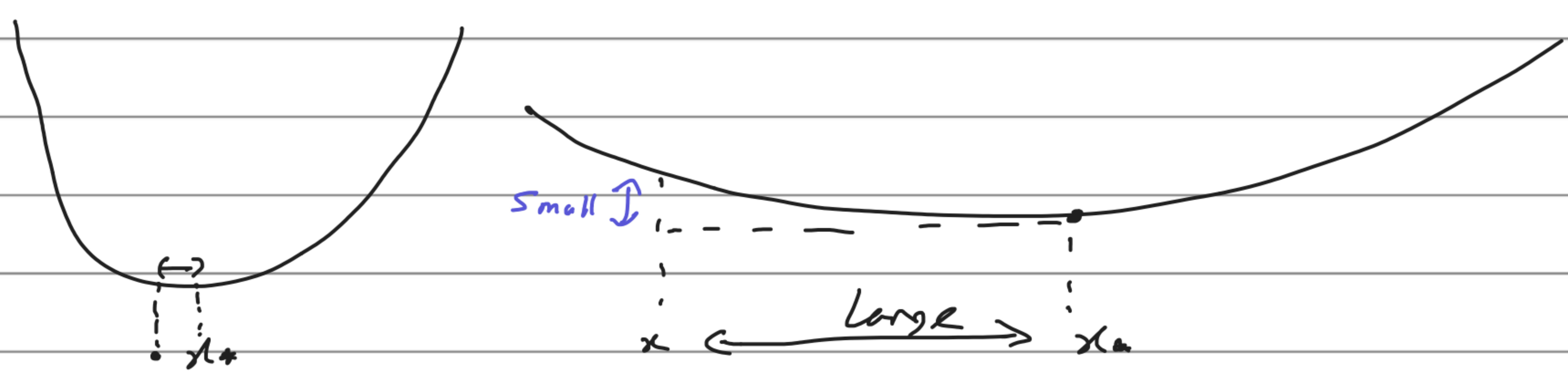
(SOC sufficient : $\nabla^2 f(x_*) \succ 0$)

Thm: If $\|\nabla f(x)\| \leq \varepsilon$ for any $x \in$ Ball

1 - $\|x - x_*\| \leq \dfrac{\varepsilon}{m}$?

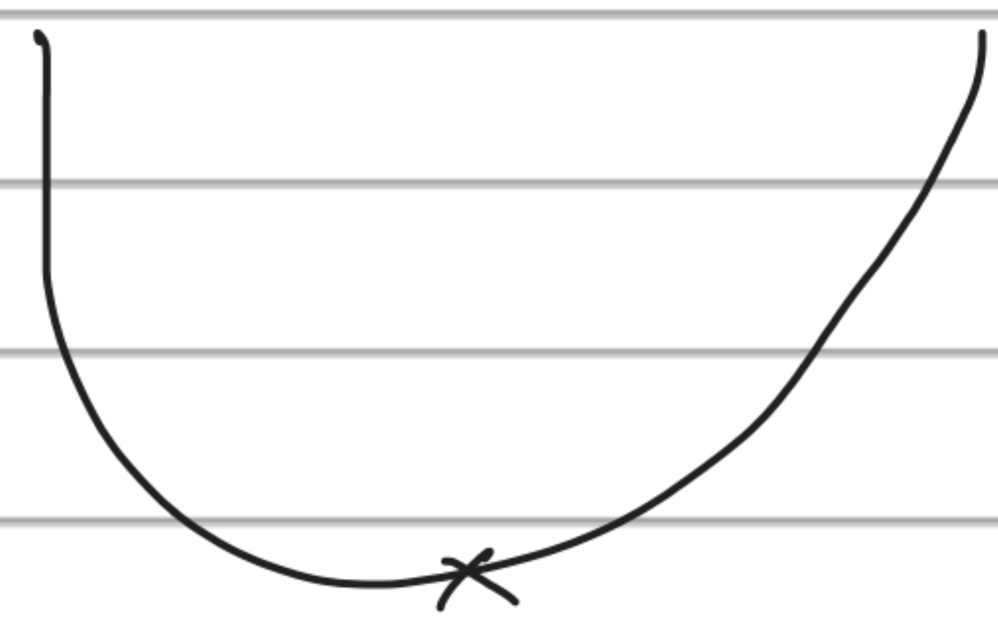
2 - $f(x) - f(x_*) \leq \dfrac{\varepsilon^2}{2m}$ ✓

gradient being small $\Longrightarrow$ close enough to solution (x_* , $f(x_*)$)



SOC necessary ✓, SOC sufficient ✗

$f(x) = x^4$

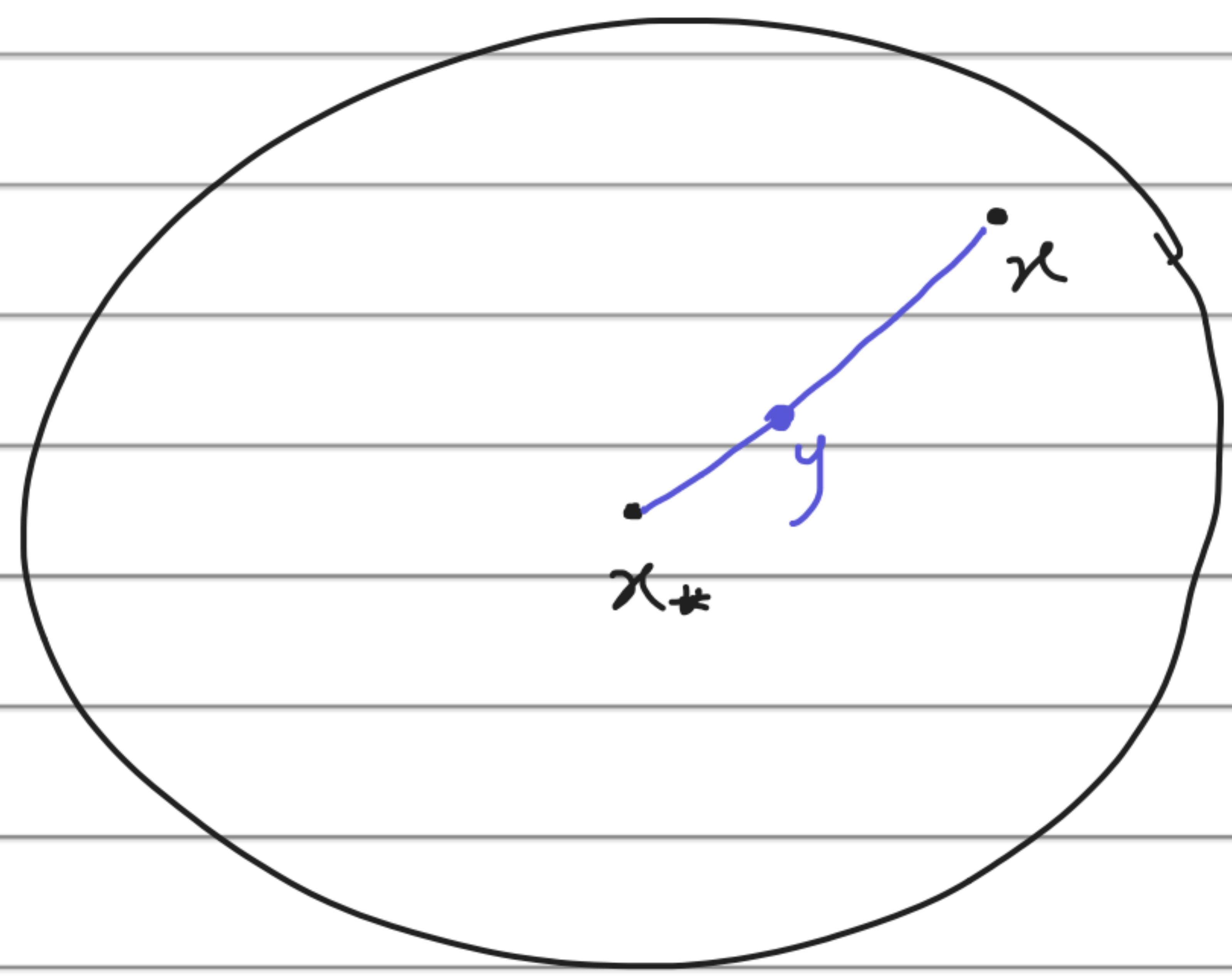


$\Longrightarrow f''(0) = 0$

≥ 0

$\Longrightarrow m = 0 \Longrightarrow \frac{\varepsilon}{m} = \infty$

proof:



$$\underbrace{f(x_*)}_{\text{new point}} = \underbrace{f(x)}_{\text{nominal}} + \nabla f(x)^T \overset{x_* - x}{\Delta x} + \frac{1}{2} \Delta x^T \nabla^2 f(\overset{y \in \text{Ball}}{y}) \Delta x$$

$$\geq f(x) + \nabla f(x)^T \Delta x + \frac{1}{2} m \|\Delta x\|^2$$

$\nabla^2 f(y) \geq m I$

$$\geq \min_{z \in \mathbb{R}^n} f(x) + \nabla f(x)^T z + \frac{1}{2} m \|z\|^2$$

$$\min_{z} \underbrace{f(x) + \nabla f(x)^T z + \frac{1}{2} m \|z\|^2}_{\text{quadratic, convex}}$$

↓

gradient w.r.t. z $= 0$

↓

$$0 + \nabla f(x) + \frac{1}{2} \times 2 \times m \times z_* = 0$$

$$\Rightarrow z_* = -\frac{\nabla f(x)}{m}$$ → plug into the quadratic term

$\Rightarrow \boxed{f(x_*)} \geq$ - - - - - - - (depends on z_*)

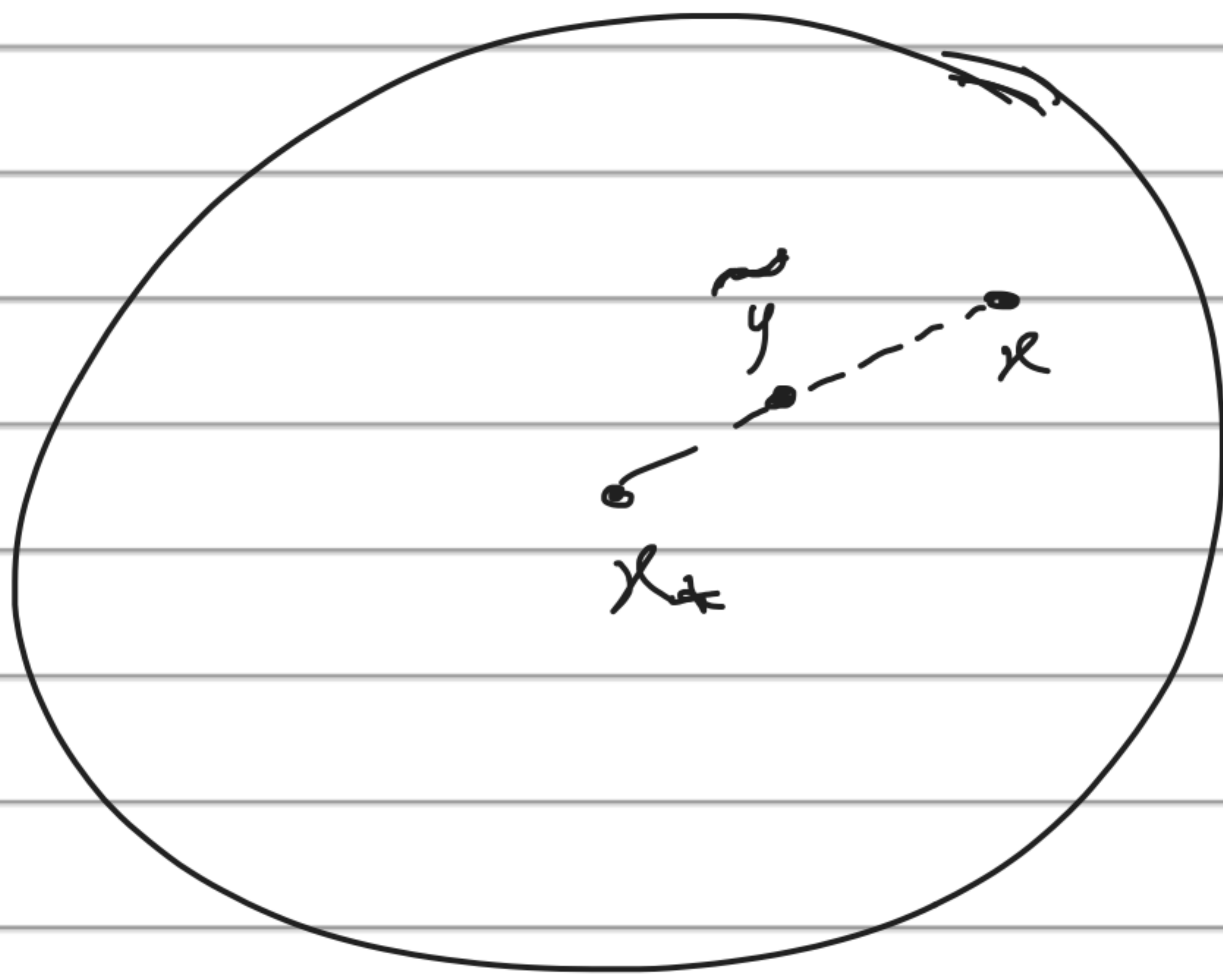
$$= f(x) - \frac{\|\nabla f(x)\|^2}{2m}$$

$$\geq \boxed{f(x) - \frac{\varepsilon^2}{2m}}$$

$$\Rightarrow 0 \overset{?}{\leq} f(x) - f(x_*) \leq \varepsilon^2/2m$$

Proof of part $\underline{\underline{1}}$:

$$f(x) = f(x_*) + \cancelto{0}{\nabla f(x_*)^T} \Delta x + \frac{1}{2} \Delta x^T \overbrace{\nabla^2 f(\tilde{y})}^{\succeq mI} \Delta x$$



$$\Longrightarrow \cancel{f(x)} \geq \cancel{f(x_*)} + \frac{1}{2} \times m \times \|\Delta x\|^2$$

$+$

$$\text{part } \underline{\underline{1}}: \quad \cancel{f(x_*)} \geq \cancel{f(x)} - \frac{\varepsilon^2}{2m}$$

$$\frac{1}{2} \frac{\varepsilon^2}{m} \geq \frac{1}{2} m \|\Delta x\|^2$$

$$\Longrightarrow \|\underbrace{\Delta x}_{x - x_*}\| \leq \frac{\varepsilon}{m}$$