Optimization

Optimization

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

Unconstrained Optimization

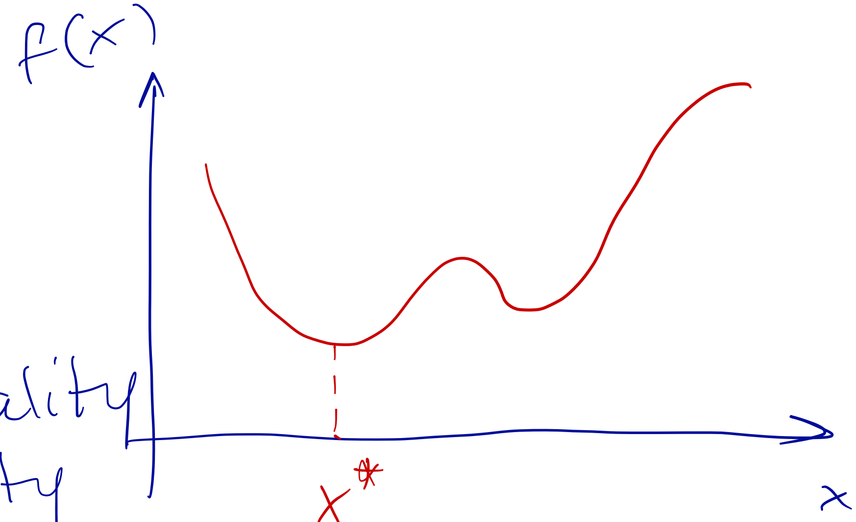
minimize $f(x)$

Find x^* such that

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

x^* : minimizer

$f(x)$: cost function



Constrained Optimization

$$\min_x f(x)$$

s.t $h(x) \geq 0 \rightarrow$ inequality

$g(x) = 0 \rightarrow$ equality

Optimization

$$f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}} f(\boldsymbol{x})$$
$$\text{s.t. } \mathbf{g}(\boldsymbol{x}) = \mathbf{0}$$
$$\mathbf{h}(\boldsymbol{x}) \le \mathbf{0}$$

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

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

$$\min_{x} \left(-f(x)\right) \Longrightarrow f = -f$$

- **What if constraint is $\mathbf{h}(\boldsymbol{x}) > \mathbf{0}$?**

$$-h \le 0 \Longrightarrow h \ge 0 \Longrightarrow h = -h$$

- **What if method only has inequality constraints?**

$$g = 0 \Rightarrow -\epsilon \le h \le \epsilon$$

for small ϵ

Calculus problem: maximize the rectangle area subject to perimeter constraint

without perimeter constraint, what would be the maximizer of the area?

area

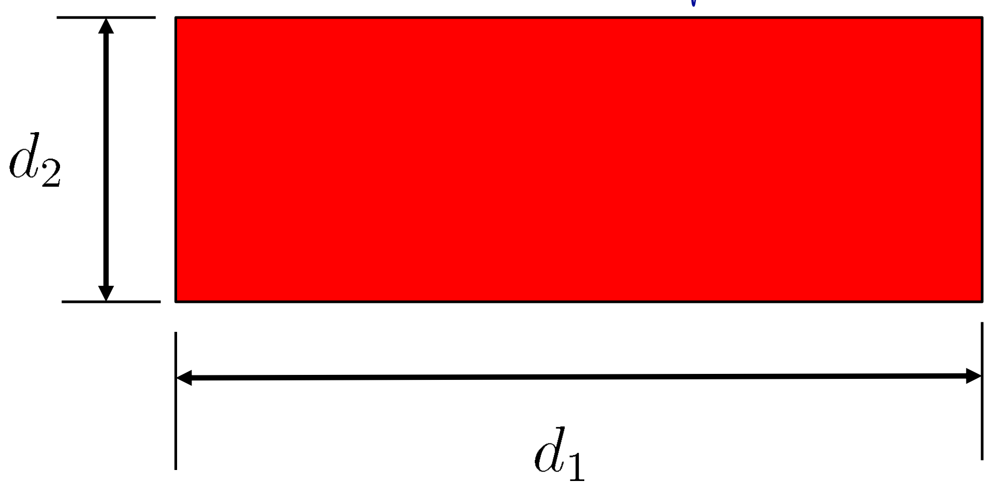
$$\max_{\boldsymbol{d} \in \mathcal{R}^2} \quad f(d_1, d_2) = d_1 \times d_2$$

$$\text{such that} \quad g(d_1, d_2) = 2\,(d_1 + d_2) - 20 \leq 0$$

perimeter

$d_1 \leq 10$

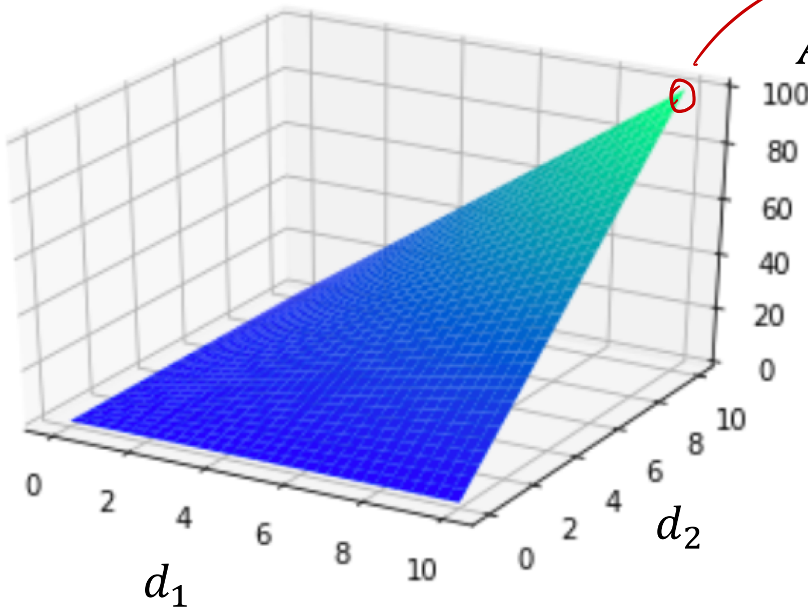
$d_2 \leq 10$



Demo: Constrained-Problem-2D

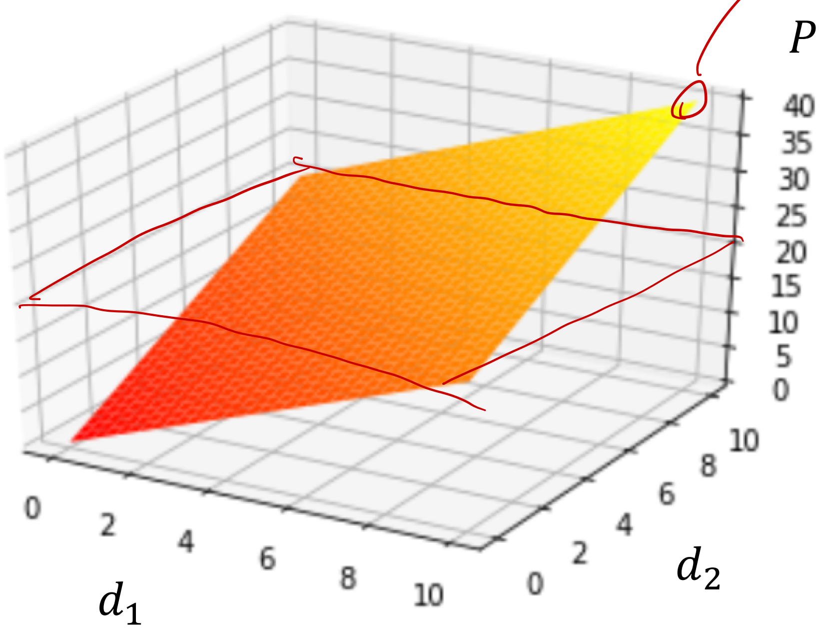
unconstrained solution

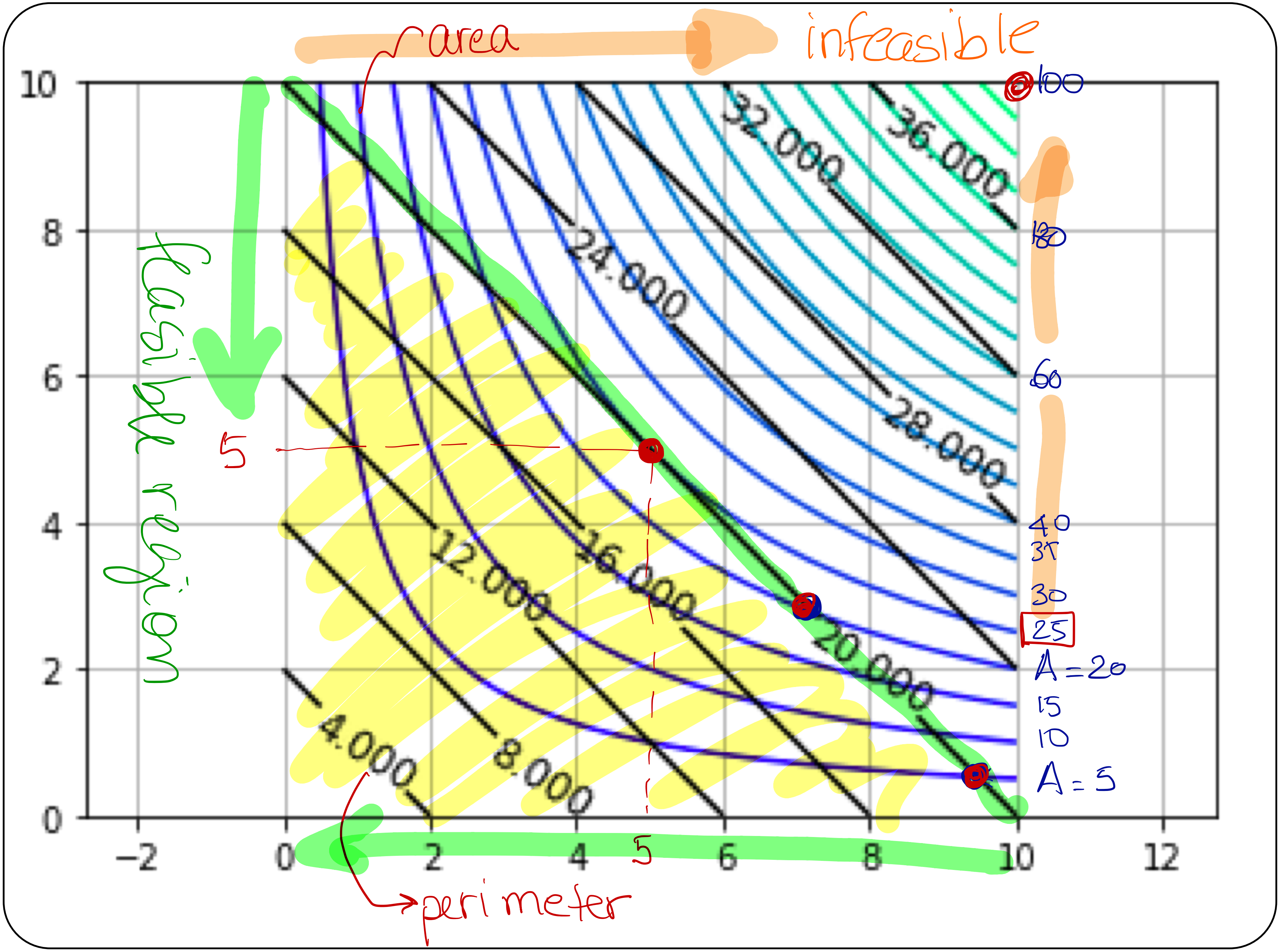
$Area = d_1 d_2$



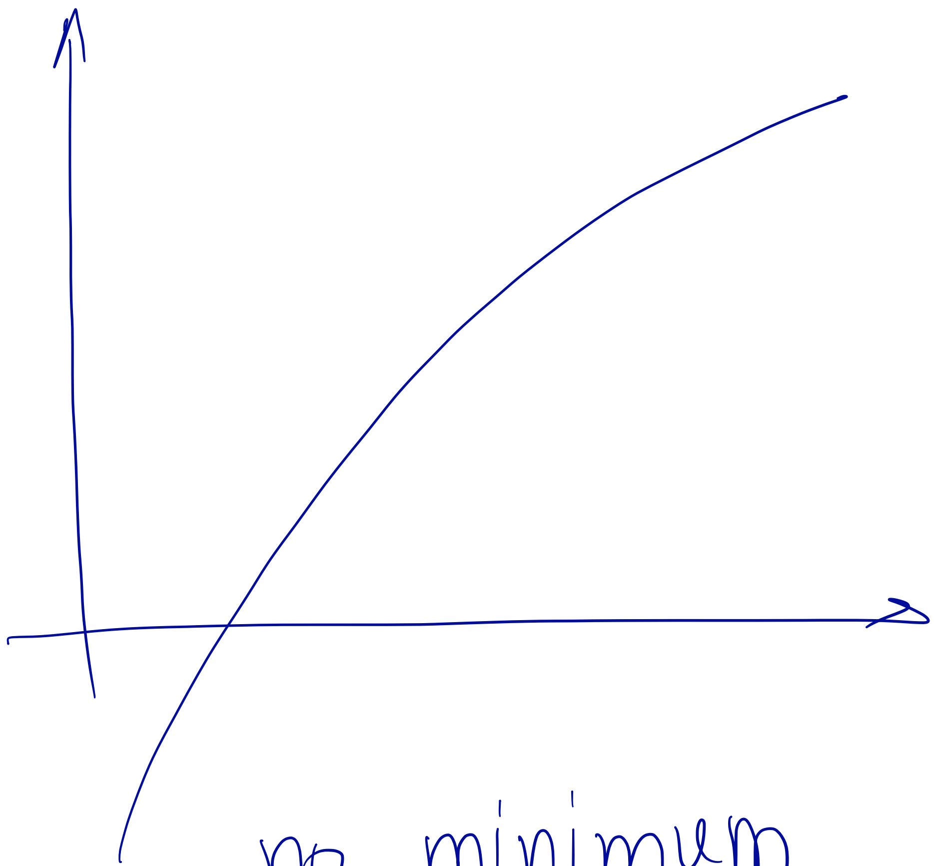
perimeter = 40 > 20 !

$Perimeter = 2(d_1 + d_2)$

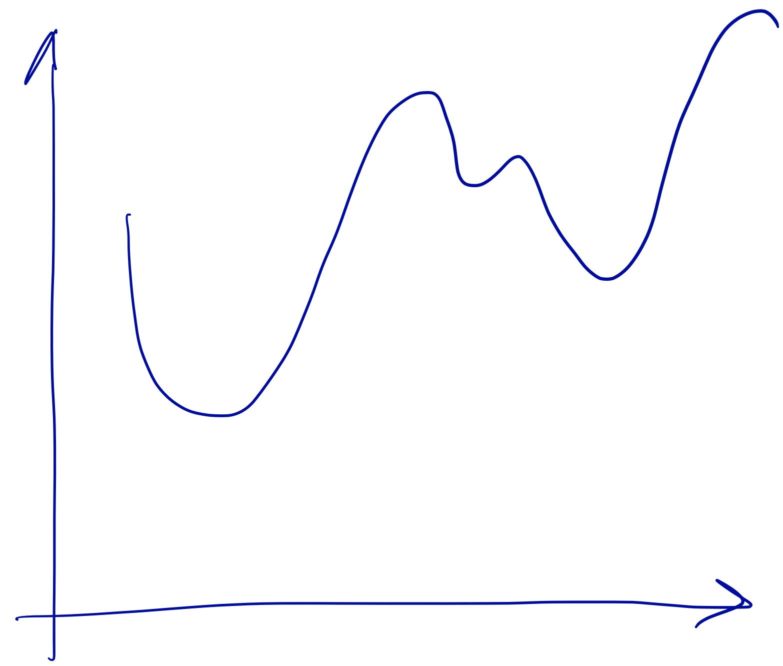



area
infeasible
feasible region
perimeter
5
5
100
80
60
40
35
30
25
A=20
15
10
A=5
4.000
8.000
12.000
16.000
20.000
24.000
28.000
32.000
36.000

Does the solution exists? Local or global solution?



no minimum



several minimizers

Types of optimization problems

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

f: nonlinear, continuous and smooth

AI : GA
SA

Stochastic methods

MC

Gradient-free methods

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

nD → $f(\underset{\sim}{x}) : \mathbb{R}^n \to \mathbb{R}$

Evaluate $f(\boldsymbol{x})$

Gradient (first-derivative) methods

1D : $f(x)$, $f'(x)$

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

$\nabla f(\underset{\sim}{x}) : \mathbb{R}^n \to \mathbb{R}^n$

Evaluate $f(\boldsymbol{x}), \boldsymbol{\nabla} f(\boldsymbol{x})$

Second-derivative methods

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

ND : $\nabla^2 f(\underset{\sim}{x}) : \mathbb{R}^n \to \mathbb{R}^{n\times n}$

Evaluate $f(\boldsymbol{x}), \boldsymbol{\nabla} f(\boldsymbol{x}), \boldsymbol{\nabla}^2 f(\boldsymbol{x})$

Taking derivatives…

$$f(\underset{\sim}{x}) = f\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \mathbb{R}^n \to \mathbb{R}$$

$$\nabla f(\underset{\sim}{x}) = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix} \qquad \mathbb{R}^n \to \mathbb{R}^n$$

$$\nabla^2 f(\underset{\sim}{x}) = H(\underset{\sim}{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \ddots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

$\underline{\underline{n \times n}}$ Hessian

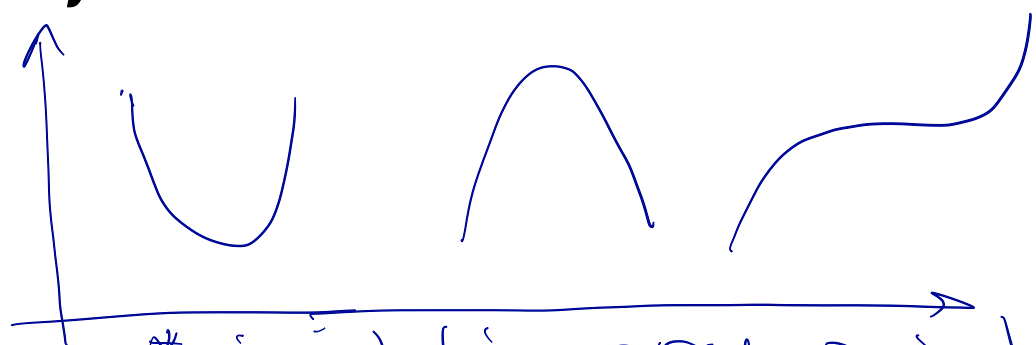
What is the optimal solution?

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

(First-order) Necessary condition

(1D) $f'(x^*) = 0$

(ND) $\nabla f(x^*) = \underline{0}$



x^* is stationary point!

(Second-order) Sufficient condition

$$\left.\begin{array}{l} \text{(1D) } f''(x^*) > 0 \\ \\ \text{(ND) } \nabla^2 f(x^*) \text{ is} \\ \quad \text{positive definite} \end{array}\right\} x^* \text{ is a minimizer}$$

$\min_{\underset{\sim}{x}} f(\underset{\sim}{x})$

First-order necessary condition

$\rightarrow \nabla f(\underset{\sim}{x}) = \underset{\sim}{0}$

Second-order sufficient condition

$\rightarrow H_f$ is positive definite

$$Hy = \lambda y \Rightarrow y^T H y = \lambda y^T y = \lambda \|y\|_2^2$$

★ $y^T H y > 0$ for all $y \neq 0 \Rightarrow \lambda_i > 0$ for all i

$\Rightarrow$ positive_definite $\Rightarrow$ minimizer

★ $y^T H y < 0$ for all $y \neq 0 \Rightarrow \lambda_i < 0$ for all i

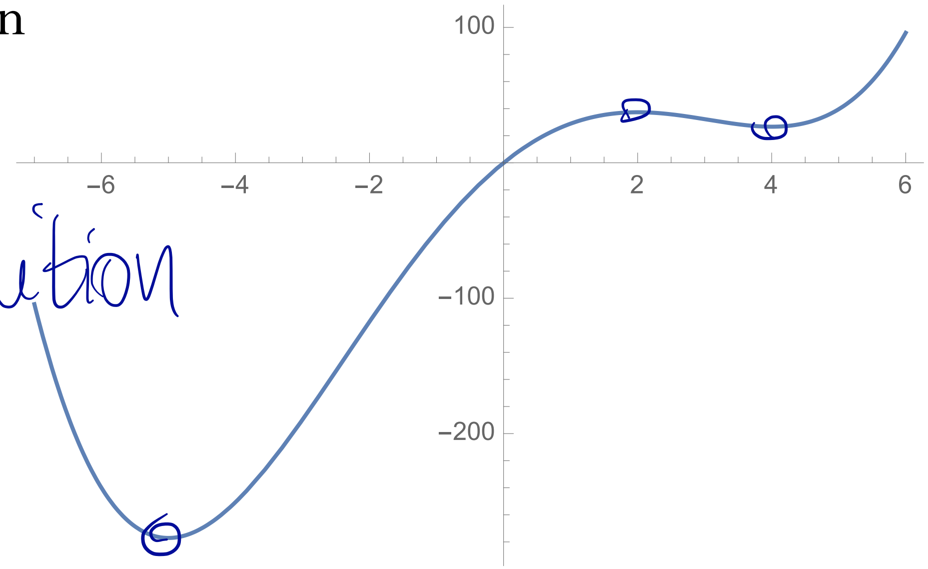
$\Rightarrow$ negative_definite $\Rightarrow$ maximizer

★ if some λ_i are positive, some negative $\Rightarrow$ indefinite

$\Rightarrow$ saddle point

Example (1D)

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



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

First order necessary condition

$f'(x) = 0$

$x = -5$, $x = 2$, $x = 4$

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

$f''(2) = 12 - 4 - 22 = -14$ (max)

$f''(4) = 48 - 8 - 22 = 18$ (min)

Example (ND)

Consider the function $f(x_1, x_2) = 2x_1^3 + 4x_2^2 + 2x_2 - 24x_1$
Find the stationary point and check the sufficient condition

$$\nabla f = \begin{bmatrix} 6x_1^2 - 24 \\ 8x_2 + 2 \end{bmatrix} \qquad H_f = \begin{bmatrix} 12x_1 & 0 \\ 0 & 8 \end{bmatrix}$$

1st order:

$$\nabla f = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6x_1^2 - 24 \\ 8x_2 + 2 \end{bmatrix}$$

$$x_1 = \pm 2$$

$$x_2 = -0.25$$

2nd order

$$H_f\begin{pmatrix} 2 \\ -0.25 \end{pmatrix} = \begin{bmatrix} 24 & 0 \\ 0 & 8 \end{bmatrix}$$

$\Rightarrow$ pos. def $\Rightarrow$ min

$$H_f\begin{pmatrix} -2 \\ -0.25 \end{pmatrix} = \begin{bmatrix} -24 & 0 \\ 0 & 8 \end{bmatrix}$$

$\Rightarrow$ indefinite $\Rightarrow$ saddle

Optimization in 1D: Golden Section Search

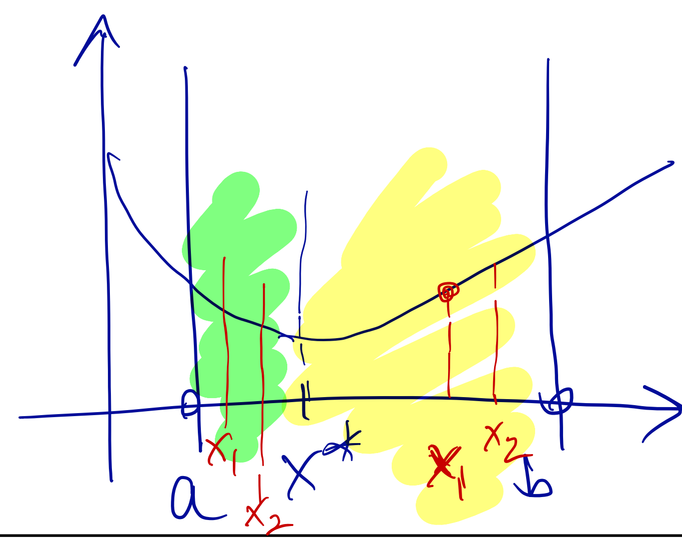
- Similar idea of bisection method for root finding
- Needs to bracket the minimum inside an interval
- Required the function to be unimodal

A function $f: \mathcal{R} \to \mathcal{R}$ 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 $[a, b]$

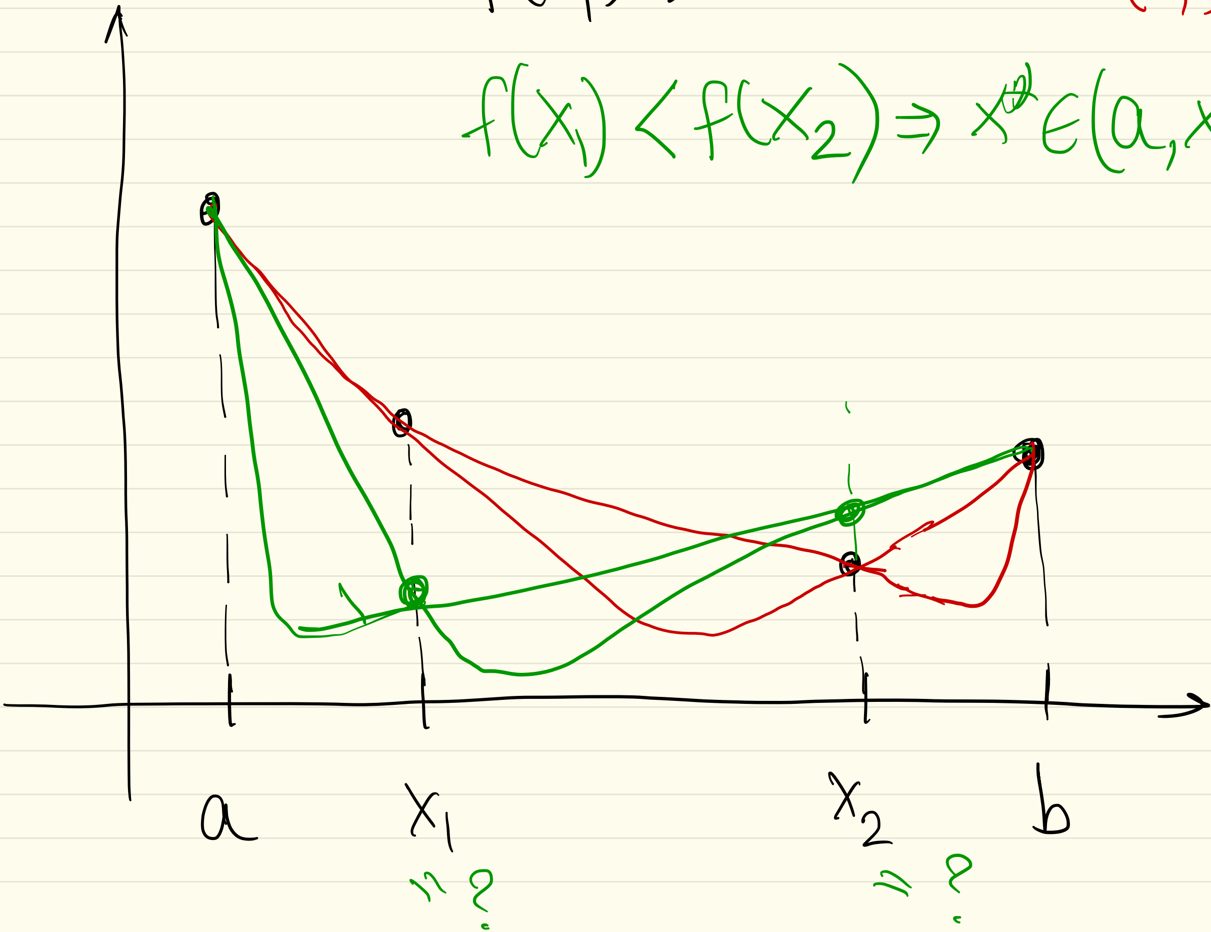
✓ For any $x_1, x_2 \in [a, b]$ with $x_1 < x_2$

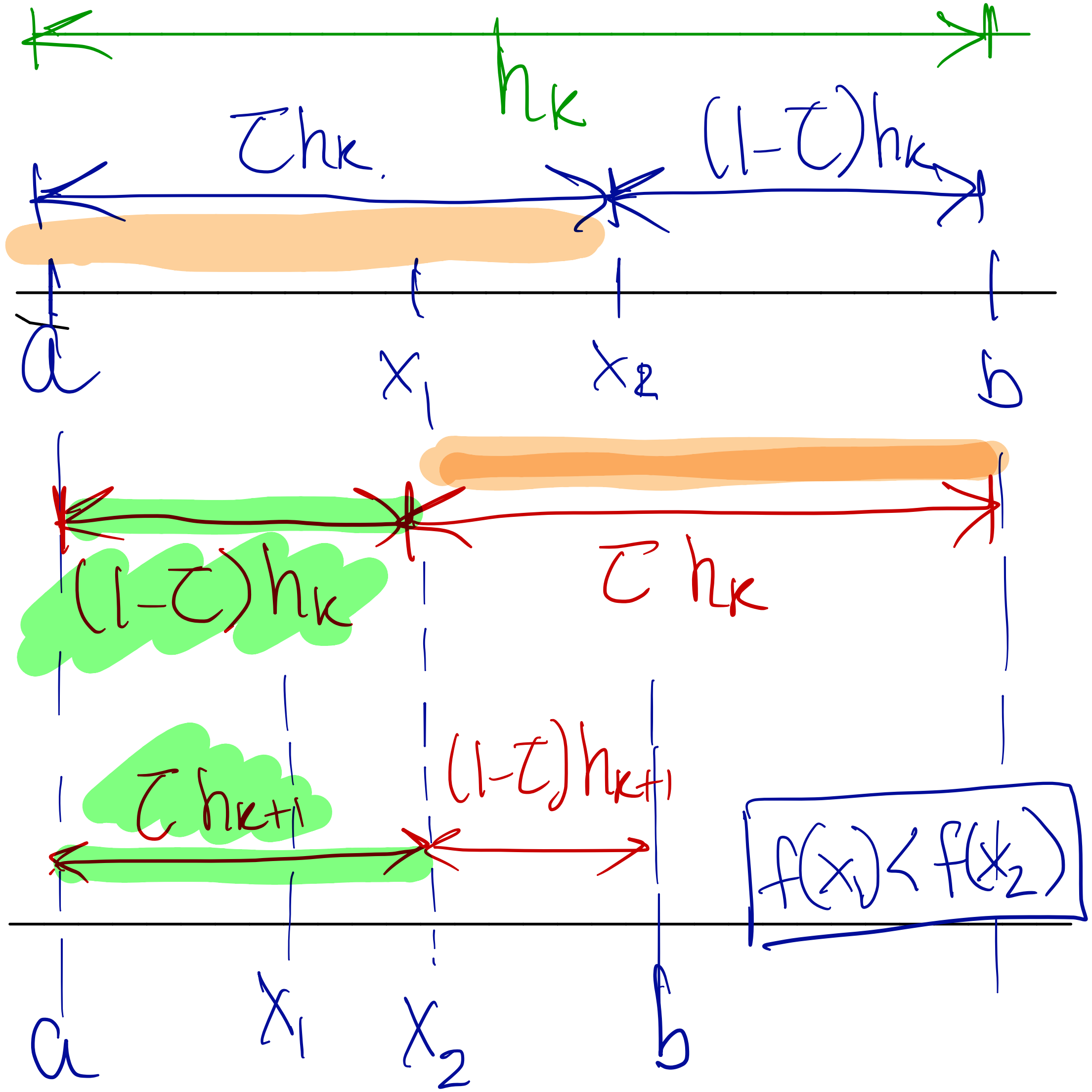
- $x_2 < \boldsymbol{x}^* \Longrightarrow f(x_1) > f(x_2)$
- $x_1 > \boldsymbol{x}^* \Longrightarrow f(x_1) < f(x_2)$



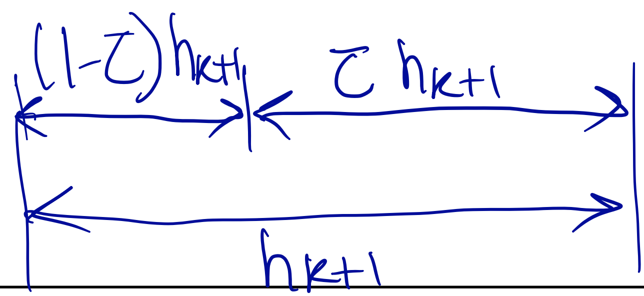
$$f(x_1) > f(x_2) \Rightarrow x^* \in (x_1, b)$$

$$f(x_1) < f(x_2) \Rightarrow x^* \in (a, x_2)$$





$$\boxed{f(x_1) < f(x_2)}$$



Suppose:

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

$x_2 = a + \tau h_k$

$h_k = b - a$

New interval:

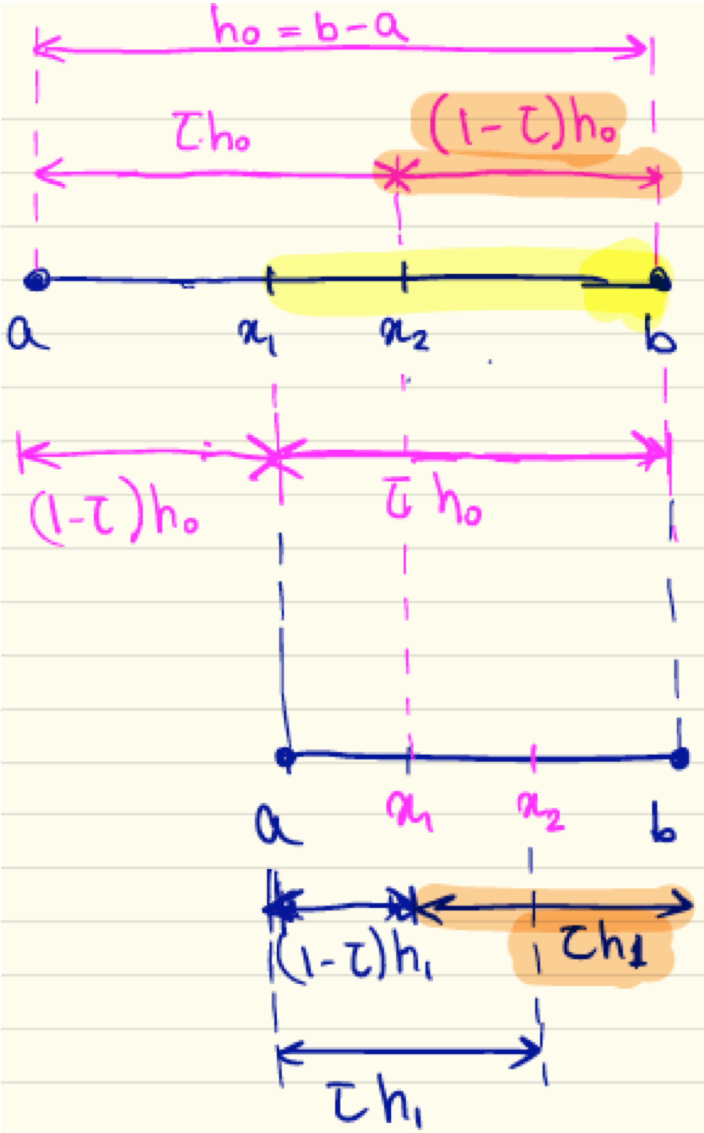
$h_{k+1} = \tau h_k$

at each step, interval gets reduced by τ

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

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

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

Demo: Golden Section Proportions

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

- Derivative free method!

- Slow convergence:

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

- Only one function evaluation per iteration

Iclicker question

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

$h_0 = 20$

$h_1 = 0.618\,(20) = 12.36$

Newton's Method

$\tilde{f}(x)$

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

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

And we want to find the minimum of the quadratic function using the first-order necessary condition $f'(x) = 0 \Longrightarrow \tilde{f}'(x) = 0$

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

$$x - x_0 = -f'(x_0)/f''(x_0)$$

$$\boxed{x = x_0 - f'(x_0)/f''(x_0)} \rightarrow \text{stationary condition}$$

Newton's Method

- **Algorithm:**

x_0 = starting guess

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

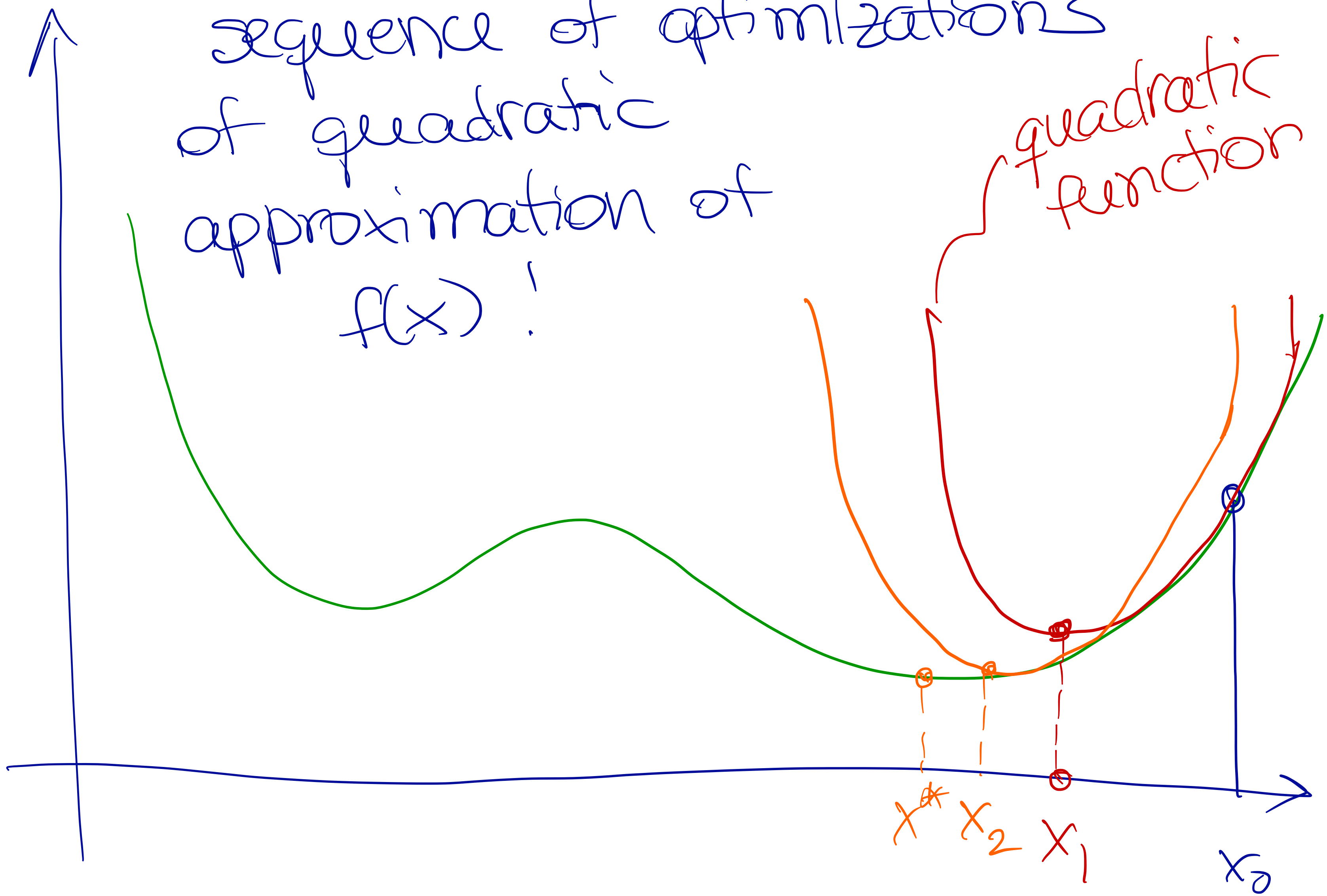
only imposes 1st order necessary condition

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

Demo: "Newton's method in 1D"
And "Newton's method Initial Guess"

Newton's Method (Graphical Representation)

sequence of optimizations of quadratic approximation of f(x)!



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?

$$f'(x) = 12x^2 + 4x + 5 \longrightarrow f'(x_0) = 48 + 8 + 5 = 61$$

$$f''(x) = 24x + 4 \longrightarrow f''(x_0) = 48 + 4 = 52$$

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)} = 2 - \frac{61}{52} = 0.8269$$

If $f(x) = x^2 + 24x - 3$

And $x_0 = 1$

How many iteration of Newton's method will take to converge to x^* such that $f(x^*) = \min f(x)$.

(A) 10 (B) 5 (C) 2 (D) 1 (E) cannot determine