EE 5606 Convex Optimization

Course homepage

https://people.iith.ac.in/shashankvatedka/html/courses/2024/EE5606/course_details.html

Timetable slot - S

Prerequisites:

- math and programming
- strong background in linear algebra/matrix theory
- programming in python - some tutorials on course webpage

Introduction

Why study this course?

Nearly every engineering problem is an optimization problem

Engineering problem $\longrightarrow$ Mathematical optimization problem

$\downarrow$

Solve this

- Objective function
- Variable
- Constraints

Examples

1. Chip design

2. Wireless communication

$$\text{Voice} \to \boxed{} \to \text{Channel} \to \boxed{} \to \text{Voice}$$

$$\downarrow$$

$$X^k \to \boxed{f} \to C^n \to \boxed{CH} \to Y^n \to \boxed{g} \to \hat{X}^k$$

Rate : $R = \dfrac{k}{n}$

Probability of error:

$$\Pr[\hat{X}^k \neq X^k]$$

$$\max_{P_e \le 10^{-3}} R$$

Examples

3. Signal denoising

$$B = A + N$$

N ↓ Gaussian

$$A \to \boxed{T} \to B \to \boxed{f} \to \hat{A}$$

A ↓ Clean image

B ↓ Noisy image

$\hat{A}$ ↓ Denoised image

Variable : f

Objective function : $\|A - \hat{A}\|_2$

Constraints : - f should be linear

- f

4. Object detection in images

Examples

5. Portfolio optimization

100 \$ 10 stocks

$x_1 \; - \; - \; x_{10}$

Objective : $x_1 \mu_1 + x_2 \mu_2 + \text{---} + x_{10} \mu_{10}$

Constraints : $x_1 + x_2 + \cdots + x_{10} \leq 100$

$x_1 \leq b_1 \quad x_2 \leq b_2 \quad \text{---}$

6. Industrial control

$x(t) \longrightarrow \boxed{} \longrightarrow y(t) \qquad \bar{y}(t)$

Variable : $x(t)$

Objective : $\int_0^\infty |y(t) - \bar{y}(t)|^2 \, dt$

Constraint : $\int |x(t)|^2 \, dt \leq P$

Formal definition of a minimization problem

Variable : $\underline{x} \in \mathbb{R}^n$

Objective fn : $g(\underline{x})$ $\quad g: \mathbb{R}^n \longrightarrow \mathbb{R}$

Constraints :

$$g_1(\underline{x}) \leq b_1$$
$$g_2(x) \leq b_2$$
$$\vdots$$
$$g_m(x) \leq b_m$$

$$\boxed{\text{PROBLEM :} \quad \begin{aligned} \min \quad & g(\underline{x}) \\ \text{ST} \quad & g_1(x) \leq b_1 \\ & g_2(x) \leq b_2 \\ & \vdots \\ & g_m(x) \leq b_m \end{aligned}}$$

$$\equiv \quad \begin{aligned} \min \quad & g(\underline{x}) \\ \text{ST} \quad & \bar{g}(\underline{x}) \leq \underline{b} \end{aligned}$$

Is this definition general enough?

(1) Maximization :

$$\max_{\underline{x}} h(\underline{x}) \quad \text{s.t.} \quad \bar{g}(\underline{x}) \le \underline{b} \qquad \equiv \qquad \min_{\underline{x}} \left(-h(\underline{x})\right) \quad \text{s.t.} \quad \bar{g}(x) \le \underline{b}$$

(2)

$$\min_{\underline{x}} h(\underline{x}) \quad \text{s.t.} \quad g_1(x) \le b_1, \quad g_2(x) \ge b_2 \qquad \equiv$$

$$\min x^2 \qquad \equiv \qquad \min x^2$$
$$x^3 \le 2 \qquad\qquad x^3 \le 2$$
$$e^x \ge 3 \qquad\qquad -e^{+x} \le -3$$

③ $\min f(x)$ $\equiv$ $\min f(x)$

$x < 1$ $\quad$ ST $g_1(x) \le b_1$

$\equiv$ $\min f(x)$

ST $g_1(x) \le 0$

$$g_1(x) = \begin{cases} 0 & x < 1 \\ 1 & x \ge 1 \end{cases}$$

$x' = x + \delta$

$\min. \ f(x' - \delta)$

ST

$x' \le 1$

$\delta \le 1 - x$

$x \le 1 - h$

$h \ge 0$

$-h$

$$\min_{\underline{x} \in S} f(x) \equiv \min_{g(\underline{x}) \leq 0} f(x)$$

$$g(\underline{x}) = \begin{cases} 0 & \text{if } x \in S \\ 1 & \text{else} \end{cases}$$

Class 2:

Problem: $\min\limits_{g(\underline{x}) \le \underline{b}} f(\underline{x}) \equiv \min\limits_{\underline{x} \in S} f(\underline{x})$

S: constraint set

① Does this have a solution?

No: $S = \{\underline{x} : g(x) \geq b\}$ may be empty.

Any $\underline{x} \in S$ is called feasible

Eg: $\{x : x \geq 0 \ \& \ x \leq -1\}$

② Is the solution unique?

Not necessary

$f(x) = |x^2 - 1|$

(3) How do we find the solution?

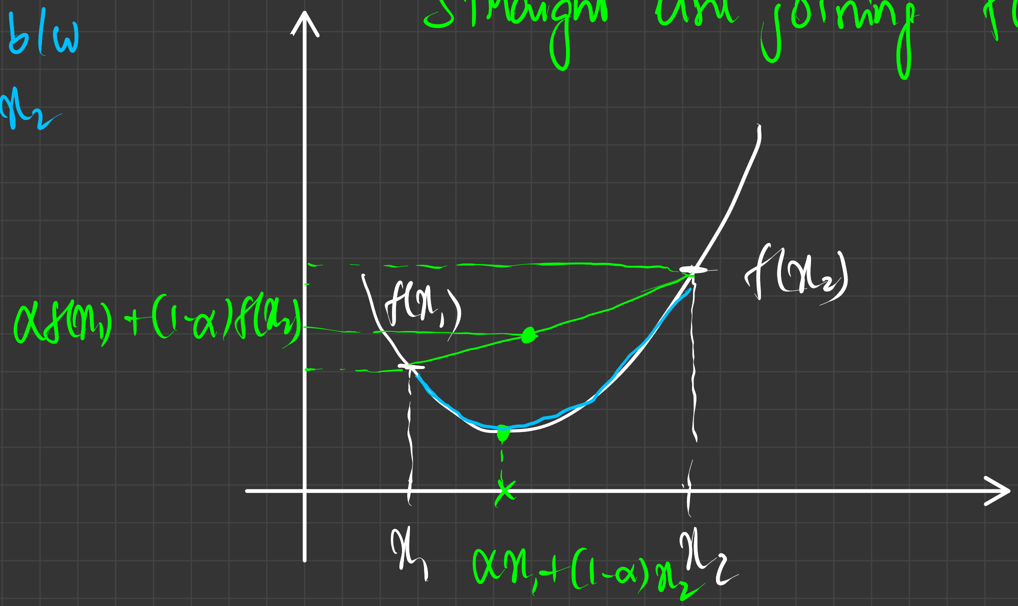
Convex functions over the reals

Defn : $f : \mathbb{R} \to \mathbb{R}$

f is convex iff

$\forall\, x_1, x_2 \in \mathbb{R}, \quad \alpha \in [0,1],$

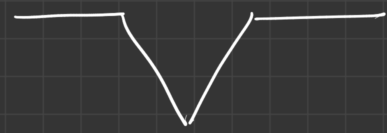
$$\underbrace{f(\alpha x_1 + (1-\alpha) x_2)}_{\text{fn value b/w } x_1 \,\&\, x_2} \le \underbrace{\alpha f(x_1) + (1-\alpha) f(x_2)}_{\text{Straight line joining } f(x_1) \,\&\, f(x_2)}$$



f is concave if $-f$ is convex

Theorem: If $f : \mathbb{R} \to \mathbb{R}$ twice differentiable

f is convex $\iff$ $f''(x) \geq 0 \quad \forall x$



Examples:

① $f(x) = x^2$

$f''(x) = 2 \quad \forall x$

≥ 0

$\therefore$ convex

② e^x

$f''(x) = e^x \geq 0 \quad \forall x \in \mathbb{R}$ convex

③ e^{-x}

$f''(x) = e^{-x} \geq 0 \quad \forall x \in \mathbb{R}$ convex

④ $\log x$, $x > 0$ $\quad f''(x) = -\frac{1}{x^2} \leq 0 \quad \forall x \in (0, \infty)$ concave

⑤ $f(x) = \dfrac{1}{1+x^2}$

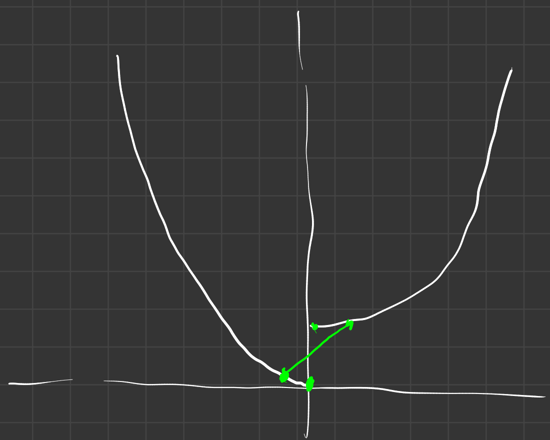
$$f'(x) = -\frac{2x}{(1+x^2)^2}$$

$$f''(x) = -2x^4 + 8x^3 - 4x^2 + 8x - 2$$

$$\frac{2}{(1+x^2)^3} \quad (3x^2-1)$$

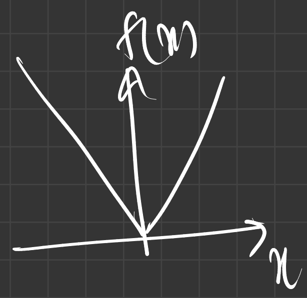
Not convex.

⑥ $f(x) = \begin{cases} x^2, & x \le 0 \\ x^2+1, & x > 0 \end{cases}$



Not convex/

Concave



7 $f(x) = x^3$ $\qquad f''(x) = 6x$

8 $f(x) = |x|$

$$
\begin{aligned}
f(\alpha x_1 + (1-\alpha) x_2) &= |\alpha x_1 + (1-\alpha) x_2| \\
&\leq |\alpha x_1| + |(1-\alpha) x_2| \quad \text{(Triangle inequality)} \\
&= \alpha |x_1| + (1-\alpha) |x_2| \\
&= \alpha f(x_1) + (1-\alpha) f(x_2)
\end{aligned}
$$

Theorem: If $f : \mathbb{R} \to \mathbb{R}$ twice differentiable

f is convex $\iff$ $f''(x) \geq 0 \quad \forall x$

Proof: Suppose f is convex.

$$f'(x) = \lim_{t \to 0} \frac{f(x+t) - f(x)}{t}$$

$$f''(x) = \lim_{t \downarrow 0} \frac{f(x+t) + f(x-t) - 2f(x)}{t^2}$$

Consider $f(x) = f\left(\frac{x+t}{2} + \frac{x-t}{2}\right)$

$$= f\Big(\underbrace{\tfrac{1}{2}(x+t) + \tfrac{1}{2}(x-t)}_{\alpha x_1 + (1-\alpha)x_2}\Big)$$

$$\leq \frac{1}{2} f(x+t) + \frac{1}{2} f(x-t)$$

$$f(x+t) + f(x-t) - 2f(x) \geq 0 \quad \forall x \in \mathbb{R}$$

$$\Rightarrow \quad f''(x) \geq 0 \quad \forall x \in \mathbb{R}.$$

Now, suppose $f''(x) \geq 0 \quad \forall x \in \mathbb{R}.$

$$\alpha f(x_1) + (1-\alpha) f(x_2) - f(\alpha x_1 + (1-\alpha) x_2)$$

Define $x = \alpha x_1 + (1-\alpha) x_2$, Say $x_1 < x_2$

$$\alpha f(x_1) + (1-\alpha) f(x_2) - f(x)$$

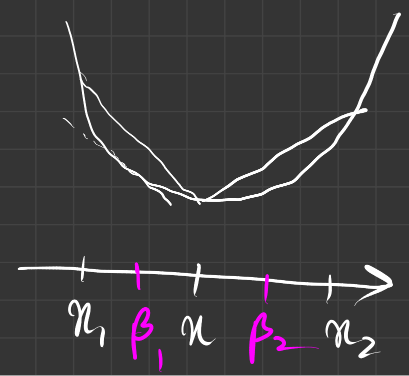
$$= \alpha\big(f(x_1) - f(x)\big) + (1-\alpha)\big(f(x_2) - f(x)\big)$$

$\exists\ \beta_1 \in (x_1, x) \qquad \beta_2 \in (x, x_2)$

MVT

$$= \alpha(x_1 - x) f'(\beta_1) + (1-\alpha)(x_2 - x) f'(\beta_2)$$

$=$



Sub. for x

$$= \alpha\left(x_1 - \alpha x_1 - (1-\alpha)x_2\right) f'(\beta_1)$$
$$+ (1-\alpha)\left(x_2 - \alpha x_1 - (1-\alpha)x_2\right) f'(\beta_2)$$
$$= -\alpha(1-\alpha)(x_2 - x_1) f'(\beta_1)$$
$$+ (1-\alpha)\alpha(x_2 - x_1) f'(\beta_2)$$
$$= \alpha(1-\alpha)(x_2 - x_1)\left(f'(\beta_2) - f'(\beta_1)\right)$$

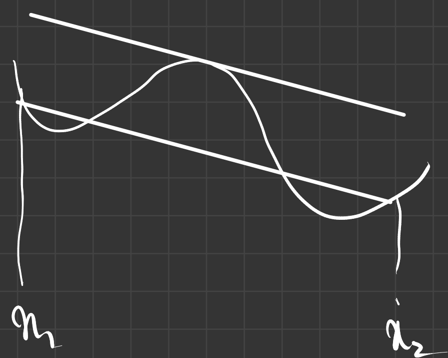
$\beta_2 > \beta_1$

$f''(x) \geq 0 \ \forall x \Rightarrow f'(x)$ is monotone increasing (nondecreasing)

$\Rightarrow f'(\beta_2) \geq f'(\beta_1)$

$\Rightarrow \alpha f(x_1) + (1-\alpha) f(x_2) - f(x) \geq 0 \Rightarrow f$ is convex

MVT



MVT: If $f'(x)$ exists for all $x \in [x_1, x_2]$ $\exists$ $x' \in (x_1, x_2)$ st

$$f'(x') = \frac{f(x_1) - f(x_2)}{x_1 - x_2}$$

Unconstrained minimization of convex $f: \mathbb{R} \to \mathbb{R}$

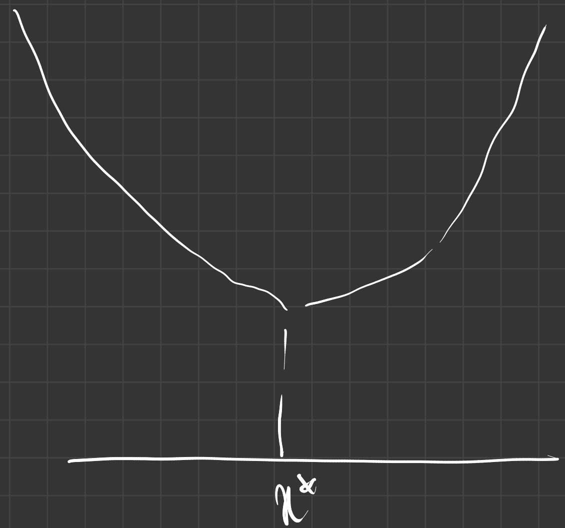
Solve for $f'(x) = 0$

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

$\forall x < x^*$, $f'(x) \leq 0$

$\Rightarrow$ f is decreasing for $x < x^*$

$\Rightarrow$ $f(x) \geq f(x^*)$



$\forall x > x^*$, $f'(x) \geq 0$

$\Rightarrow$ f is increasing

$\Rightarrow$ $f(x) \geq f(x^*)$

Capacity of a multiple-antenna (MIMO) wireless channel

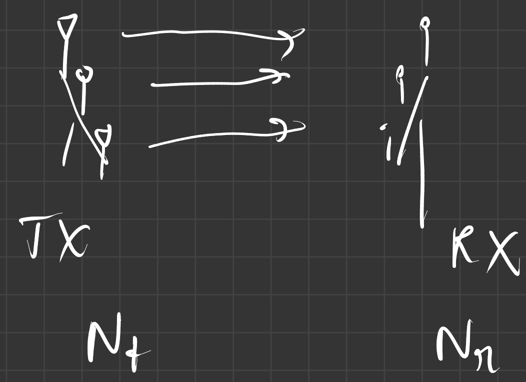
N_t-length vector

$$y_t = H\underline{x}_t + z_t$$

$N_r \times N_t$

Gaussian

$\mathcal{N}(0, \sigma^2)$



$$\mathbb{E}\|x_t\|_2^2 \le P$$

Max. rate of reliable communication

$$C = \max_{\substack{A:\ \text{P.S.D} \\ \& \ \mathrm{tr}(A) \le P}} \log\det\left(\sigma^2 I + H A H^T\right)$$

Recap

f is convex if $\forall x_1, x_2 \in \mathbb{R}$, $\alpha \in [0,1]$,

$$f(\alpha x_1 + (1-\alpha) x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$$

If $f : \mathbb{R} \to \mathbb{R}$ twice differentiable

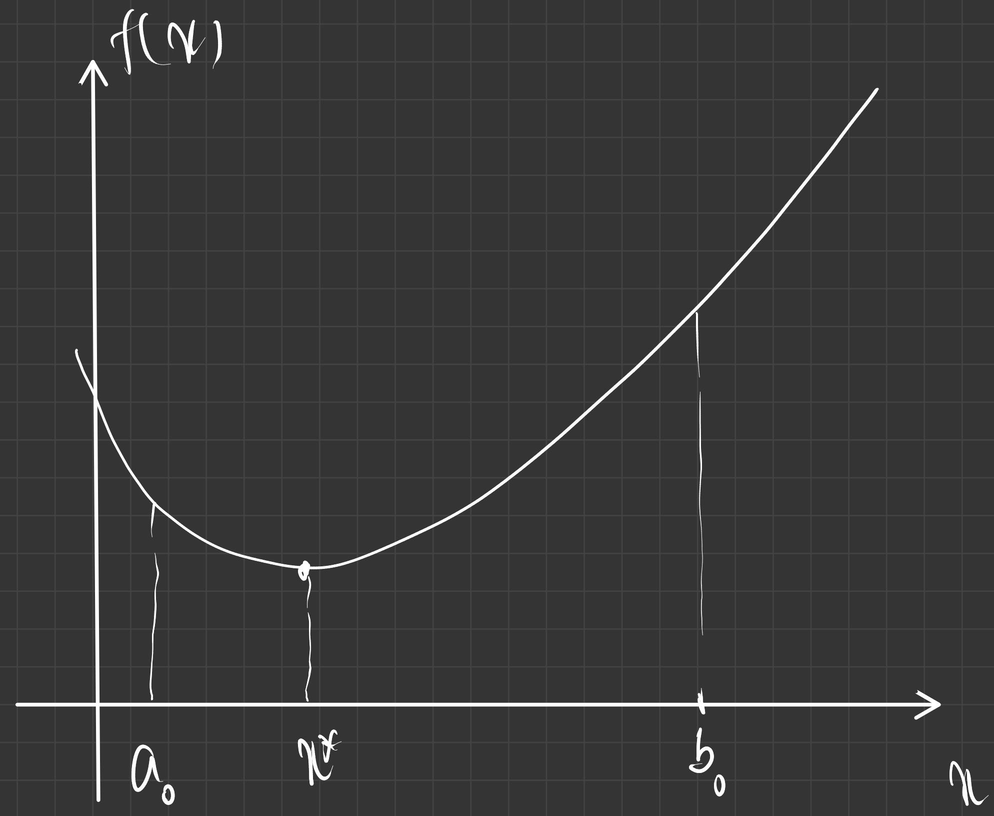
f is convex $\iff$ $f''(x) \geq 0 \quad \forall x$

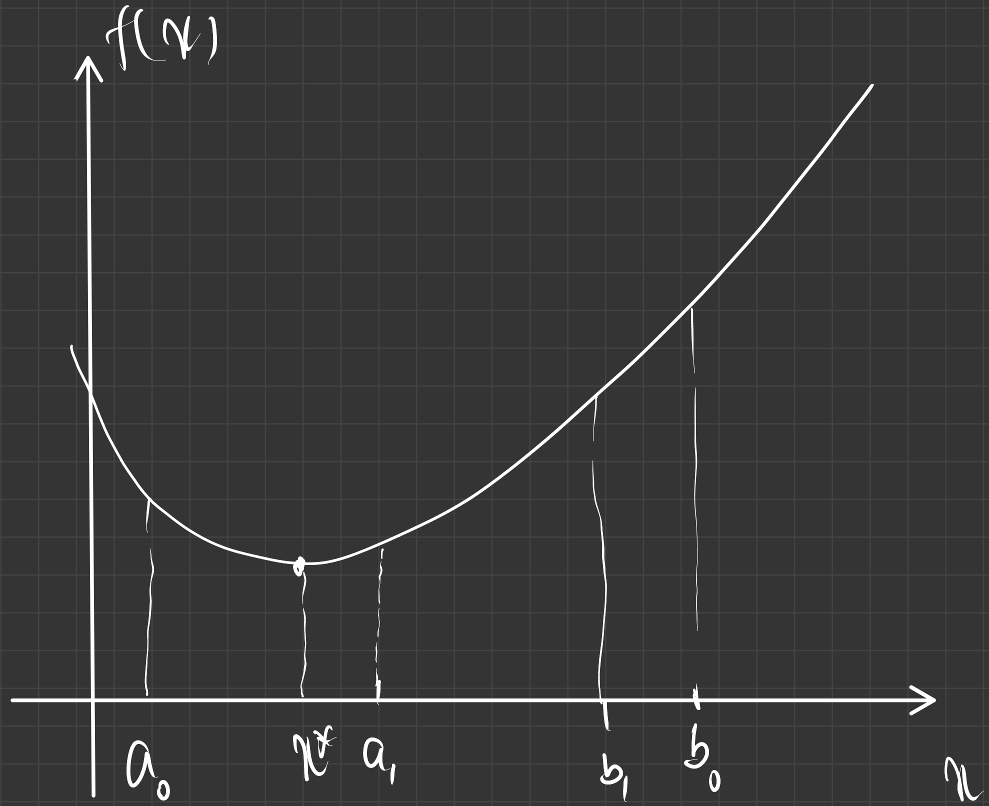
Unconstrained minimization of convex $f : \mathbb{R} \to \mathbb{R}$

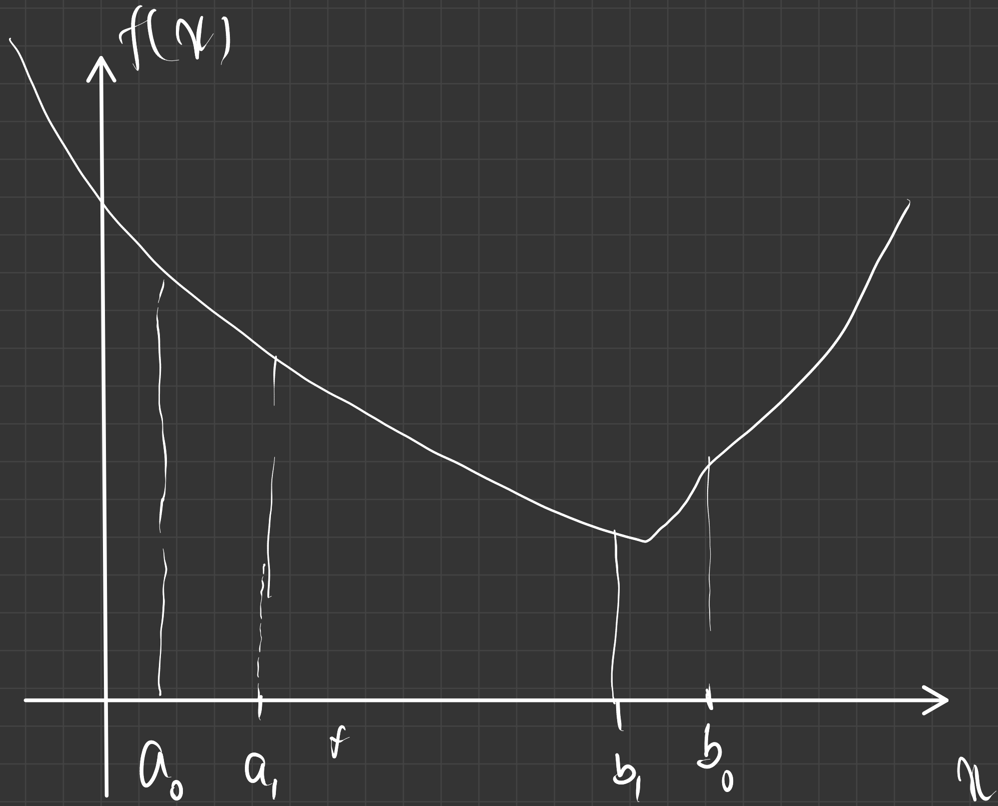
Solve for $f'(x) = 0$

Numerically solving 1-d convex optimization problems

Goal:
$$\min_{x \in [a_0, b_0]} f(x)$$




f(x)
a0
x*
a1
b1
b0
x


f(x)
a0
a1
b1
b0
x


a0
a1
b1
b0

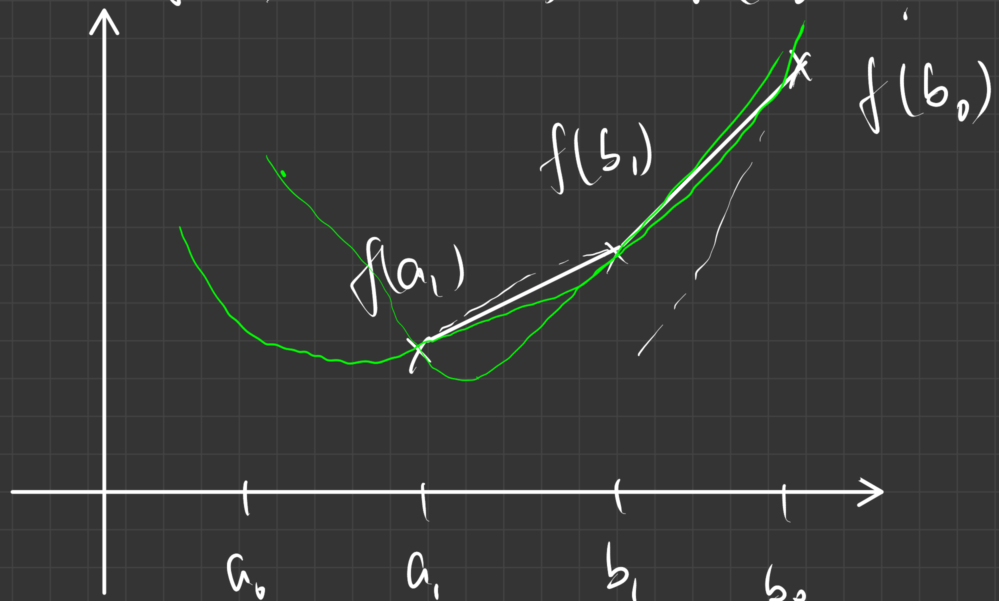
Observations: If $x^* \in [a_0, a_1]$

$$f(a_1) \leq f(b_1) \leq f(b_0)$$

If $x^* \in [b_1, b_0]$

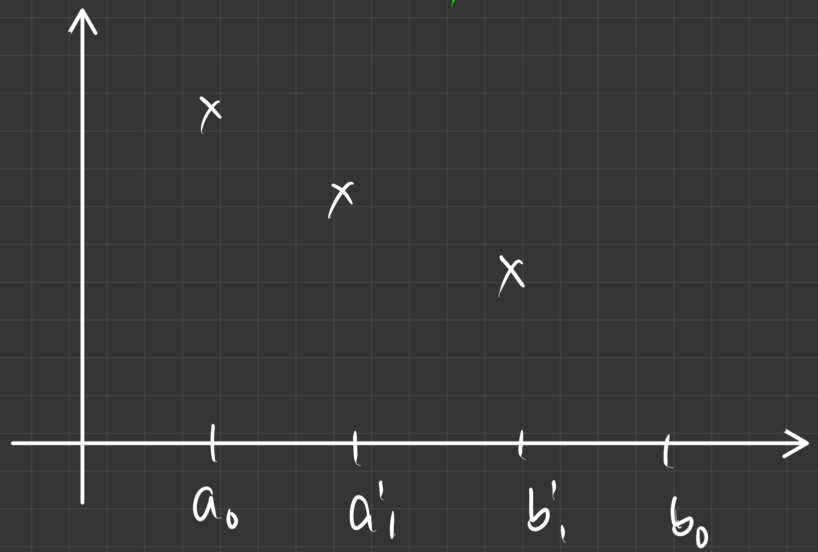
can we have

$$f(a_1) < f(b_1) < f(b_0) \ ?$$



If $f(a_1) \leq f(b_1) \leq f(b_0)$, then $x^* \in [a_0, b_1]$

If $f(a_0) \geq f(a_1) \geq f(b_1)$, then $x^* \in [a_1, b_0]$



$[a_{i-1}, b_{i-1}]$

Algo:

a_0, b_0

for $i = 1, 2, 3 \ldots$

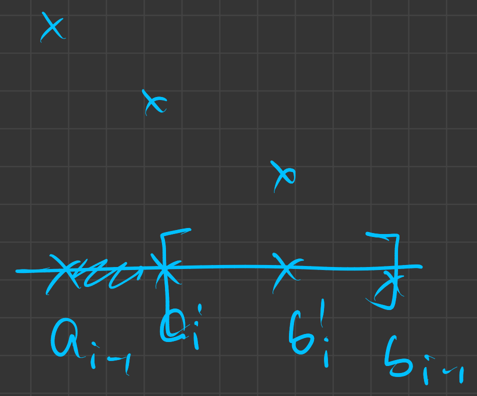
* $a_i' = a_{i-1} + \delta$
* $b_i' = b_{i-1} - \delta$
* if $f(a_{i-1}) \geq f(a_i') \geq f(b_i')$

  $a_i = a_i'$

  $b_i = b_{i-1}$

  else: $a_i = a_{i-1}, \quad b_i = b_i'$

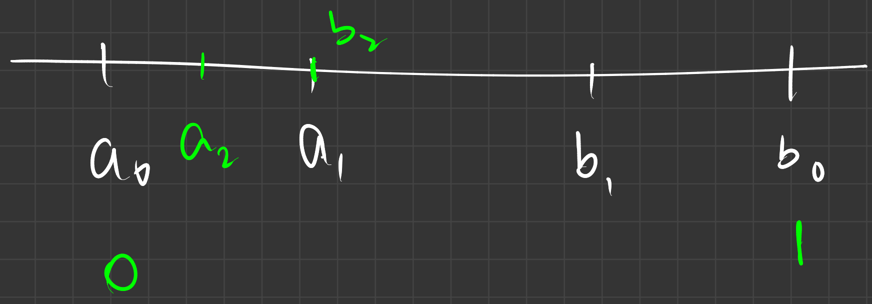
$[a_{i-1}, b_{i-1}]$



How to choose δ?

* $a_{i-1} + \delta < b_{i-1} - \delta \Rightarrow \delta < \frac{b_{i-1} - a_{i-1}}{2}$

* Can we reduce to 1 eval per iteration?



$b_2 = a_1$ OK $a_2 = b_1$

Suppose: $a_1 = a_0 + \rho(b_0 - a_0)$

$b_1 = b_0 - \rho(b_0 - a_0)$

$a_2 = a_0 + \rho(b_1 - a_0)$

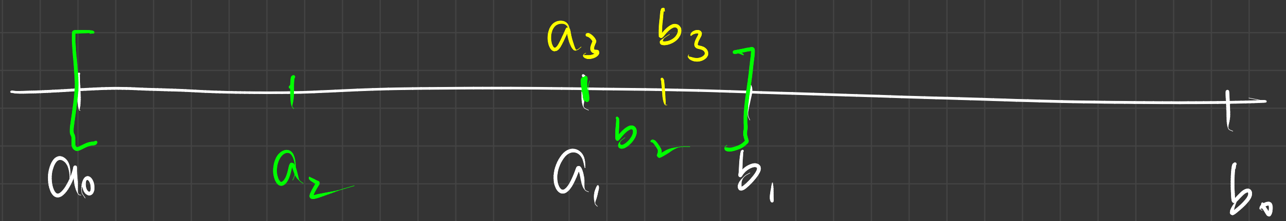
$b_2 = b_1 - \rho(b_1 - a_0)$

Want: $b_2 = a_1$

$b_1 - \rho(b_1 - a_0) = a_0 + \rho(b_0 - a_0)$

$b_0 - \rho - \rho(b_0 - \rho - a_0) = a_0 + \rho$

$1 - \rho - \rho(1 - \rho) = \rho$



In each iteration,

$$a_i' = a_{i-1} + \rho\,(b_{i-1} - a_{i-1})$$

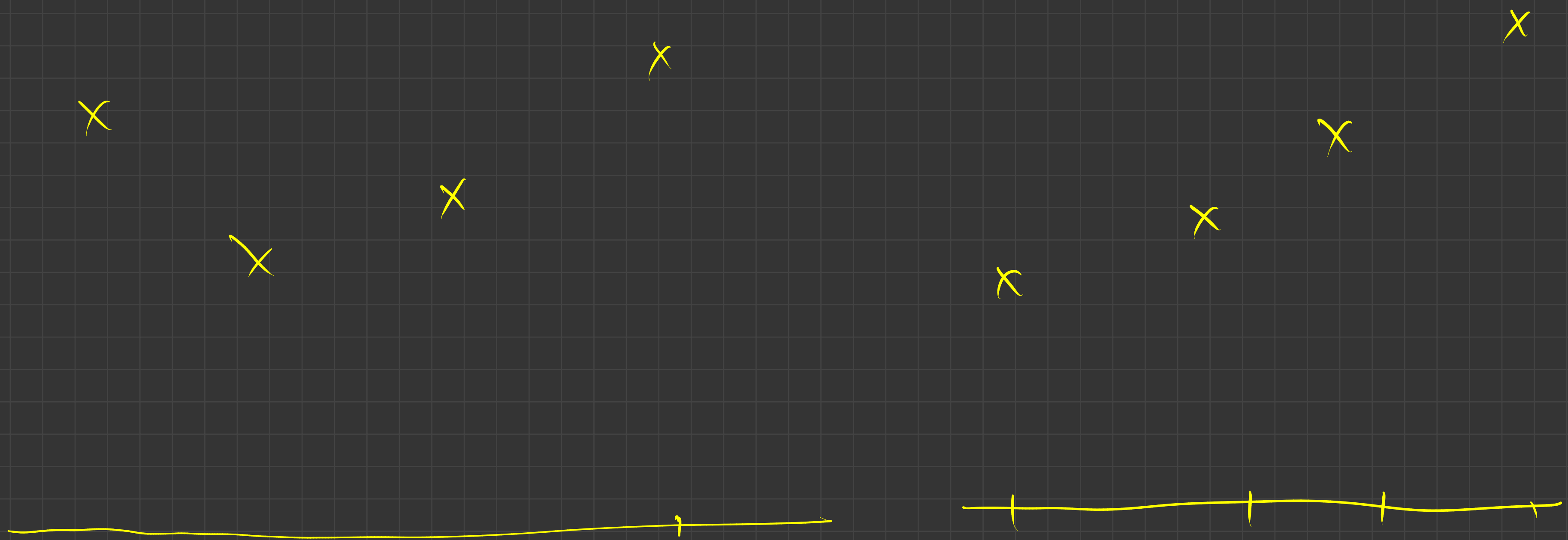
$$b_i' = b_{i-1} - \rho\,(b_{i-1} - a_{i-1})$$

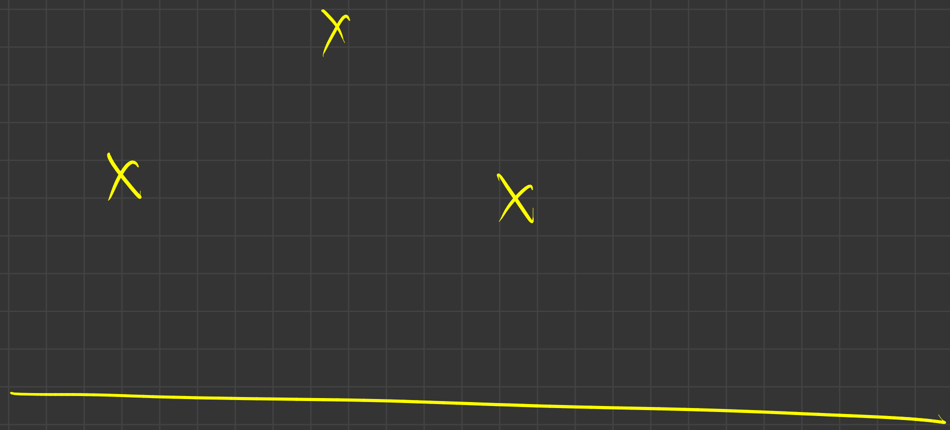
$$1-2\rho = \rho(1-\rho)$$

$$\rho^2 - 3\rho + 1 = 0$$

$$\rho = \frac{3 \pm \sqrt{5}}{2}$$

$$\rho = \frac{3-\sqrt{5}}{2} \approx 0.382\ldots$$





$$\frac{b_i - a_i}{b_{i-1} - a_{i-1}} = 1-\rho \approx 0.62$$

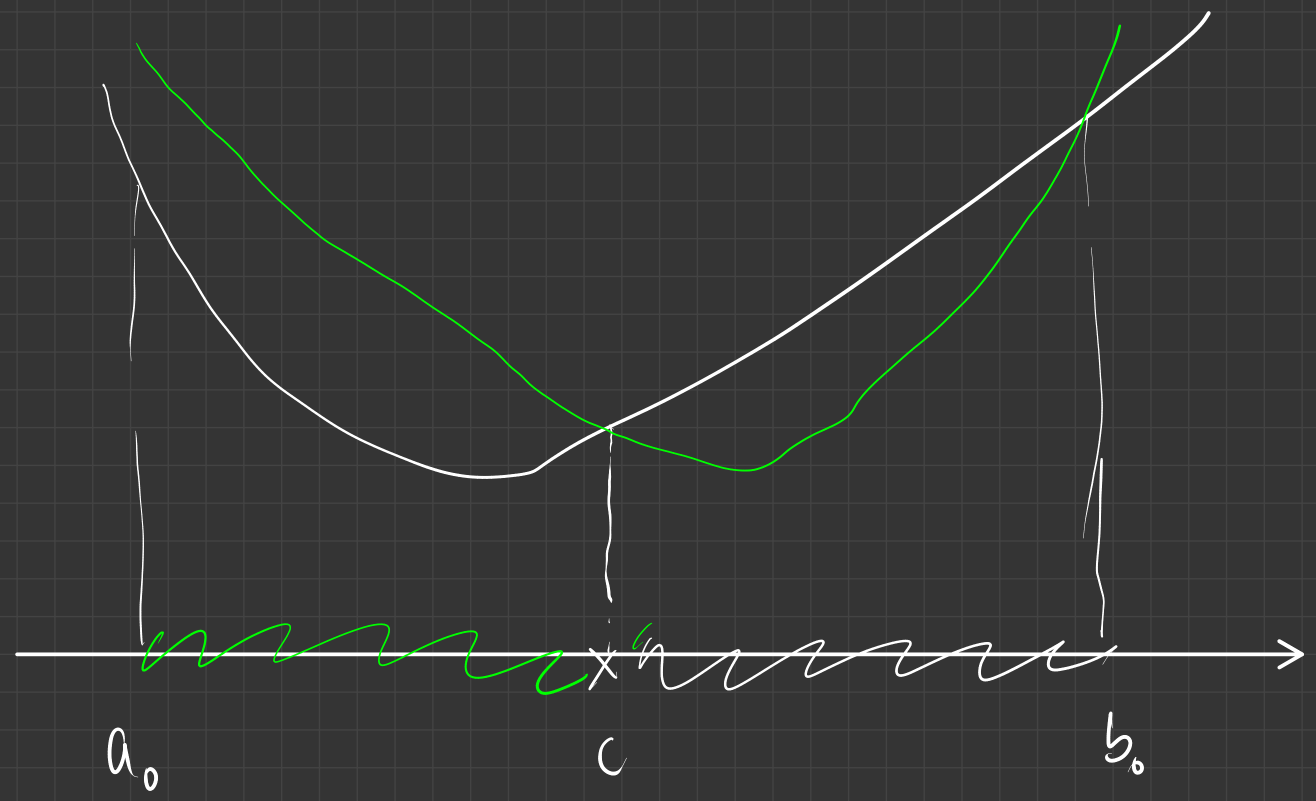
$$\frac{b_i - a_i}{b_0 - a_0} = (1-\rho)^i$$

To compute x^* to accuracy ε, No of iterations:

$$(1-\rho)^N \le \varepsilon$$

$$N \ge (\log \varepsilon)/\log(1-\rho) = \frac{\log \frac{1}{\varepsilon}}{\log 1/(1-\rho)}$$

Bisection method



Algorithm : a_0 , b_0

for $i = 1, 2, 3, \dots$

$$c = \frac{a_{i-1} + b_{i-1}}{2}$$

if $f'(c) > 0$:

$b_i = c$

$a_i = a_{i-1}$

if $f'(c) < 0$

$b_i = b_{i-1}$

$a_i = c$

else :

stop & o/p c

Recap: Algorithms for 1-d convex optimization

1. Golden section search: needs only $f(x)$

2. Bisection method: needs $f'(x)$

Newton's method

$f : \mathbb{R} \to \mathbb{R}$ convex & twice differentiable.

$$g(x) = a_0 + a_1(x - x_1) + a_2(x - x_1)^2$$

$g(x)$ is convex iff $a_2 \geq 0$

$x_t \to$ current pt

$$g_t(x) = f(x_t) + (x - x_t) f'(x_t) + \frac{(x - x_t)^2}{2} f''(x_t)$$

surrogate function

Is $g_t(x)$ also convex?

Yes, since $f''(x_t) \geq 0$

$$\min_{x \in \mathbb{R}} g_t(x)$$

$$g_t'(x) = 0 \Rightarrow f'(x_t) + (x - x_t) f''(x_t) = 0$$

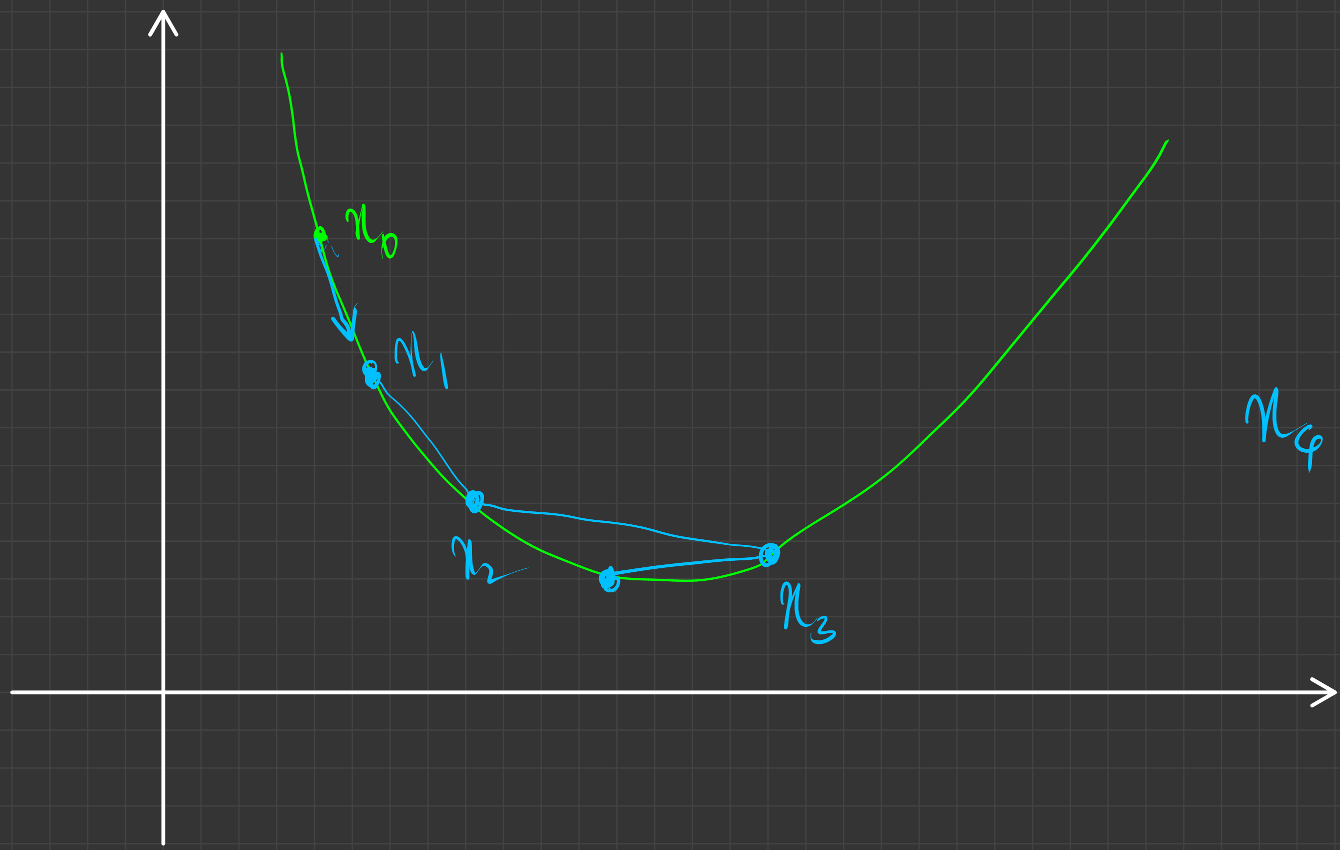
$$x = x_t - \frac{f'(x_t)}{f''(x_t)}$$

Algorithm: f, x_0

for $i = 1, 2, 3, \ldots$

$$x_i = x_{i-1} - \frac{f'(x_{i-1})}{f''(x_{i-1})}$$

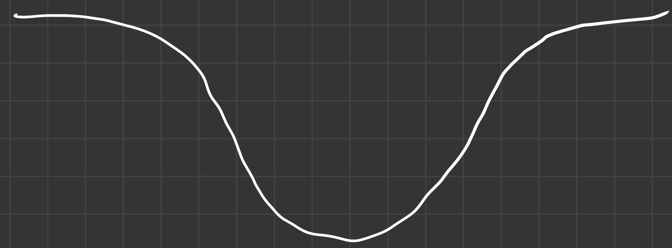
$|x_i - x_{i-1}| \le \varepsilon$, terminate.



$$x_4 = x_3 - \frac{f'(x_3)}{f''(x_3)}$$

Gradient descent : $x_{t+1} = x_t - \gamma_t f'(x_t)$

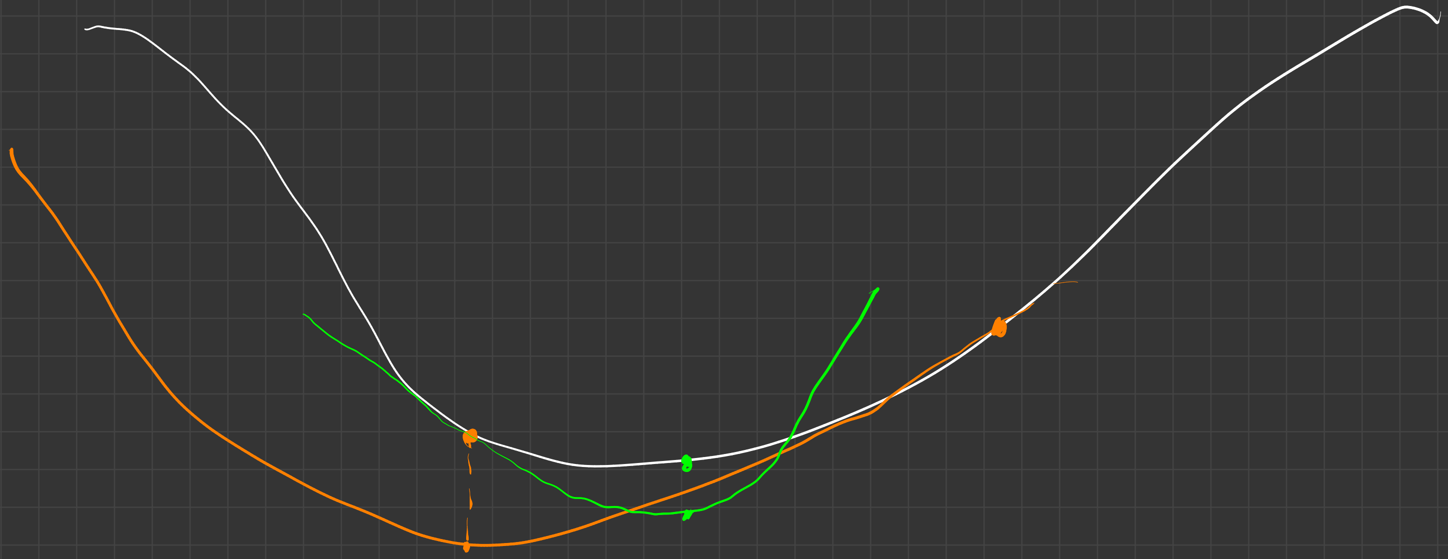
$$f(x) = -e^{-x^2/2}$$



$$f'(x) = x e^{-x^2/2}$$

$$f''(x) = e^{-x^2/2} - x^2 e^{-x^2/2}$$

$$g_t(x) = -e^{-x_t^2/2} + (x - x_t)\left(x e^{-x_t^2/2}\right) + \frac{(x-x_t)^2}{2}\left(e^{-x_t^2/2} - x^2 e^{-x_t^2/2}\right)$$



Some basics of topology and real analysis

Upper and lower bounds; sup and max; inf and min

$S \subseteq \mathbb{R}$

* We say that x is an upper bound for S if

$$y \leq x \quad \forall y \in S$$

lower bound if $\quad y \geq x \quad \forall y \in S$

* We say that x is the supremum of S if

least upper bound.

① x is an upper bound for S

② if y is an upper bound for S, $y \geq x$

* Similarly infimum is the greatest lower bound.

$\sup [1, 2] = 2$ $\max [1, 2] = 2$

$\sup (1, 2) = 2$ $\max [1, 2)$ does not exist

* If $\sup S$ lies in S, we call it the maximum

  $\inf S \in S$, we call it the minimum.

* If S is not bounded from above, $\sup S = \infty$

  below, $\inf S = -\infty$

* Every nonempty $S \subseteq \mathbb{R}$ has sup & inf.

Consider $\underset{x \in (0,1)}{\text{maximize}} \quad x^2$

$$\sup_{x \in (0,1)} x^2 = 1$$

Countable and uncountable sets

A set S is countable if $\exists$ a one-one map from S to $\mathbb{N}$.

Eg: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q} \cong (\mathbb{Z}, \mathbb{Z})$ $\mathbb{Q}^k$

$2\mathbb{Z}$, $2\mathbb{Z}+1$

Eg: $\mathbb{R}$, $\mathbb{C}$, $[0,1]$, etc.

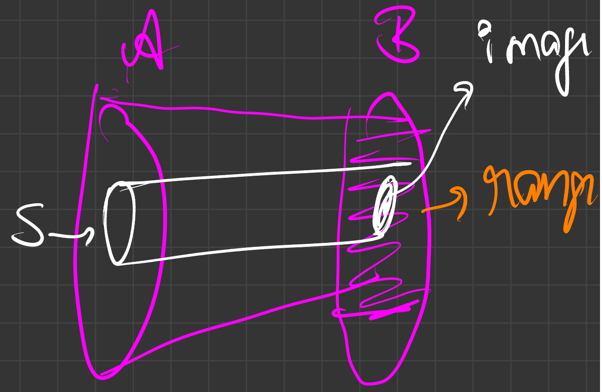
Functions: domain, co-domain, range, image, inverse image

$f : A \to B$

A ↓ domain

B ↘ codomain

$f(A) = \{ f(x) : x \in A \} \longrightarrow$ range



For $S \subseteq A$,

$f(S) = \{ f(x) : x \in S \} \longrightarrow$ image of S under f

$y \in B$,

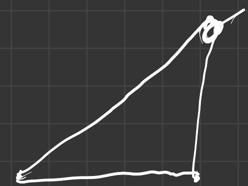
$f^{-1}(y) = \{ x \in A : f(x) = y \}$ inverse image

Metric :

A

A metric $d : A \times A \to \mathbb{R}$

1. $d(x,y) \geq 0 \quad \forall x,y \in A$
2. $d(x,y) = 0$ if & only if $x = y$
3. $d(x,y) = d(y,x)$
4. $d(x,z) \leq d(x,y) + d(y,z)$



Eg :

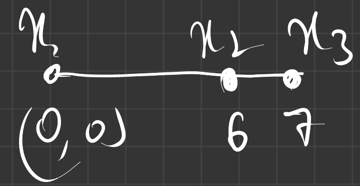
$$d(\underline{x}, y) = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2}$$ is a metric

$$d^2(\underline{x}, y) = \sum_{i=1}^{n} (x_i - y_i)^2$$

$$(x_1 - x_3)^2 = 49$$

$$(x_1 - x_2)^2 = 36$$

$$(x_2 - x_3)^2 = 1$$



$$d_p(\underset{\sim}{x}, \underset{\sim}{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p} \rightarrow L_p \text{ metric}$$

Norm and inner product

Norm : $f : \mathcal{A} \to \mathbb{R}$ satisfying ($\mathcal{A}$ is a vector space over $\mathbb{R}$)

① $f(\underline{x}) \geq 0 \quad \forall x \in \mathcal{A}$

② $f(\underline{x}) = 0$ iff $x = 0$

③ $f(\alpha \underline{x}) = |\alpha| f(\underline{x}) \quad \forall \underline{x} \in \mathcal{A}$, $\alpha \in \mathbb{R}$

④ $f(x+y) \leq f(x) + f(y)$

If f is a norm, then $d(x,y) = f(x-y)$ is a metric.