# Quadratic Form

$A \in \mathbb{R}^{d \times d}$, $x \in \mathbb{R}^{d}$

$x^T A x$ _ Quadratic

We assume A is symmetric

$x^T B x = x^T A x$ where B is symmetric

$$B = \frac{1}{2}(A + A^T)$$

## Definitiveness

$x^T A x > 0 \ \forall x \neq 0$ - A is positive definite

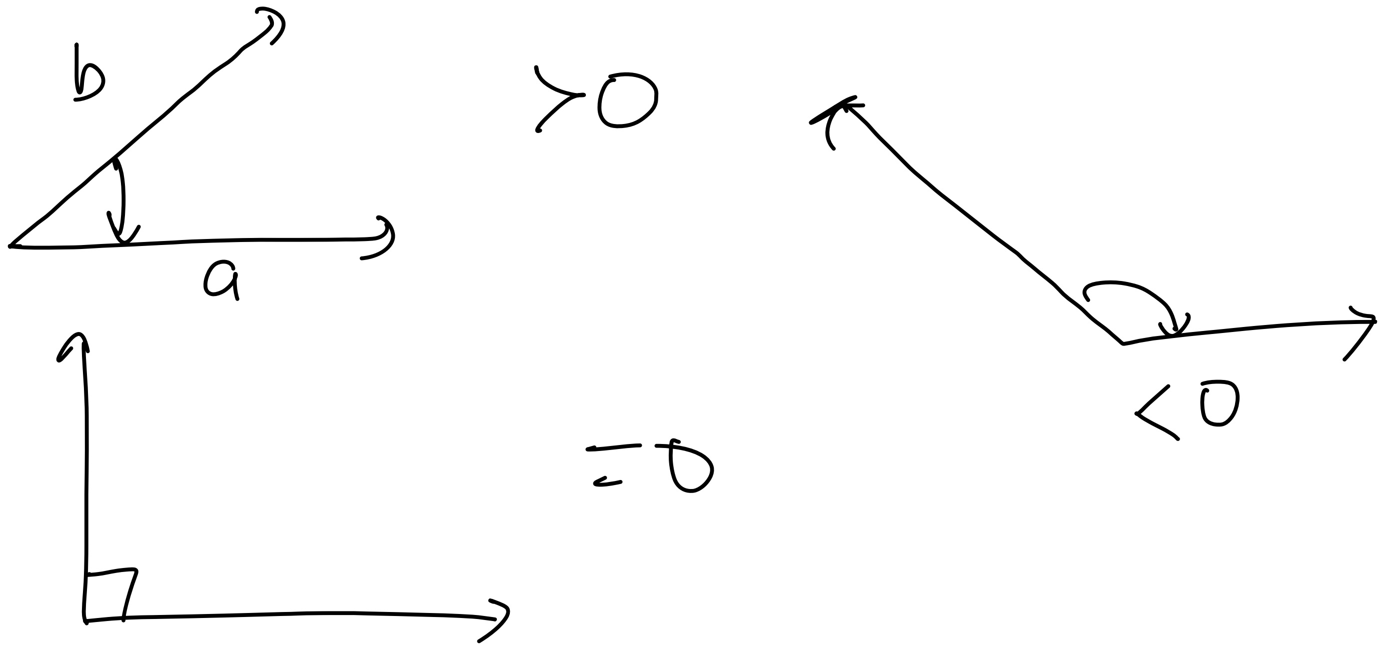
$x^T A x \geq 0 \ \forall x \neq 0$ - A is P.S.D.

$x^T A x < 0$ " A is N.D.

$\leq 0$ " A is NSD.

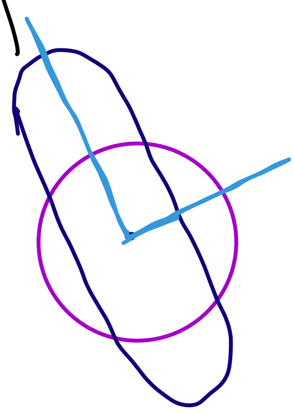
$> 0, < 0$ " A is indefinite

$a^Tb$




$x^T(Ax)$ → Dot product of input & output.

if all eigen values positive, then $x^TAx > 0$, think of eigen vector as pivot



# DECOMPOSITION

| | A | Decom. |
| --- | --- | --- |
| SVD | ANY | $A = USV^T$<br>$\begin{bmatrix} \vert & \vert & \vert \end{bmatrix}\begin{bmatrix} \ddots & 0 \\ 0 & \ddots \end{bmatrix}\begin{bmatrix} - \\ - \\ - \end{bmatrix}$ |
| EVD | SQUARE | $A = UDU^{-1}$<br>$\begin{bmatrix} \vert & \vert & \vert \end{bmatrix}\begin{bmatrix} \ddots & 0 \\ 0 & \ddots \end{bmatrix} u^{-1}$ |

$$A(X) = U(S(V^T(X)))$$

$$A(X) = U(D(U^{-1}(X)))$$

$U, V \rightarrow$ orthonormal

$S, D \rightarrow$ diagonal

| A | Step 1 | Step 2 | Step 3 |
| --- | --- | --- | --- |
| | | Diagonal | Rotation |
| SVD | $V^T$ | Scaling along axes | $U$ |
| | Rotation | | |
| Eig | $U^{-1}$ | Scaling along axis (If eigen values complex, rotation also involved) | $U$ |

For square symmetric → SVD = EVD

# Matrix Calculus

| function | eg | Value | 1st, | 2nd, |
|---|---|---|---|---|
| $f: \mathbb{R} \to \mathbb{R}$ | $x^2$ | $\mathbb{R}$ | $\mathbb{R}$ | $\mathbb{R}$ |
| $f: \mathbb{R}^d \to \mathbb{R}$ | Loss function | $\mathbb{R}$ | $\mathbb{R}^d$ [Gradient] | $\mathbb{R}^{d \times d}$ [Hessian] $S^d$ |
| $f: \mathbb{R}^d \to \mathbb{R}^p$ | NN-Layer | $\mathbb{R}^p$ | $\mathbb{R}^{d \times p}$ [Jacobian] | $\mathbb{R}^{d \times p \times p}$ [Tensor] |

If Hessian is P.D / P.S.D. $\to$ (strictly) convex

If Hessian is N.D./N.S.D $\to$ (strictly) concave

If Hessian is Indefinite $\to$ Saddle

(Along 1 plane convex
Along other plane concave)

$$\nabla_x f(x) = \nabla_x f(x_1, x_2, \ldots x_d) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_d}(x) \end{bmatrix}$$

Gradient points to direction of steepest ascent

$$\nabla_A f(A) = \begin{bmatrix} \frac{\partial f(A)}{\partial a_{11}} & \cdots & \\ & & \frac{\partial f(A)}{\partial a_{mn}} \end{bmatrix}$$

$$\nabla_x^2 f(x) \qquad f : \mathbb{R}^d \to \mathbb{R}$$

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

(Q) $\nabla_x b^T x$

$\rightarrow x \in \mathbb{R}^d \quad b^T x \in \mathbb{R}$

$$\nabla_x b^T x \in \mathbb{R}^d = \begin{bmatrix} \frac{\partial}{\partial x_1} b^T x \\ \vdots \\ \vdots \\ \frac{\partial}{\partial x_n} b^T x \end{bmatrix}$$

$$= \begin{bmatrix} \vdots \\ \frac{\partial}{\partial x_i} b_1 x_1 + \cdots b_i x_i + \cdots b_d x_d \\ \vdots \end{bmatrix}$$

$$= \begin{bmatrix} \vdots \\ b_i \\ \vdots \end{bmatrix} = \underline{\underline{b}}$$

$$\nabla_x x^T A x = \nabla_x x^T A x + \nabla_x x^T A x$$
$$= Ax + x^T A$$
$$= (A + A^T) x$$
If $A = A^T$, $2Ax$

$$\nabla_x \log|A| = A^{-1}$$

For ABC such that it is square
$$tr(ABC) = tr(BCA) = tr(CAB)$$

Norm

$$f(x+y) \leq f(x) + f(y)$$

L2 norm $\|x\|_2 = \sqrt{\sum_{x=1}^{x=n} x_i^2}$

l1 norm $\|x\|_1 = \sum_{x=1}^{n} |x_i|$

$l_\infty$ norm $||x||_\infty = \max_{x_i} |x_i|$

$$||x||_p = \left(\sum_{i=1}^{i=p} |x_i|^p\right)^{1/p}$$

Frobenius Norm

$$||A||_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2} = \sqrt{tr(A^TA)}$$

★ $||Ux||_2 = ||x||_2$ where $x \in \mathbb{R}^n$

$U \in \mathbb{R}^{n \times n}$ orthogonal

Proj of vector $y \in \mathbb{R}^m$ onto span of $\{x_1 \ldots x_n\}$ is vector $v \in \{x_1, \cdots x_n\}$ span such that $v$ is closest to $y$ as measured by Euclidean norm $||v-y||_2$

$$Proj(y; \{x_1 \ldots x_n\}) = \arg\min_{v \in span(\{x_1, x_2 .. x_n\})} ||y-v||^2$$

$$x^T A x = (x^T A x)^T = x^T A^T x \quad \left(\text{Since scalar}\right)$$

$$2x^T A x = x^T A x + x^T A^T x$$

$$x^T A x = \frac{1}{2} x^T (A + A^T) x$$

Hence by default, we assume them to be symmetric

P.D & N.D. are always full rank, hence invertible

Given $A \in \mathbb{R}^{m \times n}$ $\quad G = A^T A$ (Gram Matrix)

$\downarrow$

always P.S.D.

if $m \geq n$ & $A$ is full rank

$G = A^T A$ is P.D.

$$U = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}$$ orthonormal matrix

$$\Lambda = \mathrm{diag}(\lambda_1 \ldots \lambda_n)$$

$$AU = \begin{bmatrix} | & | & & | \\ Au_1 & Au_2 & \cdots & Au_n \\ | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} | & | & & | \\ \lambda_1 u_1 & \lambda_2 u_2 & \cdots & \lambda_n u_n \\ | & | & & | \end{bmatrix}$$

$$\begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} = U\Lambda$$

$$A = AUU^T = U\Lambda U^T$$

Let $x$ be a vector, $\hat{x}$ is rep. in basis $U$

$$\hat{z} = U^T z = U^T A x = U^T U \Lambda U^T x$$
$$= \Lambda \hat{x}$$

We can see left multiplying by $A$ in original space $\Rightarrow$ $\Lambda$ multiplication in new space

$$q = AAAx$$

$$\hat{q} = U^T q = U^T A A A x$$

$$= U^T U \Lambda U^T U \Lambda U^T U \Lambda U^T x$$

$$= \Lambda^3 \hat{x}$$

Diagonalizing Quadratic form

$$x^T A x = x^T U \Lambda U^T x$$

$$= \hat{x}^T \Lambda \hat{x}$$

$$= \sum_{i=1}^{n} \lambda_i \hat{x}_i^2 \qquad \text{n term}$$

$$x^T A x = \sum_{i=1, j=1}^{n} x_i x_j A_{ij} \qquad n^2 \text{ term}$$

$$\nabla_x b^T x = b \qquad \nabla_x x^T A x = 2Ax \ \text{(sym)}$$

$$\nabla_x^2 b^T x = 0 \qquad \nabla_x^2 x^T A x = 2A \ \text{(sym)}$$

# LEAST SQUARES

$A \in \mathbb{R}^{m \times n} \rightarrow$ full rank $\quad b \in \mathbb{R}^m \quad b \notin \mathcal{R}(A)$

Find matrix $x$ $\quad Ax$ is as close to $b$,

$$\|Ax - b\|_2^2$$

$$\|x\|_2^2 = x^T x$$

$$\|Ax - b\|_2^2 = (Ax - b)^T (Ax - b)$$

$$= (x^T A^T - b^T)(Ax - b)$$

$$= x^T A^T A x - x^T A^T b - b^T A x + b^T b$$

$$= x^T A^T A x - 2 b^T A x + b^T b$$

$$\nabla_x (\qquad) = 2A^T A x - 2A^T b$$

$$A^T A x = A^T b$$

$$x = (A^T A)^{-1} A^T b$$

Eigen values - optimization

$$\max_{x \in \mathbb{R}^n} (x^T A x) \qquad \|x_2\|^2 = 1$$

$$\mathcal{L}(x, \lambda) = x^T A x - \lambda (x^T x - 1)$$

$$\nabla_x (\mathcal{L}(x, \lambda)) = 2Ax - 2\lambda x = 0$$

$$\boxed{Ax = \lambda x}$$

Those points which max/min $x^T A x$ are eigen values/eigen vectors.

# PROBABILITY

$A \perp B \rightarrow$ independent

Random variable mapping of outcome to real value

$\Omega \rightarrow$ sample space

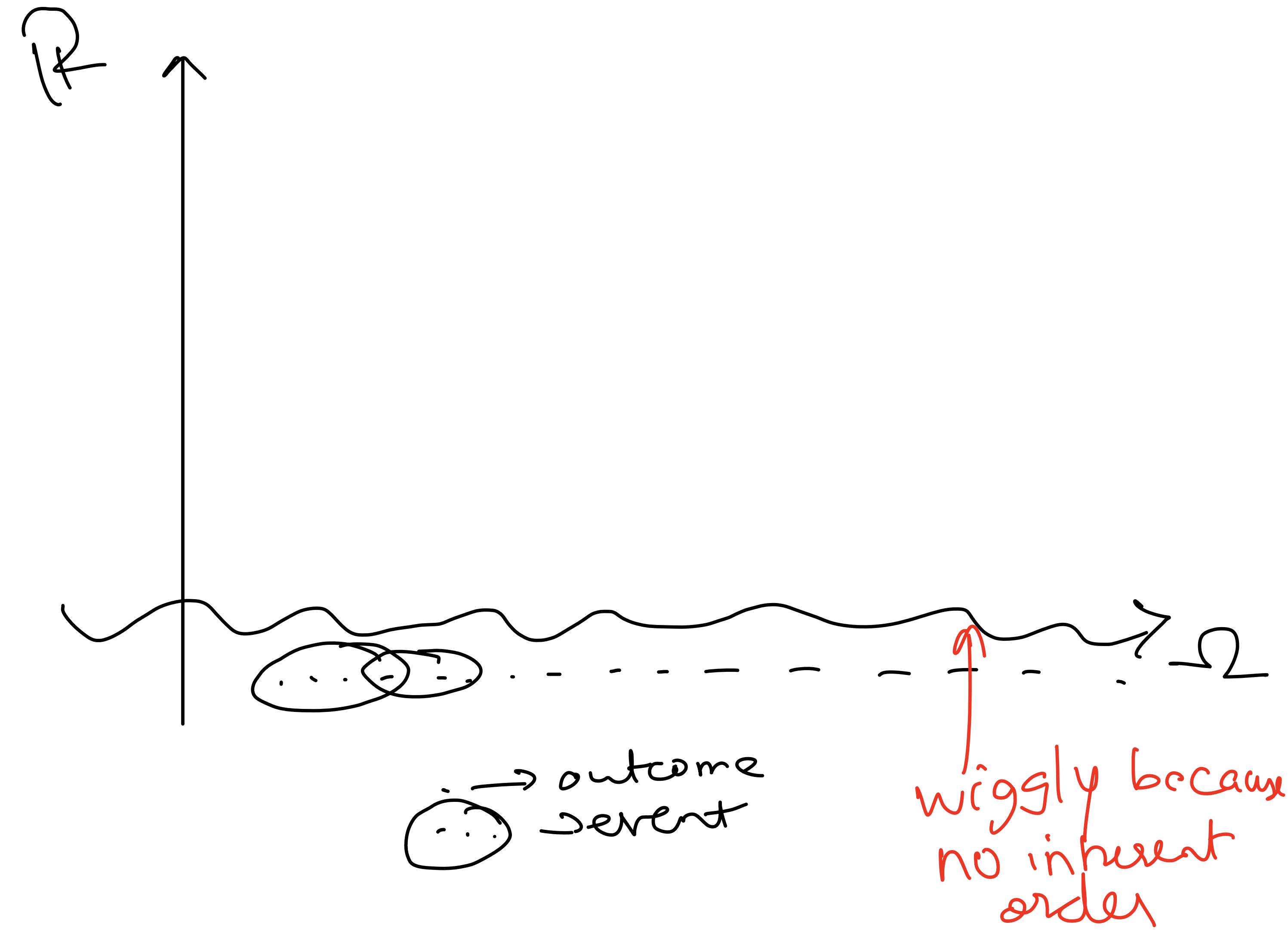
RV $X : \Omega \rightarrow \mathbb{R}$

\# of heads: $X(\omega_0) = 5$

set of outcomes $\rightarrow$ event

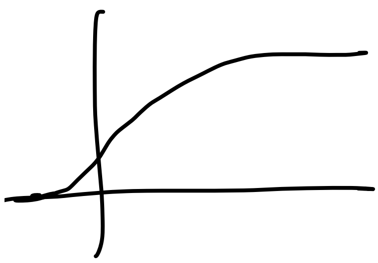
Roll a dice $\rightarrow$ 6 possible outcomes

Event $\rightarrow$ S.O. outcomes to which P assigned




C.D.F.

$$F_X(x) = P(X < x)$$



$$P[\{\omega : X(\omega) < t]$$

$\omega \rightarrow$ event

$\omega$: HHHTT

$X(\omega)$ = no. of heads

bet$^n$ 0 & 1 $\leftarrow \infty$

$\downarrow -\infty$

U

# Discrete vs. Continuous RV

**Discrete RV**: $Val(X)$ countable

$P(X = k) := P(\{\omega | X(\omega) = k\})$

Probability Mass Function (PMF)

$p_X : Val(X) \to [0, 1]$

$$p_X(x) := P(X = x)$$

$$\sum_{x \in Val(X)} p_X(x) = 1$$

**Continuous RV**: $Val(X)$ uncountable

$P(a \leq X \leq b) := P(\{\omega | a \leq X(\omega) \leq b\})$

Probability Density Function (PDF)

$f_X : \mathbb{R} \to \mathbb{R}$

$$f_X(x) := \frac{d}{dx} F_X(x) \leftarrow \text{CDF}$$

$$f_X(x) \neq P(X = x)$$

$$\int_{-\infty}^{\infty} f_X(x) dx$$

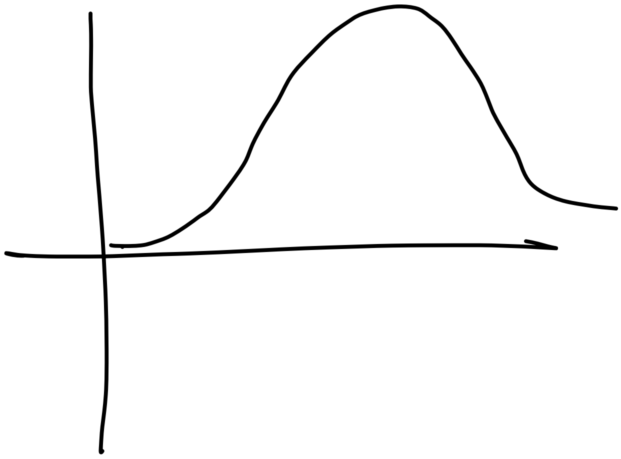
CDF:

$P(a \leq \cancel{x} \leq b)$

$= F_X(b) - F_X(a)$

$= \int_a^b f_X(x)\, dx$

$\uparrow$ PDF



Val (X) → set of all possible values

# Expected Value & Variance

$g: \mathbb{R} \to \mathbb{R}$

Let $X$ be discrete RV with PMF $p_X$

$$E(g(X)) = \sum g(x) p_X(x) \quad \forall x \text{ in } Val(X)$$

Continuous

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

Monte Carlo Estimate

$$\lim_{N \to \infty} \boxed{\frac{1}{N} \sum_{i=1}^{N} g(x^{(i)})} \to \int_{-\infty}^{\infty} g(x) p(x)\, dx$$

Law of Large Numbers

$$\mathrm{Var}(X) = E\left[(X - E[X])^2\right]$$

$$= E\left[X^2 + E^2[X] - 2XE[X]\right]$$

$$= E[X^2] + E^2[X] - 2E^2[X]$$

$$= E[X^2] - E^2[X]$$

parameter → no. which tries to summarize shape of the distribution

GAUSSIAN