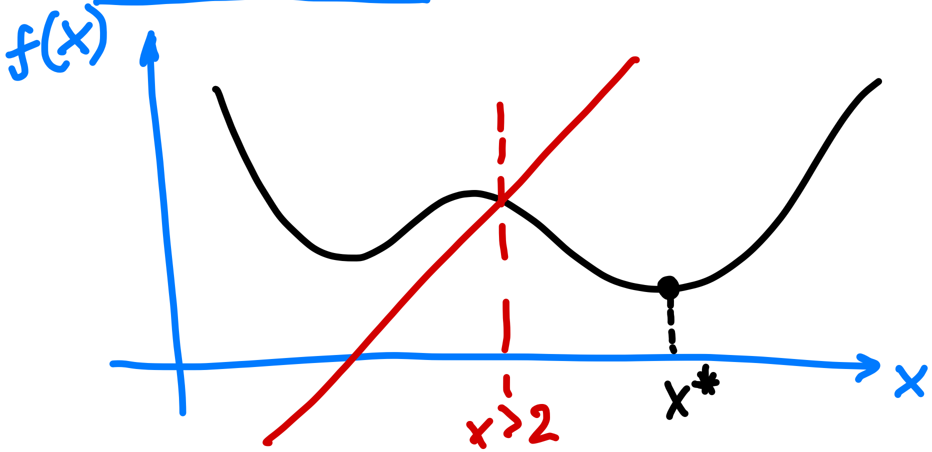
Optimization (Introduction)

Optimization

1D $f(x): \mathbb{R} \rightarrow \mathbb{R}$

ND $f(\underset{\sim}{x}): \mathbb{R}^n \rightarrow \mathbb{R}$

Goal: Find the **minimizer** $\boldsymbol{x}^*$ that minimizes the **objective (cost) function** $f(\boldsymbol{x}): \mathcal{R}^n \rightarrow \mathcal{R}$



Unconstrained Optimization

$$f(x^*) = \min_{x} f(x)$$

or

$$x^* = \arg \min_{x} f(x)$$

Optimization

Goal: Find the **minimizer** $\boldsymbol{x}^*$ that minimizes the **objective (cost) function** $f(\boldsymbol{x}): \mathcal{R}^n \rightarrow \mathcal{R}$

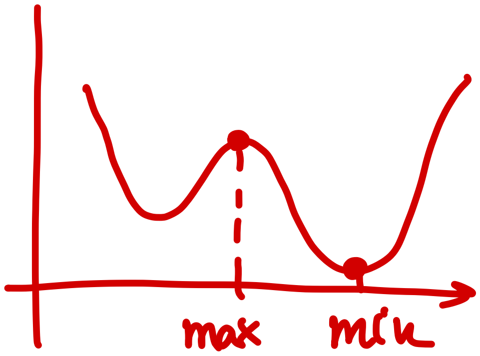
Constrained Optimization

$$\begin{cases} f(x^*) = \min_{x} f(x) \\ \text{s.t. } h_i(x) = 0 \rightarrow \text{equality} \\ \qquad g_j(x) \leq 0 \rightarrow \text{inequality} \end{cases}$$

$i = 1, n$

$j = 1, m$

Unconstrained Optimization



- **What if we are looking for a maximizer $\boldsymbol{x}^*$?**

$$f(\boldsymbol{x}^*) = \max_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$f(x^*) = \min_{x} (-f(x))$$

Calculus problem: maximize the rectangle area subject to perimeter constraint

$$\max_{\boldsymbol{d}\in\mathcal{R}^2} \quad f(d_1, d_2) = d_1 \times d_2$$

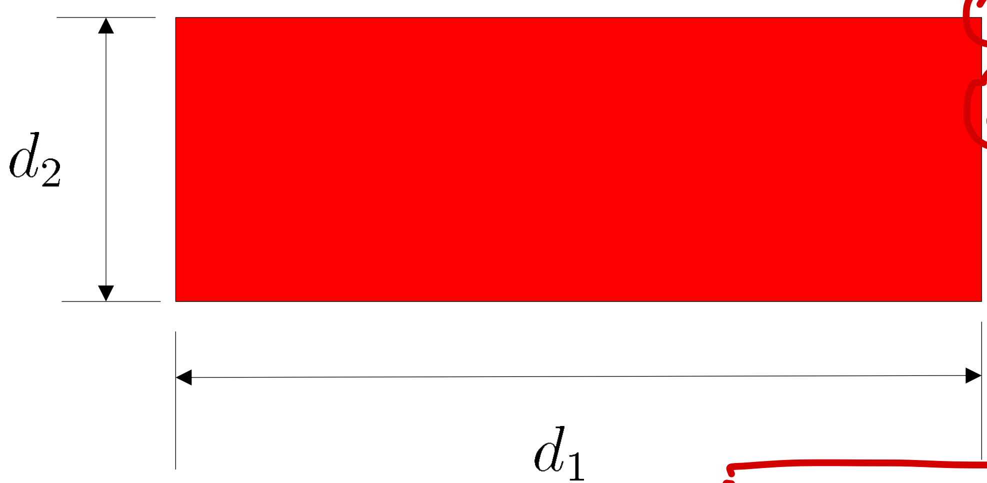
area → max Area

such that ① $g(d_1, d_2) = 2\,(d_1 + d_2) - 20 \le 0$ → perimeter constraint

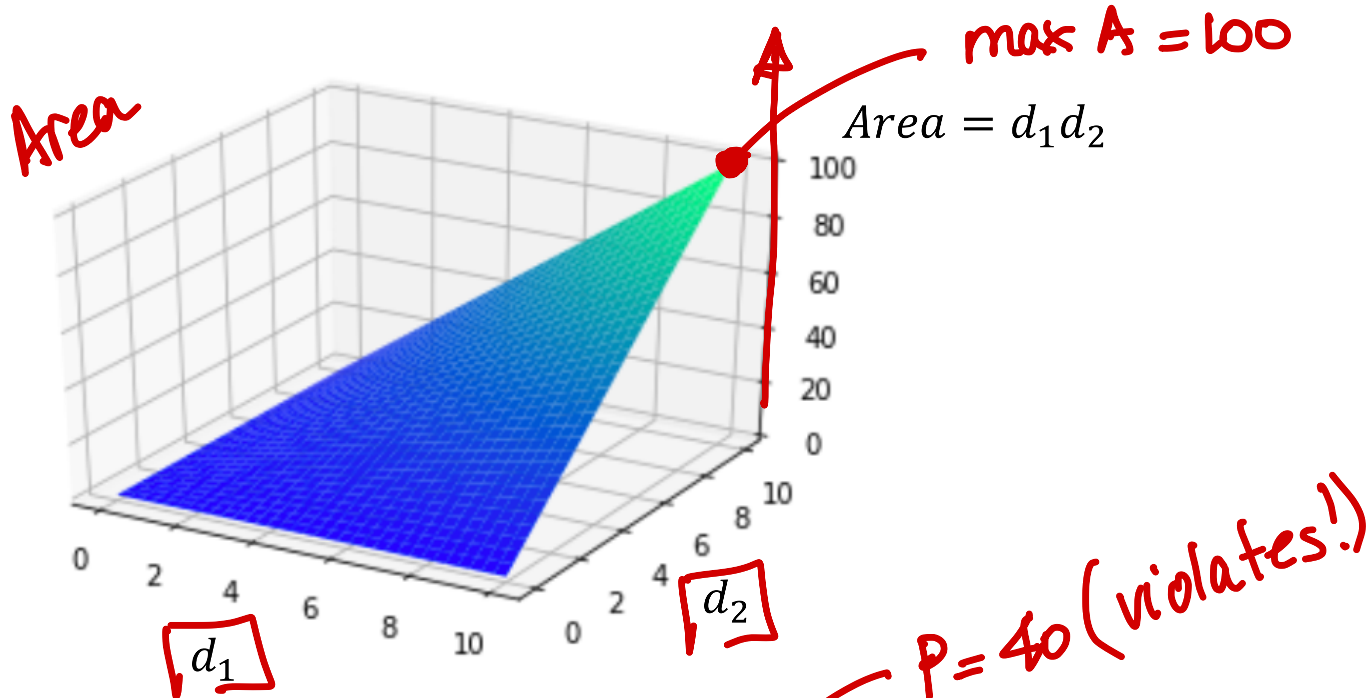
($2\,(d_1 + d_2)$: perimeter)

② $d_1 \le 10$

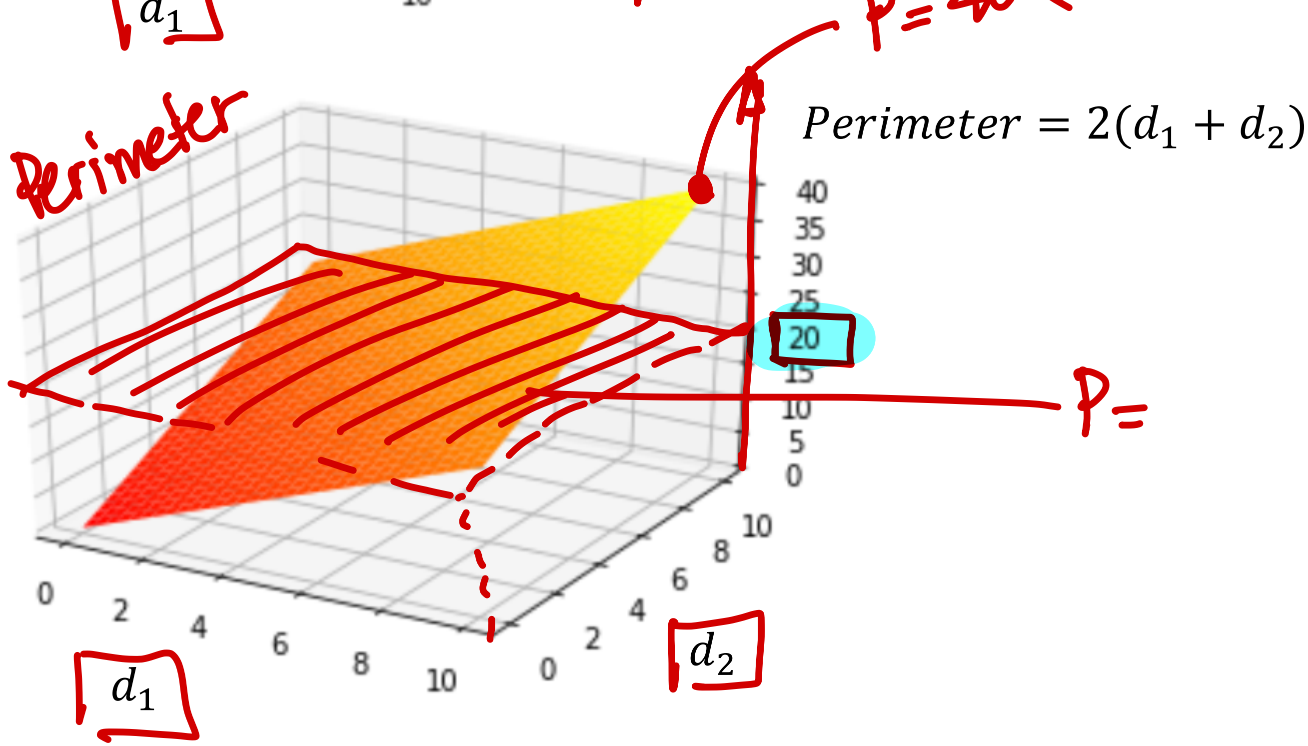
③ $d_2 \le 10$



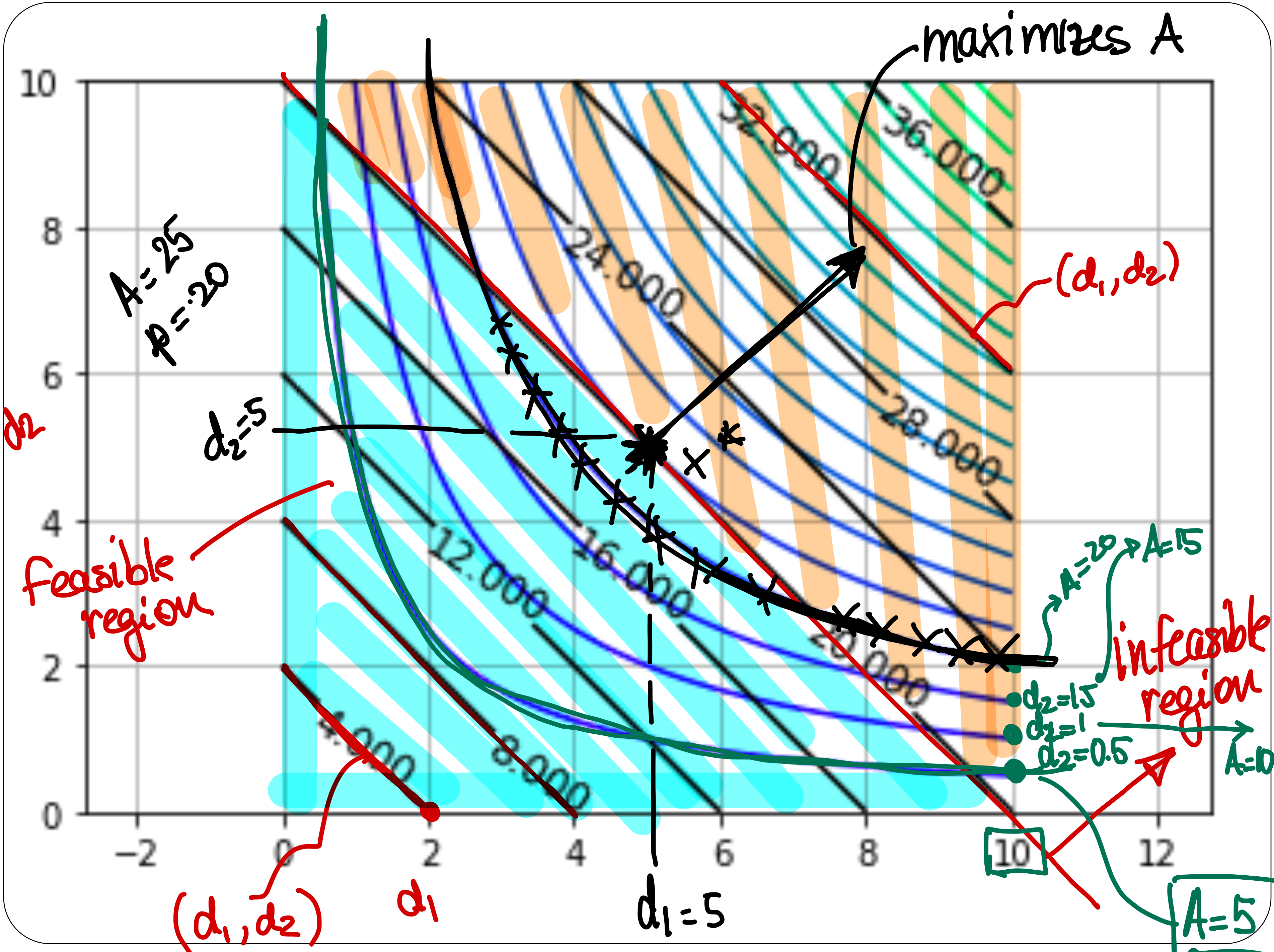
d_1^*, d_2^* (without peri const) $\Longrightarrow$ $d_1 = d_2 = 10 \rightarrow A = 100$



Area
max A = 100
$Area = d_1 d_2$
100
80
60
40
20
0
0
2
4
6
8
10
d_1
0
2
4
6
8
10
d_2



Perimeter
P= 40 (violates!)
$Perimeter = 2(d_1 + d_2)$
40
35
30
25
20
15
10
5
0
P=
0
2
4
6
8
10
d_1
0
2
4
6
8
10
d_2


maximizes A
A=25
p=.20
d2
d2=5
d1=5
d1
feasible region
infeasible region
(d1, d2)
(d1, d2)
A=20
A=15
A=10
A=5
d2=1.5
d2=1
d2=0.5
4.000
8.000
12.000
16.000
20.000
24.000
28.000
32.000
36.000

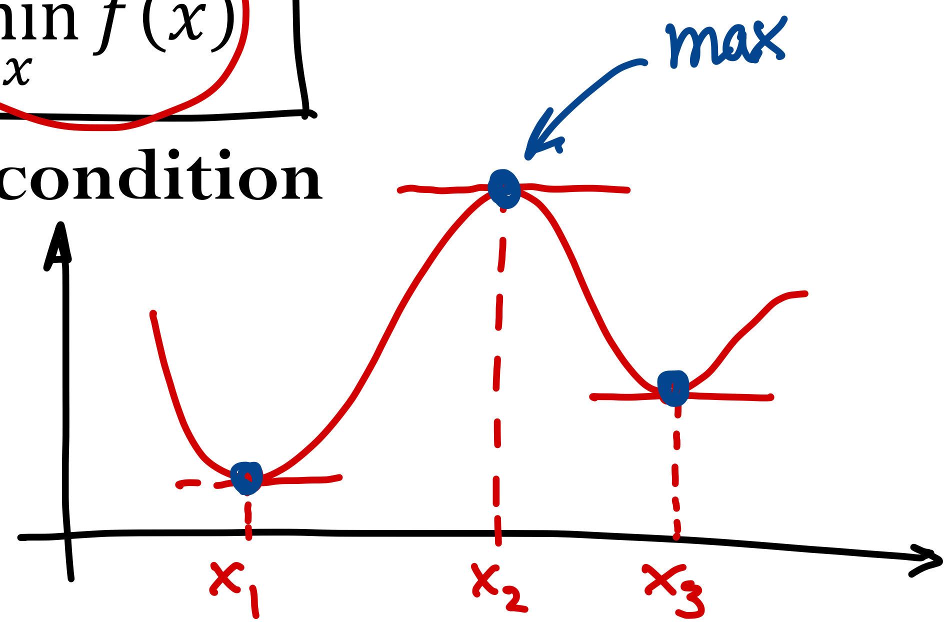
What is the optimal solution? (1D)

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

(First-order) Necessary condition

$$f'(x^*) = 0$$

gives stationary points



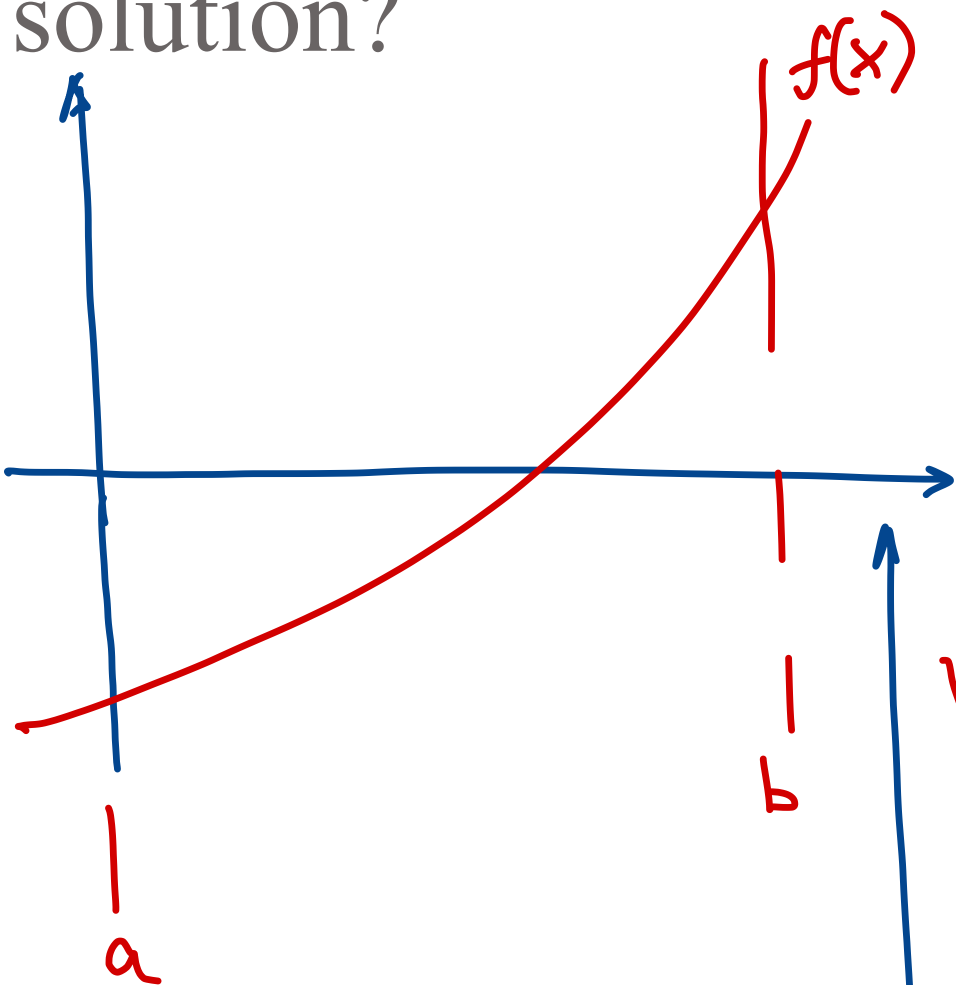
(Second-order) Sufficient condition

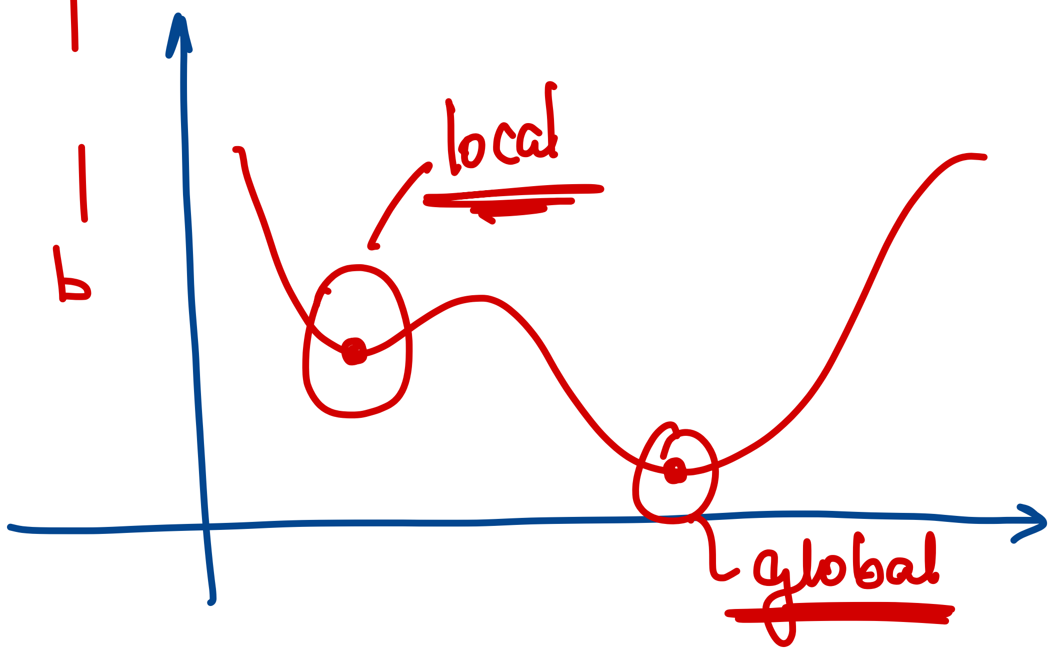
$f''(x^*) > 0 \longrightarrow x^*$ is minimum

$f''(x^*) < 0 \longrightarrow x^*$ is maximum

Does the solution exists? Local or global solution?

no stationary

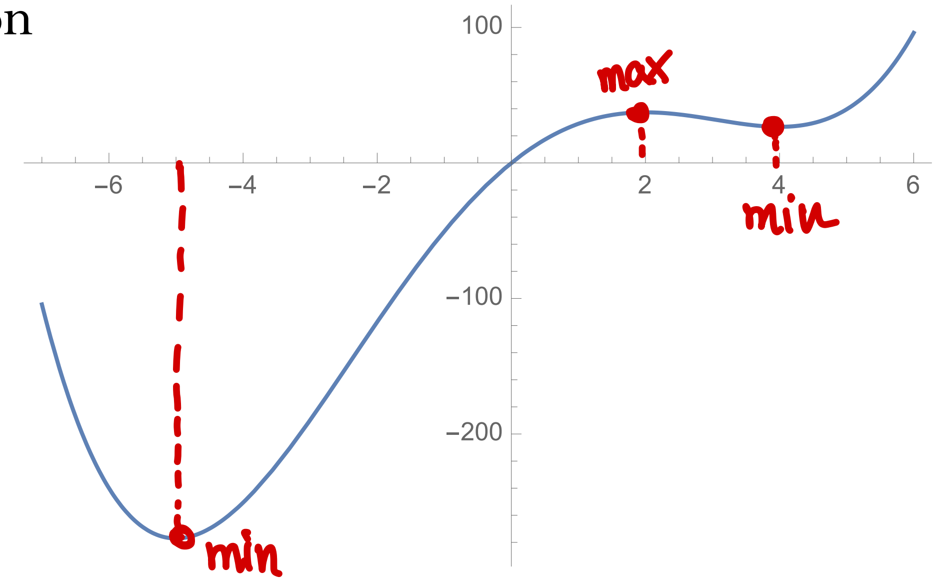




Example (1D)

min f(x)
x

Consider the function $f(x) = \frac{x^4}{4} - \frac{x^3}{3} - 11x^2 + 40x$. Find the stationary point and check the sufficient condition



* 1st order necessary condition

$f'(x) = \frac{4x^3}{4} - \frac{3x^2}{3} - 22x + 40$

$f'(x) = 0 \Rightarrow x^3 - x^2 - 22x + 40 = 0$

solutions $\Rightarrow x = \begin{cases} -5 \\ 2 \\ 4 \end{cases}$

* 2nd order condition:

$f''(x) = 3x^2 - 2x - 22$

$f''(-5) = 3(25) + 10 - 22 > 0$ (MIN)

$f''(2) = 12 - 4 - 22 < 0 \longrightarrow$ (MAX)

$f''(4) = 3(16) - 8 - 22 > 0 \rightarrow$ (MIN)

Types of optimization problems

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

f: nonlinear, continuous and smooth

Gradient-free methods

Evaluate $f(x)$

Gradient (first-derivative) methods

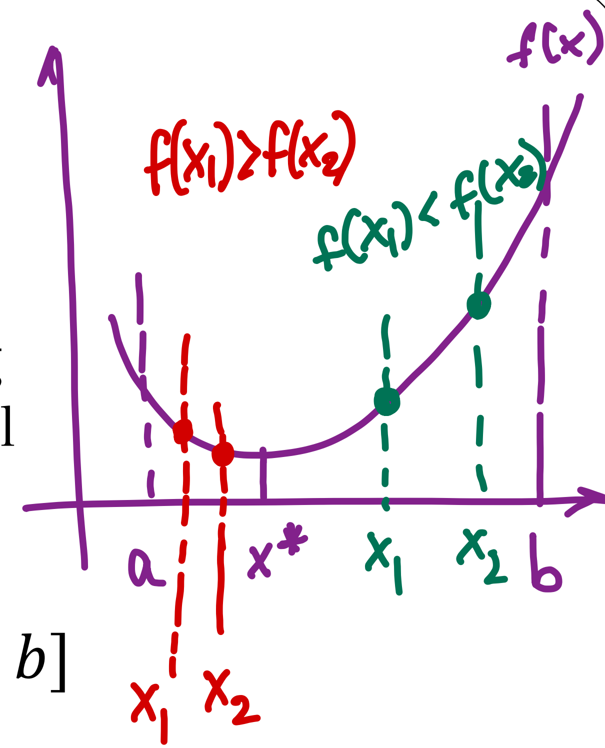
Evaluate $f(x), f'(x)$

Second-derivative methods

Evaluate $f(x), f'(x), f''(x)$

Optimization in 1D: Golden Section Search

- Similar idea of bisection method for root finding
- Needs to bracket the minimum inside an interval
- Required the function to be unimodal

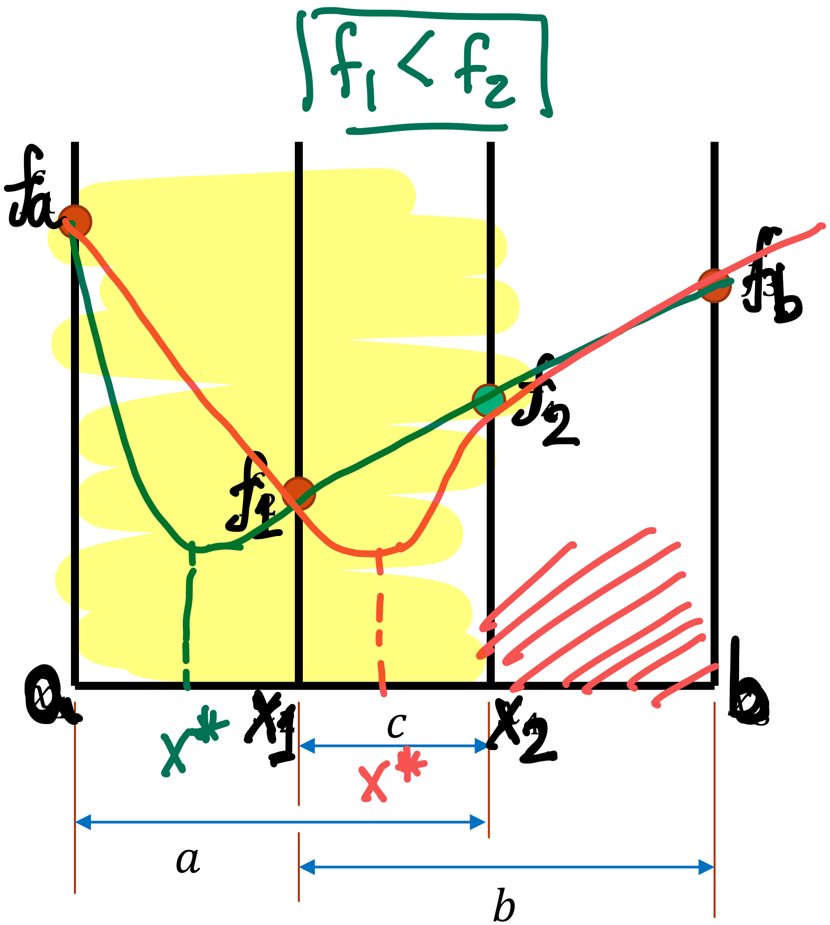


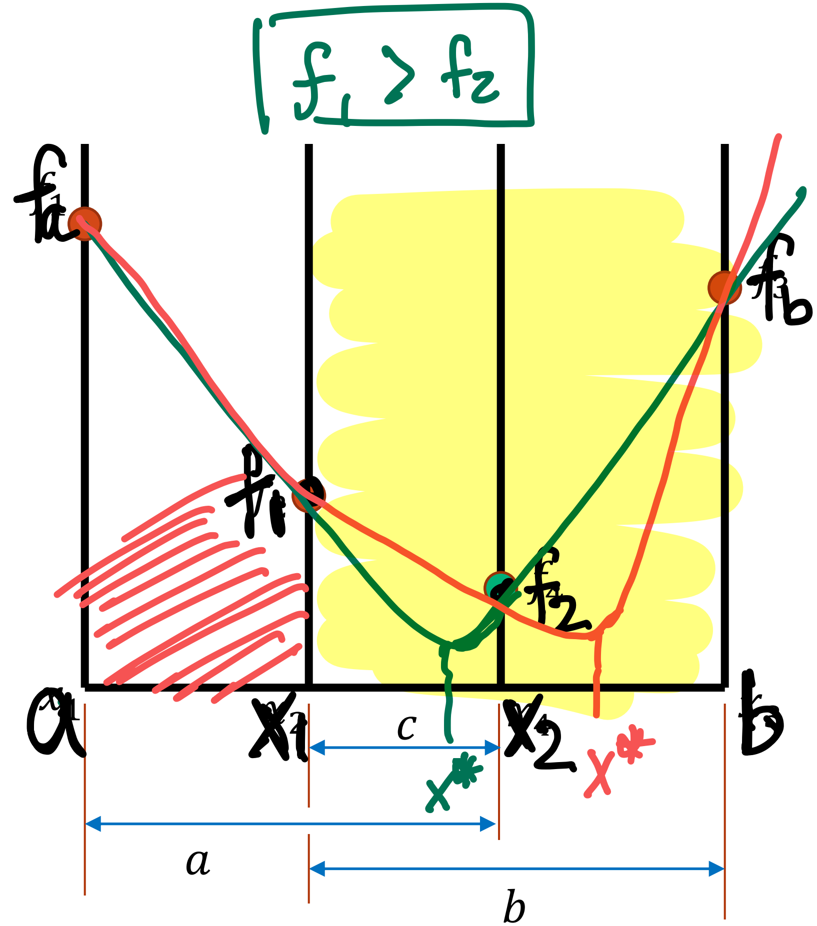
A function $f: \mathcal{R} \rightarrow \mathcal{R}$ is unimodal on an interval $[a, b]$

✓ There is a unique $\boldsymbol{x}^* \in [a, b]$ such that $f(\boldsymbol{x}^*)$ is the minimum in $[a, b]$

✓ For any $x_1, x_2 \in [a, b]$ with $x_1 < x_2$

- $x_2 < \boldsymbol{x}^* \Rightarrow f(x_1) > f(x_2)$
- $x_1 > \boldsymbol{x}^* \Rightarrow f(x_1) < f(x_2)$





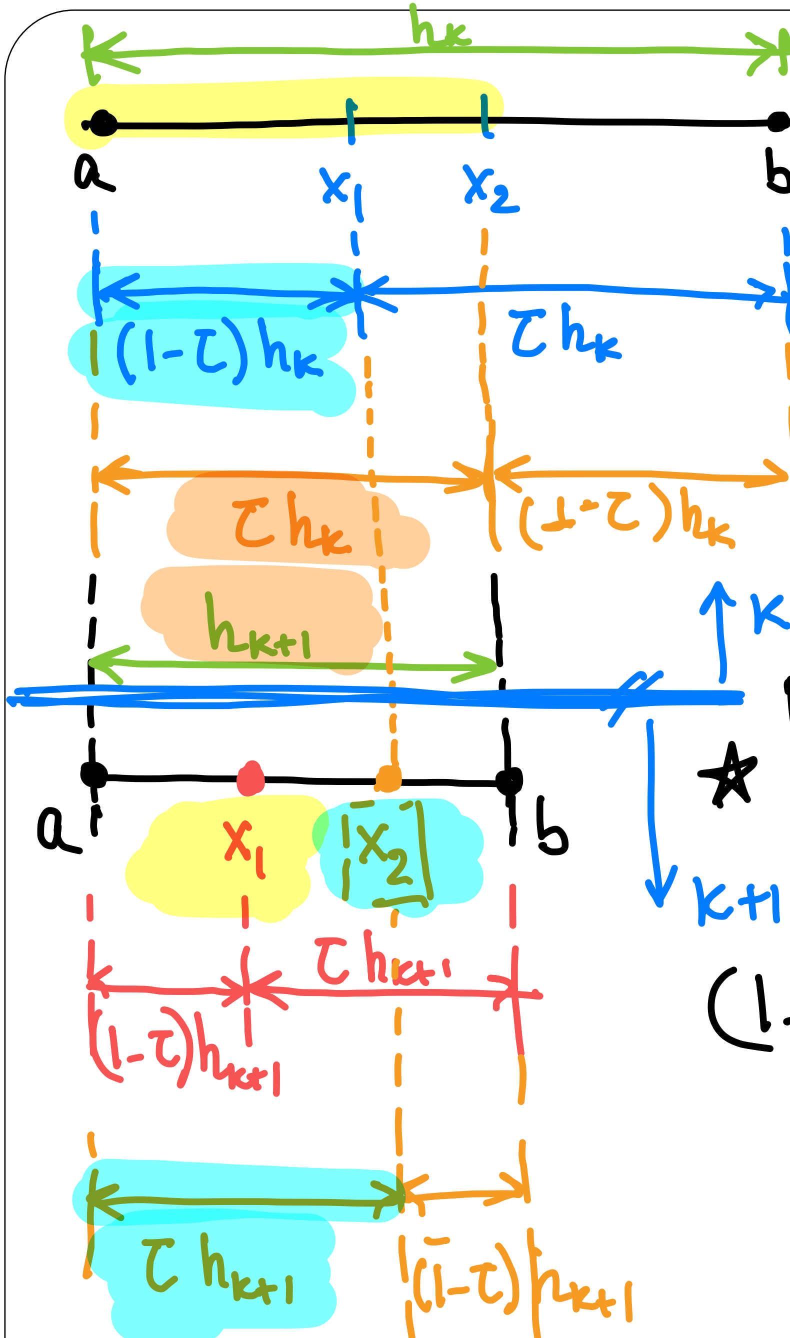
$f_1 < f_2$ $\quad$ $x_1 < x_2$

$f_1 > f_2$ $\quad$ $x_1 < x_2$

$x^* \in [a, x_2]$

$x^* \in [x_1, b]$

$x_1, x_2 = ?$



Propose the point x_1, x_2 s.t.

$$x_1 = a + (1-\tau) h_k$$

$$x_2 = a + \tau h_k$$

at the start $h_k = (b-a)$

$f_1 > f_2$ or $f_1 < f_2$

$[x_1, b]$ $[a, x_2]$

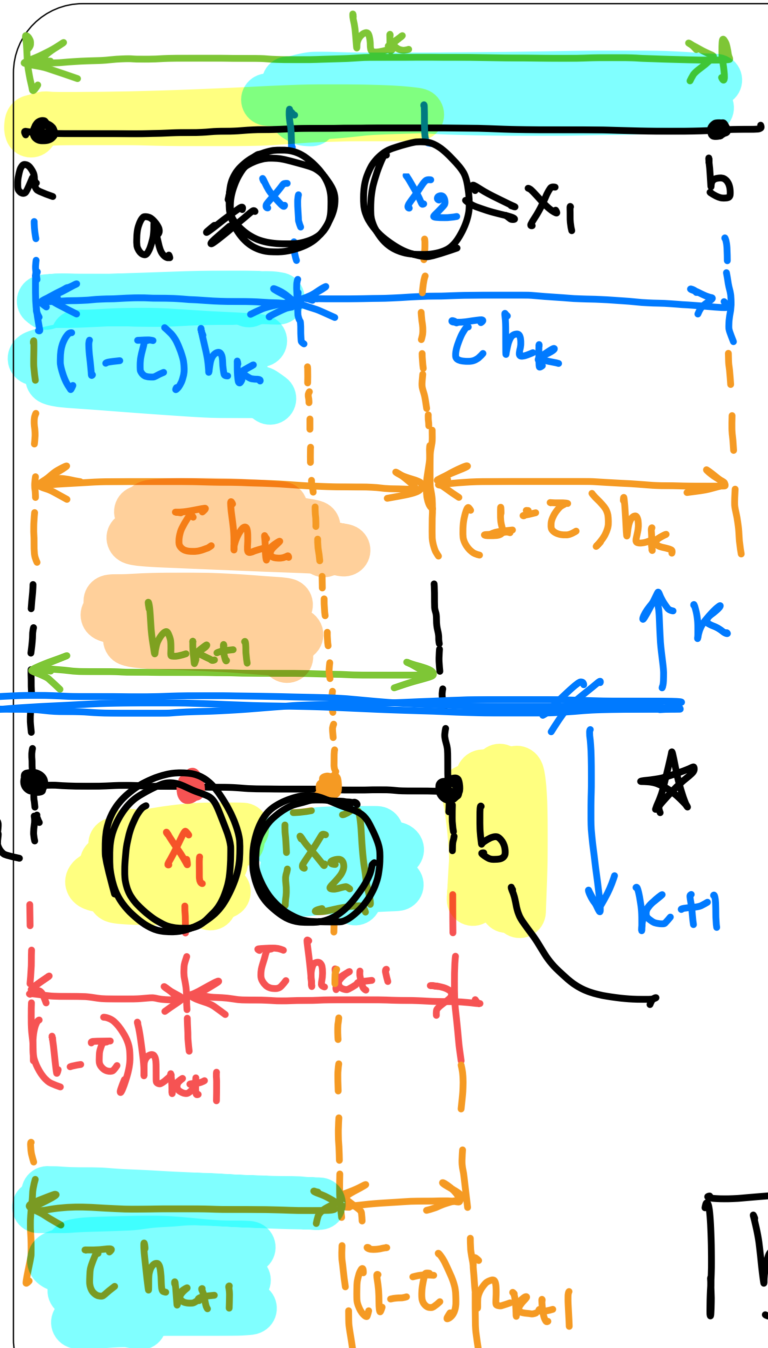
★ $$h_{k+1} = \tau h_k$$

Every iteration interval gets reduced by τ

$$(1-\tau)h_k = \tau h_{k+1} = \tau(\tau h_k)$$

$$(1-\tau)h_k = \tau^2 h_k$$

$$(1-\tau) = \tau^2 \rightarrow \tau = 0.618$$



interval (a,b) $\tau = 0.618$

$h_0 = (b-a)$

$\rightarrow$ $x_1 = a + (1-\tau) h_0$

$x_2 = a + \tau h_0$

$f_1 = f(x_1)$ $f_2 = f(x_2)$

if $f_1 < f_2$: $\longrightarrow x^* \in [a, x_2]$

$b = x_2$

$x_2 = x_1 \longrightarrow f_2 = f_1$

$h_{k+1} = \tau h_k$

$x_1 = a + (1-\tau) h_{k+1}$

$f_1 = f(x_1)$

if $f_1 > f_2$: $\longrightarrow x^* \in [x_1, b]$

$a = x_1$

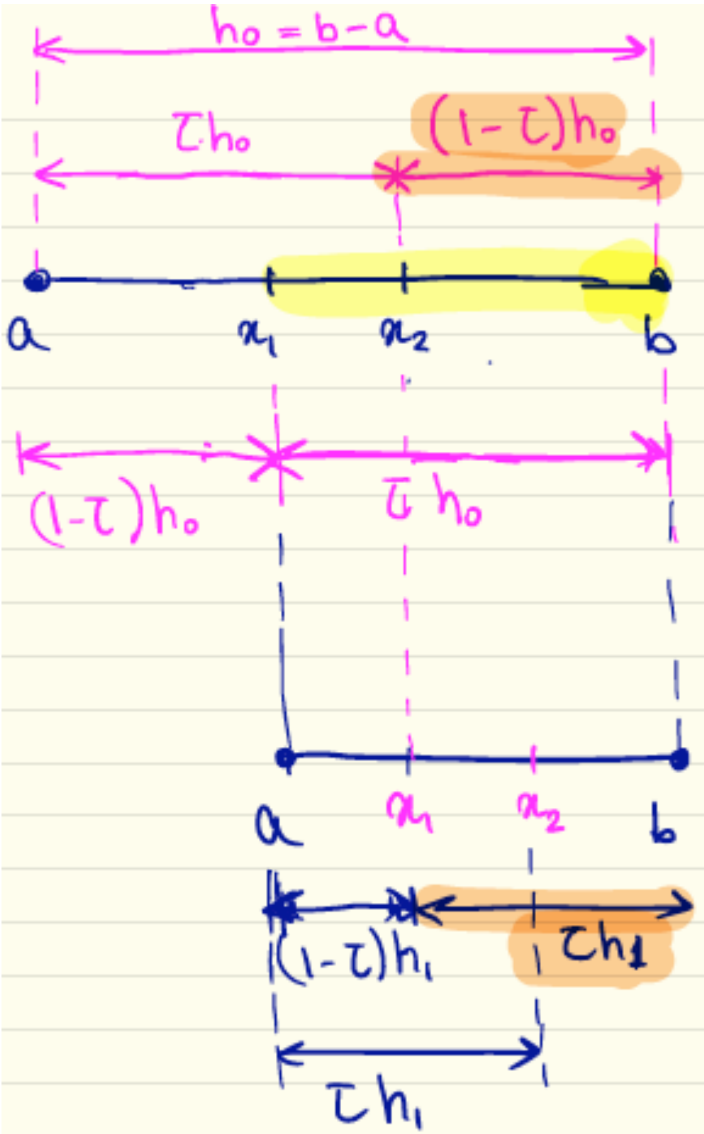
$x_1 = x_2 \longrightarrow f_1 = f_2$

$h_{k+1} = \tau h_k$

$x_2 = a + \tau h_k$ $f_2 = f(x_2)$

$h_{k+1} < tol$

Golden Section Search



Propose:

$x_1 = a + (1-\tau)h_0$

$x_2 = a + \tau h_0$

Evaluate $f_1 = f(x_1)$

$f_2 = f(x_2)$

if $(f_1 > f_2)$:

$a = x_1$

$x_1 = x_2$ → already have func. value!

$h_1 = b - a$

$x_2 = a + \tau h_1$

$f_2 = f(x_2)$ → only one

if $(f_1 < f_2)$:

$b = x_2$

$x_2 = x_1$

$x_1 = a + (1-\tau)h_1$

$f_1 = f(x_1)$

Golden Section Search

What happens with the length of the interval after one iteration?

$$h_1 = \tau\, h_o$$

Or in general: $h_{k+1} = \tau\, h_k$

Hence the interval gets reduced by $\boldsymbol{\tau}$
(for bisection method to solve nonlinear equations, τ=0.5)

For recursion:

$$\tau\, h_1 = (1 - \tau)\, h_o$$
$$\tau\, \tau\, h_o \;= (1 - \tau)\, h_o$$
$$\tau^2 = (1 - \tau)$$
$$\boldsymbol{\tau = 0.618}$$

Golden Section Search

x^* → $h_k < tol$

$x^* \in h_k$

- Derivative free method! → evaluate $f(x)$

- Slow convergence:

$e_k = h_k$ $\qquad \frac{e_{k+1}}{e_k^r} = \frac{h_{k+1}}{h_k^r} = \frac{\tau h_k}{h_k^r}$

$r = 1 \rightarrow \tau$

$$\lim_{k\to\infty} \frac{|e_{k+1}|}{|e_k|} = 0.618 \quad r = 1 \; (linear\ convergence)$$

- Only one function evaluation per iteration

x_1, x_2 cheap,

Example

Consider running golden section search on a function that is unimodal. If golden section search is started with an initial brakcet of $[-10, 10]$, what is the length of the new bracket after 1 iteration?

A) 20
B) 10
C) 12.36
D) 7.64

$a = -10$ $\Longrightarrow$ $h_0 = 20$

$b = 10$

$h_1 = ?$

$h_1 = \tau h_0 \Longrightarrow 0.618 \times 20 = 12.36$

Newton's Method

$x_{k+1} = x_k + h$

Using Taylor Expansion, we can approximate the function f with a quadratic function about x_0

quadratic approximation

$$\underbrace{f(x)}_{\text{nonlinear}} \approx f(x_0) + f'(x_0)\,(x - x_0) + \frac{1}{2} f''(x_0)\,(x - x_0)^2 = \hat{f}$$

And we want to find the minimum of the quadratic function using the first-order necessary condition

$$f'(x) = 0 \longrightarrow \hat{f}' = 0$$

$$f'(x_0) + \frac{1}{2} f''(x_0)(x - x_0)\,2 = 0$$

$$f'(x_0) + f''(x_0)(x - x_0) = 0$$

$$x - x_0 = -\frac{f'(x_0)}{f''(x_0)} \Longrightarrow \boxed{x = x_0 - \frac{f'(x_0)}{f''(x_0)}}$$

stationary point

Newton step

Newton's Method

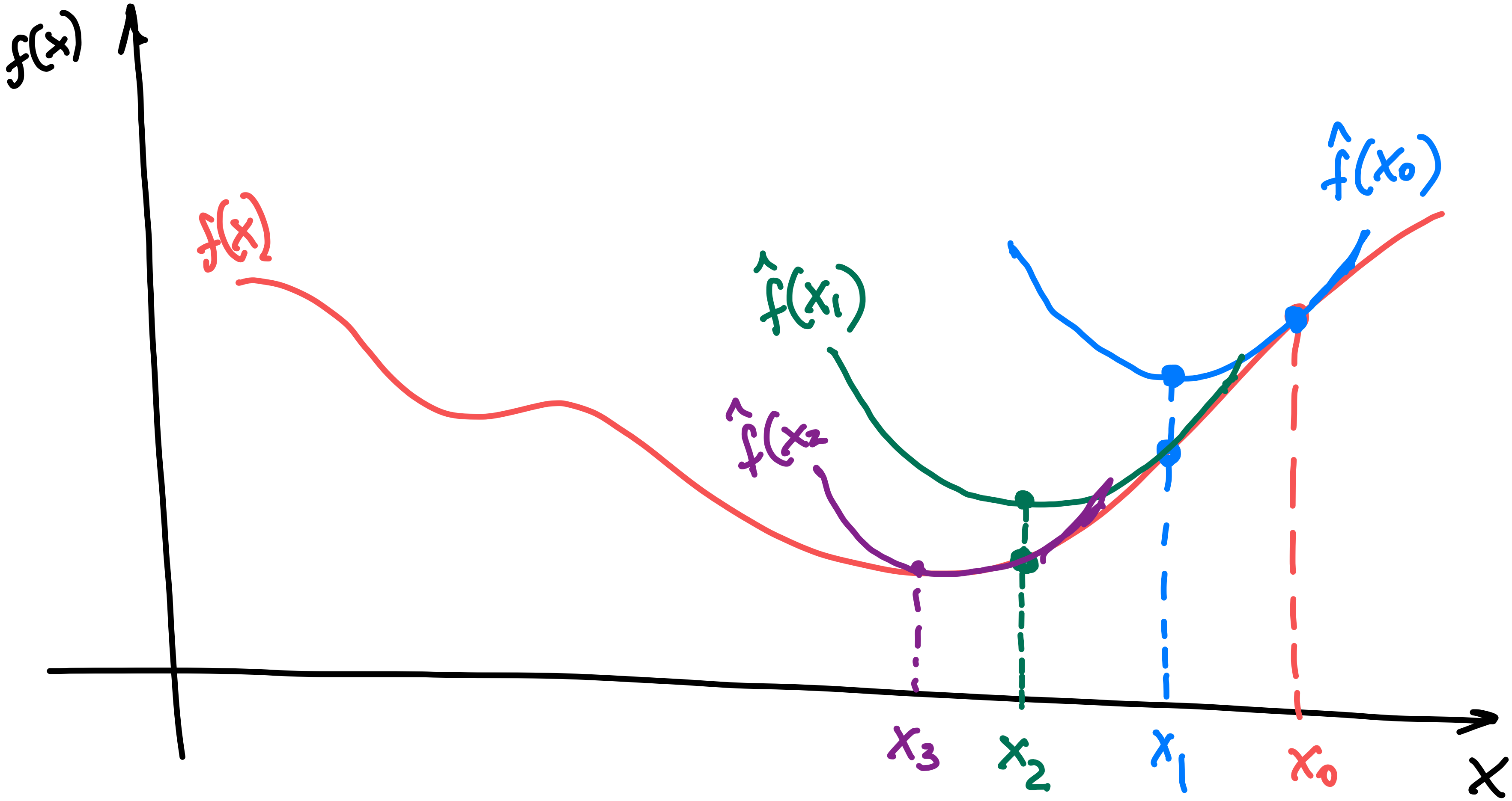
- **Algorithm:**

x_0 = starting guess

$x_{k+1} = x_k - f'(x_k)/f''(x_k)$

- **Convergence:**
  - Typical quadratic convergence
  - Local convergence (start guess close to solution)
  - May fail to converge, or converge to a maximum or point of inflection

Newton's Method (Graphical Representation)



sequence of opt.
using quad. approx $\hat{f}$

Example

Consider the function $f(x) = 4\,x^3 + 2\,x^2 + 5\,x + 40$

If we use the initial guess $x_0 = 2$, what would be the value of x after one iteration of the Newton's method?

$x_1 = ?$

$f'(x) = 12x^2 + 4x + 5$

$f''(x) = 24x + 4$

$$h = -\frac{f'(x)}{f''(x)} = -\frac{(12(4) + 4(2) + 5)}{24(2) + 4} = -\frac{61}{52}$$

$$x_1 = x_0 + h \implies x_1 = 2 - \frac{61}{52} \rightarrow \boxed{x_1 = 0.8269}$$