Announcements

GSI Office Hours:

Haozhu Wang: Wednesday 6:00pm-9:00pm in GGBL1025
Ankit Bansal: Friday 3:00pm-6:00pm in GGBL 1025

Recommended readings (Canvas or Google Drive)

Other Events

Michigan Data Science Team Information Session

September 14, 2017
6-7pm

1013 DOW

subsequent project meetings will take place every Thursday from 5:30-6:30pm in 3725 BBB

midas.umich.edu/mdst

twitter: @mdst_umich

facebook: mdst.umich

MICHIGAN DATA SCIENCE TEAM
UNIVERSITY OF MICHIGAN

Come learn about the Michigan Data Science Team & how to join!

Whether you're new to data science or have years of experience, the Michigan Data Science Team has opportunities for you. Some of these include:

- data science tutorials (intro level & beyond)
- competitions hosted on Kaggle—we are partnering with Quicken Loans for this semester's intra-team contest
- projects partnering with local organizations ranging from the City of Detroit to local startups
- project meetings with guest speakers
- writing papers and presenting at conferences

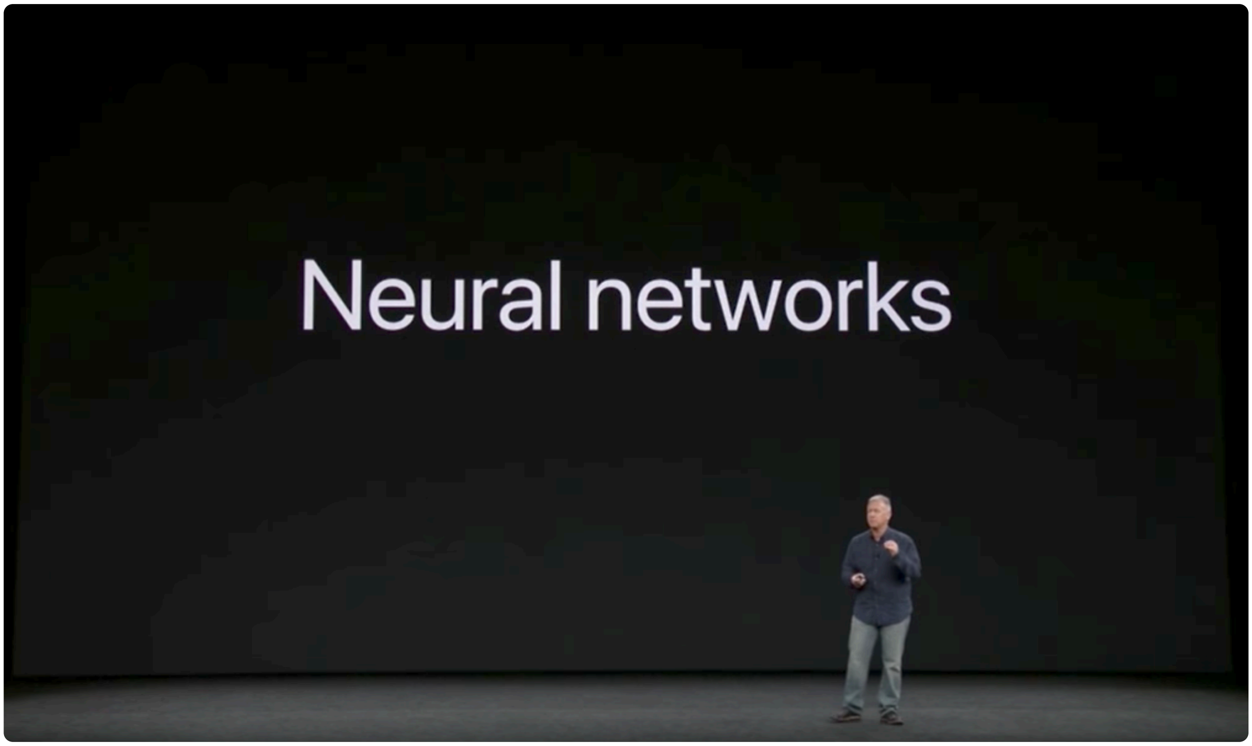
Discover Apple.
We're looking for people like you.

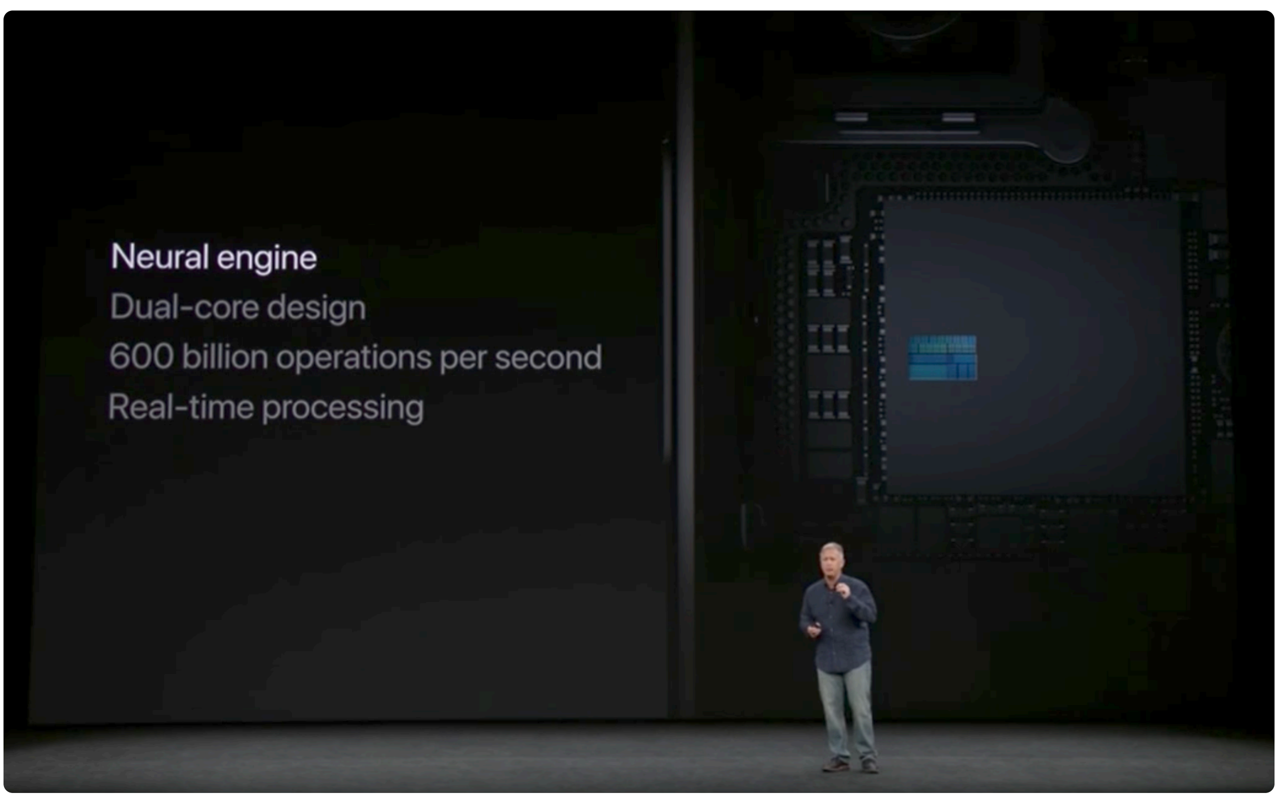
Apple is a place where people from all backgrounds get together to do their life's best work. Come join us.

Engineering Networking Day

Thursday, September 21
3:00 p.m. to 6:00 p.m.
Duderstadt Center Atrium


9:41
Back
Move your head slowly to
complete the circle.
Accessibility Options


Neural networks


Neural engine
Dual-core design
600 billion operations per second
Real-time processing

Outline

Recap EVD SVD

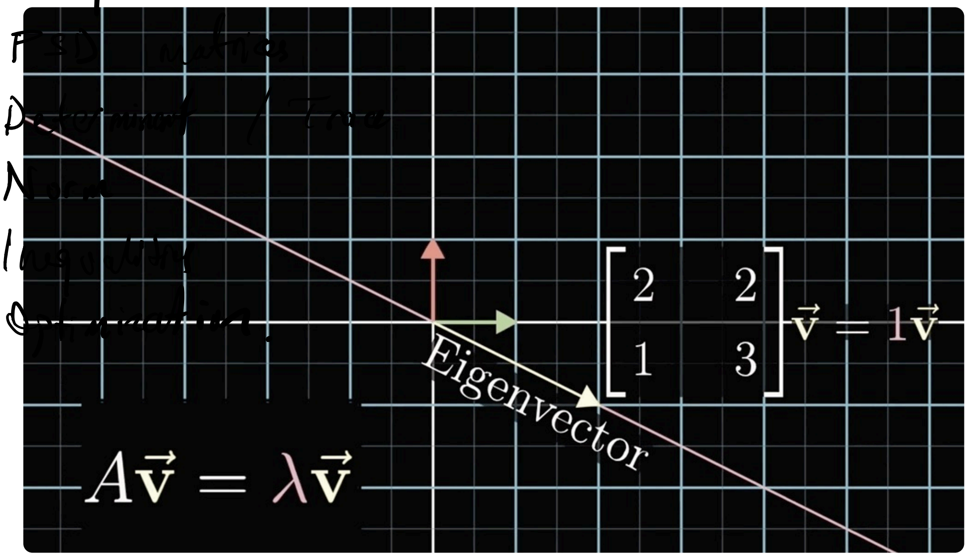
PSD matrices

Determinant / Trace

Norm

Inequalities

Optimization



EVD: $\boxed{A = U\Lambda U^T}$ $\quad A \in \mathbb{R}^{d\times d}$

$Ax = U\Lambda U^T x$

$\boxed{U^T U = U U^T = I}$

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix}$$

$\rightarrow \boxed{A u_i = \lambda_i \cdot u_i}$

$\rightarrow \boxed{A u_j = \lambda_j u_j}$

$$U = \begin{bmatrix} \underline{u_1} & \underline{u_2} & \cdots & u_d \end{bmatrix}$$

$\boxed{A = A^T} \Rightarrow u_i^T u_j = 0$

$$u_j^T A u_i = \lambda_i \cdot \overbrace{u_j^T u_i}$$
$$= u_i^T A u_j = \lambda_j \cdot \underbrace{u_i^T u_j} \Rightarrow (\lambda_i - \lambda_j) \cdot u_i^T u_j = 0$$

if $\neq \Rightarrow \langle u_i, u_j \rangle = 0$

$\boxed{A = \sum_i \lambda_i u_i u_i^T}$

$= \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \cdots +$

$$\underbrace{\dim(\operatorname{colspan}(\lambda_1 u_1 u_1^T))}_{\text{rank}}$$

$$\dim\left(\left\{ \lambda_1 u_1 u_1^T x \;\middle|\; x \in \mathbb{R}^d \right\}\right) = 1$$

SVD : $A \in \mathbb{R}^{n \times d}$

$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_d \end{bmatrix}$$

$A = U \Sigma V^T$:

$$\sigma_i = \boxed{\lambda_i(A^T A)}^{1/2}$$
$$= \boxed{\lambda_i(A A^T)}^{1/2}$$

$$A^T A v = \lambda v$$
$$A A^T \cdot A v = \lambda \cdot A v$$

U eigenvectors of AA^T

V " $A^T A$

(Symmetric) Positive Semidefinite Matrices (PSD)

$A \in \mathbb{R}^{d \times d}$, symmetric is positive semidefinite

iff $x^T A x \geq 0 \quad \forall x \in \mathbb{R}^d$

iff $x^T A x > 0$ positive definite

- PSD iff all eigenvalues are nonnegative
  $\lambda_i(A) \geq 0 \quad \forall i$

- PSD $\Rightarrow$ admit a decomposition
  $A = R^T R$

$$x^T R^T R x = \langle Rx, Rx \rangle = \|Rx\|_2^2 \geq 0$$

E.g. Cholesky Decomposition. R upper triangular.

- EVD $A = U \Lambda U^T = U \Sigma V^T$ $\quad \Lambda = \Sigma$, $V = U$

$$\sigma_i = \lambda_i(A^T A)^{1/2}$$
$$= \lambda_i(A^2)^{1/2} = \left((\lambda_i(A))^2\right)^{1/2} = \lambda_i(A)$$

- Examples: $\quad A = X^T X$

  Covariance matrices

  Kernel matrices

  Hessian Matrices depending on function.

Trace: sum of diagonal entries of a square matrix

$$\mathrm{Tr}[A] = \sum_{i=1}^{d} A_{ii} \longrightarrow \mathrm{Tr}[A+B] = \mathrm{Tr}[A] + \mathrm{Tr}[B]$$
$$\longrightarrow \mathrm{Tr}[\alpha A] = \alpha \cdot A$$

$$\mathrm{Tr}[A] = \sum_{i=1}^{d} \lambda_i(A) \qquad \boxed{\mathrm{Tr}\, A \cdot B = \mathrm{Tr}\, B \cdot A} = \mathrm{Tr}\, B^T A^T$$

symmetric A: $\quad A = U \Lambda U^T = \sum_{i=1}^{d} \lambda_i u_i u_i^T$

$$\mathrm{Tr}[a a^T] = \mathrm{Tr}[a^T a]$$
$$= \mathrm{Tr}[\langle a, a \rangle]$$

$$\mathrm{Tr}\, A = \sum_{i=1}^{d} \lambda_i \cdot \mathrm{Tr}[u_i u_i^T] \longleftarrow$$
$$= \sum_{i} \lambda_i \underbrace{\mathrm{Tr}[u_i^T u_i]}_{1}$$
$$= \sum_{i=1}^{d} \lambda_i$$

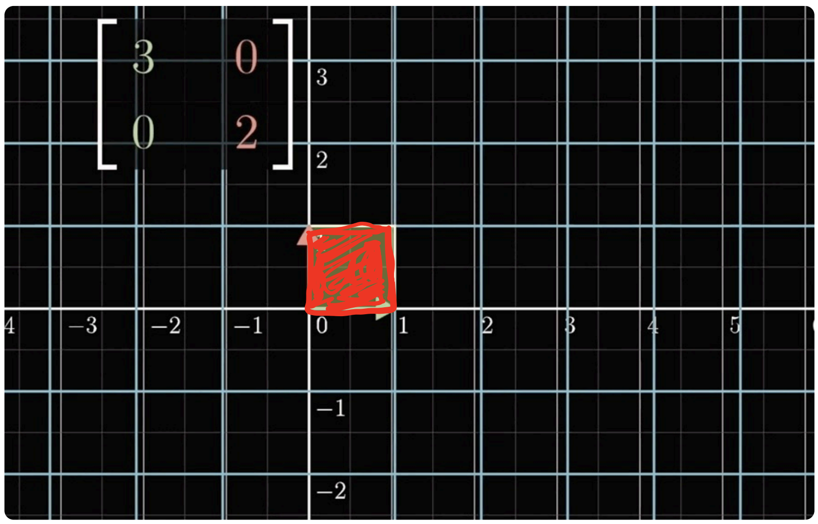
Determinant: $\det(A)$

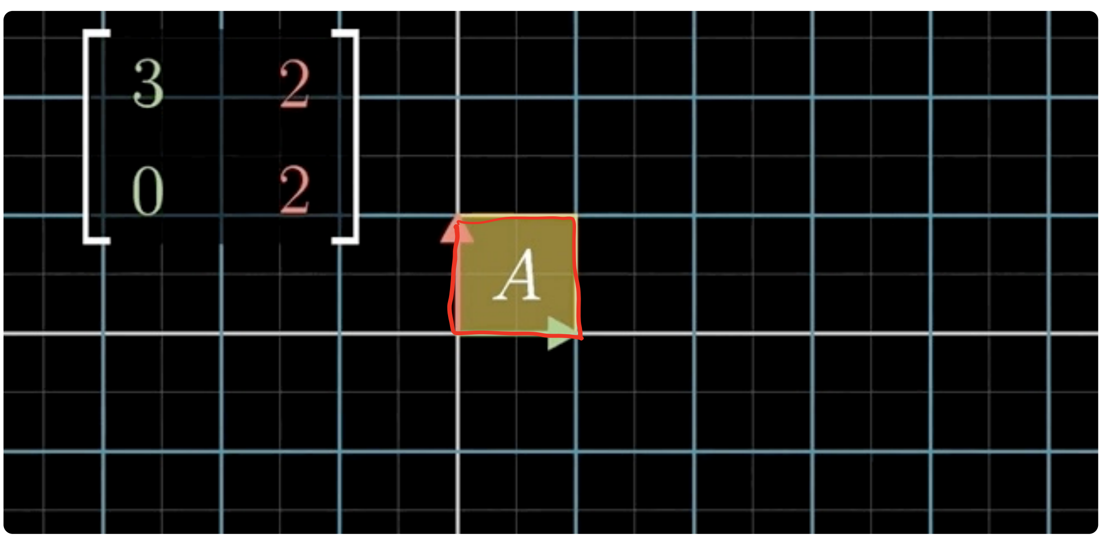
e.g. $\quad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow \det A = ad - bc$

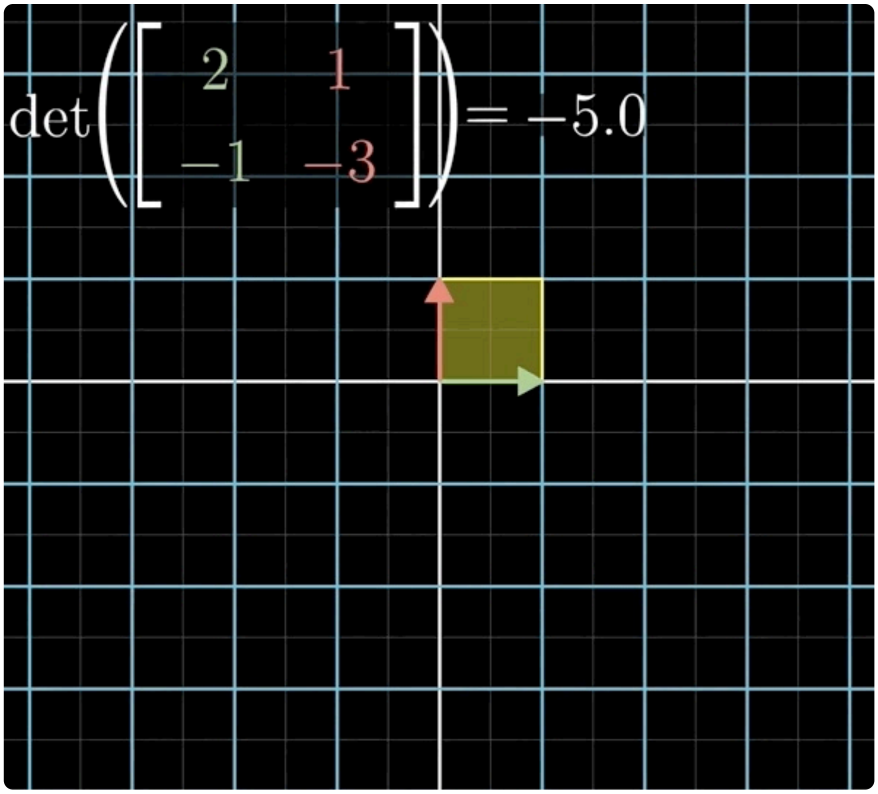
$\begin{bmatrix} c & d \end{bmatrix}$

$d \times d \Rightarrow$ Leibniz / Laplace formula

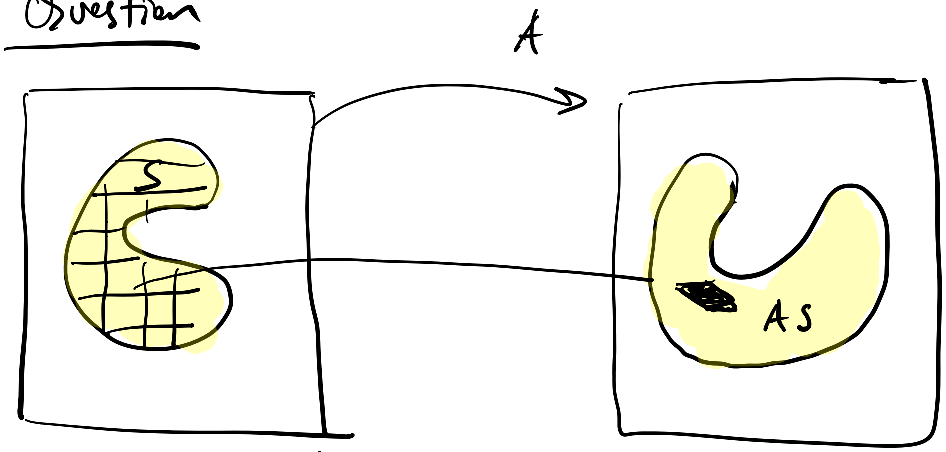
$$\det A = \prod_{i=1}^{d} \lambda_i(A)$$







Question



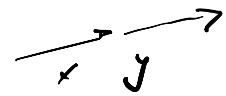
$$\mathrm{Vol}(AS) = |\det A| \, \mathrm{Vol}\, S$$

Cauchy-Schwarz Inequality

$$|\langle x, y \rangle| \le \|x\|_2 \cdot \|y\|_2 \quad \forall x, y \in \mathbb{R}$$

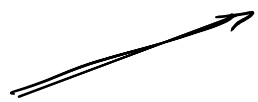
with equality iff x and y are collinear.

Another way



$$\iff \left| \left\langle \frac{x}{\|x\|_2}, \frac{y}{\|y\|_2} \right\rangle \right| \le 1 \iff |\langle u, v\rangle| \le 1 \quad \forall \|u\|_2 = 1, \ \forall \|v\|_2 = 1$$

$$\iff \max_{\|u\|_2 = 1} |\langle u, v\rangle| \le 1 \quad \forall \|v\|_2 = 1$$

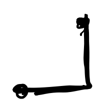


p-Norms

$$\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \longrightarrow \|\underline{x}\|_p = \left(\sum_{i=1}^{d} |x_i|^p \right)^{1/p} \qquad (p \ge 1)$$

$$\|x\|_2 = \left(\sum_i x_i^2 \right)^{1/2}$$

$$\|x\|_1 = \sum_i |x_i|$$

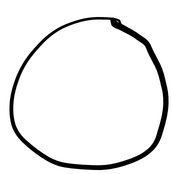


$$\|x\|_\infty = \max_i |x_i|$$

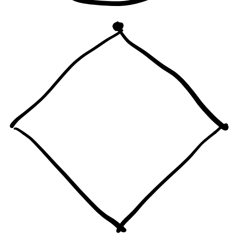
$$\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}$$

$$\left(\sum x_i^p \right)^{1/p} \approx \left(5^p \right)^{1/p} \approx 5$$

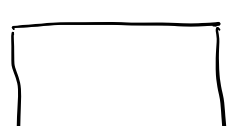
$\|x\|_2 = 1$

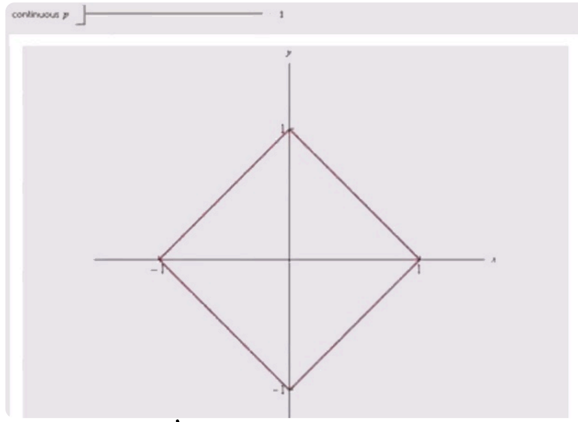


$\|x\|_1 = 1$



$\|x\|_\infty = 1$





$$\|x\|_0 = \sum_{i=1}^{d} 1\cdot\{x_i \neq 0\} \qquad \text{Cardinality}$$

$$\Downarrow$$

$$\begin{cases} 1 & \text{if } x_i \neq 0 \\ 0 & \text{o.w.} \end{cases}$$

$$x = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad \|x\|_0 = 2, \quad \|x\|_\infty = 1, \quad \|x\|_2 = \sqrt{2}$$

Hölder's Inequality: Let $p, q \geq 1$ and be conjugate, i.e., $\frac{1}{p} + \frac{1}{q} = 1$

e.g. $p=1,\ q=\infty$
$p=2,\ q=2$

$$|\langle \underline{x}, \underline{y} \rangle| \leq \|\underline{x}\|_p \cdot \|\underline{y}\|_q$$

With equality when $\underline{x}^p$ and $\underline{y}^q$ are linearly dependent

$\underline{\underline{A}}\ \underline{x} = \underline{y}$

Unconstrained Optimization

An unconst. opt problem is of the form

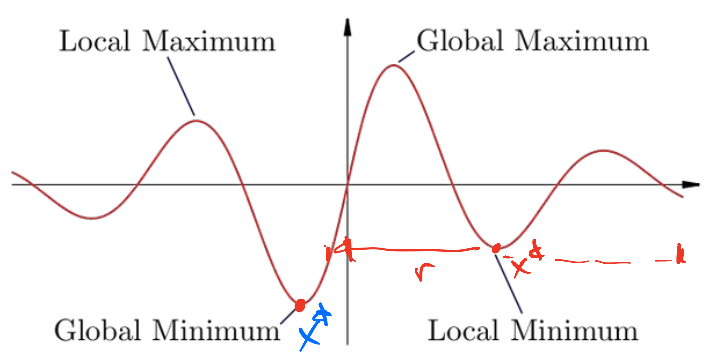
$$\min_{\underline{x} \in \mathbb{R}^d} f(\underline{x})$$

$f: \mathbb{R}^d \to \mathbb{R}$ objective function

→ A point $x^* \in \mathbb{R}^d$ is called a <u>local minimizer</u>

if $\exists r > 0$ s.t $f(x^*) \leq f(x)$ $\forall x$ s.t $\|x - x^*\| \leq r$

→ A <u>global minimizer</u> if $f(x^*) \leq f(x)$ $\forall x$



<u>First and second Order Necessary Conditions</u>

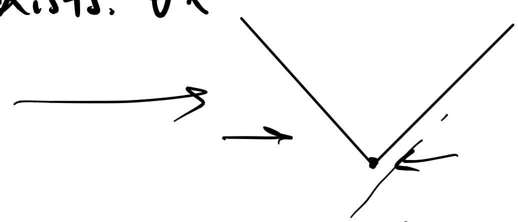
Given $f: \mathbb{R}^d \to \mathbb{R}$

Gradient and Hessian at $\underline{x} = \begin{bmatrix} \vdots \\ x_d \end{bmatrix}$ are defined as:

$$\nabla f(\underline{x}) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_d} \end{bmatrix} \qquad \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \cdots & \\ \vdots & \cdots & \end{bmatrix}$$

→ We say that f is differentiable if $\nabla f(x)$ exists. $\forall x$

$\boxed{f(x) = |x|}$



~~not~~ diff'ble at 0.

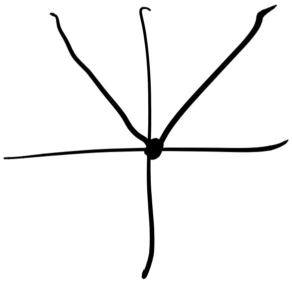
→ We say f is twice diff'ble if $\nabla^2 f(x)$ exists $\forall x$.

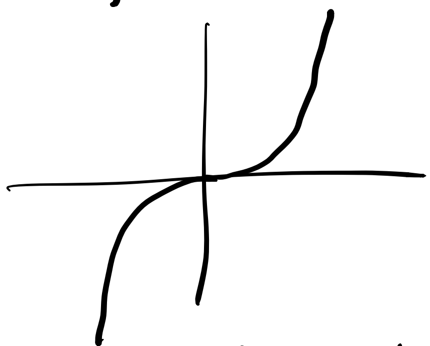
→ f is twice continuously diff'ble if it is twice diff'ble and second derivatives are continuous

- If f is twice cont. diff'ble then $\underline{\nabla^2 f(x)}$ is symmetric.

$$\frac{\partial^2 f(x)}{\quad} = \frac{\partial^2 f(x)}{\quad}$$

$\partial x_j \partial x_i$ $\qquad$ $\partial x_i \partial x_j$





not twice diff'ble at origin.

Property 1: If f is diff'ble and x^* is a local minimizer (of f), then $\nabla f(x^*) = 0$.