Recursive Least Squares Motivation for Kalman Filter

ROB 501

Necmiye Ozay

- Recursive Least Squares
- Sneak peek: Kalman Filter
- ~~(if time)~~ Intro probability

Rob 501 Handout: Grizzle
Recursive Least Squares

Model:

$$y_i = C_i x + e_i, \;\; i = 1, 2, 3, \cdots$$

$$\begin{aligned} C_i &\in \mathbb{R}^{m \times n} \\ i &= \text{time index} \\ x &= \text{an unknown } \underline{\text{constant}} \text{ vector} \in \mathbb{R}^n \\ y_i &= \text{measurements} \in \mathbb{R}^m \\ e_i &= \text{model "mismatch"} \in \mathbb{R}^m \end{aligned}$$

Objective 1: Compute a least squared error estimate of x at time k, using all available data at time k, $(y_1, \cdots, y_k)$!

Objective 2: Discover a computationally attractive form for the answer.

Solution:

$$\begin{aligned} \hat{x}_k := \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} & \left(\sum_{i=1}^{k} (y_i - C_i x)^\top S_i (y_i - C_i x) \right) \\ = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} & \left(\sum_{i=1}^{k} e_i^\top S_i e_i \right) \end{aligned}$$

where $S_i = m \times m$ positive definite matrix. ($S_i > 0$ for all time index i)

Batch Solution:

$$Y_k = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}, A_k = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_k \end{bmatrix}, E_k = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_k \end{bmatrix}$$

$$R_k = \begin{bmatrix} S_1 & & & \mathbf{0} \\ & S_2 & & \\ & & \ddots & \\ \mathbf{0} & & & S_k \end{bmatrix} = diag(S_1, S_2, \cdots, S_k) > 0$$

$Y_k = A_k x + E_k$, [model for $1 \leq i \leq k$]
$\|Y_k - A_k x\|^2 = \|E_k\|^2 := E_k^\top R_k E_k$

Since $\hat{x}_k$ is the value minimizing the error $\|E_k\|$, which is the unexplained part of the model,

$$\hat{x}_k = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \|E_k\| = \underbrace{\underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \|Y_k - A_k x\|}_{},$$

which satisfies the Normal Equations $(A_k^\top R_k A_k)\hat{x}_k = A_k^\top R_k Y_k$.

$\therefore$ $\underline{\hat{x}_k = (A_k^\top R_k A_k)^{-1} A_k^\top R_k Y_k}$, which is called a <u>Batch Solution</u>.

Drawback: $A_k = km \times n$ matrix, and grows at each step!

Solution: Find a recursive means to compute $\hat{x}_{k+1}$ in terms of $\hat{x}_k$ and the new measurement y_{k+1}!

Normal equations at time k, $(\underbrace{A_k^\top R_k A_k}_{})\hat{x}_k = \underbrace{A_k^\top R_k Y_k}_{}$, is equivalent to

$$\underbrace{\left(\sum_{i=1}^{k} C_i^\top S_i C_i \right)}_{M_k} \underbrace{\hat{x}_k}_{\hat{x}_k} = \underbrace{\sum_{i=1}^{k} C_i^\top S_i y_i}_{*}.$$

We define

$$M_k = \underbrace{\sum_{i=1}^{k} C_i^\top S_i C_i}_{}$$

so that

$$\underline{M_{k+1}} = \underline{M_k} + \underbrace{C_{k+1}^\top S_{k+1} C_{k+1}}_{}.$$

At time $k+1$, Normal equation

$$\underbrace{(\sum_{i=1}^{k+1} C_i^\top S_i C_i)}_{M_{k+1}}\,\hat{x}_{k+1} = \sum_{i=1}^{k+1} C_i^\top S_i y_i$$

$$\left(\sum_{i=1}^{k} C_i^\top S_i y_i\right) + C_{k+1}^\top S_{k+1} y_{k+1}$$

or

$$M_{k+1}\hat{x}_{k+1} = \underbrace{\sum_{i=1}^{k} C_i^\top S_i y_i}_{M_k\hat{x}_k} + C_{k+1}^\top S_{k+1} y_{k+1}.$$

$\therefore M_{k+1}\hat{x}_{k+1} = M_k\hat{x}_k + C_{k+1}^\top S_{k+1} y_{k+1}$

Good start on recursion! Estimate at time $k+1$ expressed as a linear combination of the estimate at time k and the latest measurement at time $k+1$.

Continuing,

$$\hat{x}_{k+1} = M_{k+1}^{-1}\left[M_k\hat{x}_k + C_{k+1}^\top S_{k+1} y_{k+1}\right].$$

Because

$$M_k = M_{k+1} - C_{k+1}^\top S_{k+1} C_{k+1},$$

we have

$$\hat{x}_{k+1} = \hat{x}_k + \underbrace{M_{k+1}^{-1} C_{k+1}^\top S_{k+1}}_{\text{Kalman gain}}\underbrace{(y_{k+1} - C_{k+1}\hat{x}_k)}_{\text{Innovations}}.$$

Innovations $y_{k+1} - C_{k+1}\hat{x}_k$ = measurement at time $k+1$ minus the "predicted" value of the measurement = "new information".

In a real-time implementation, computing the inverse of M_{k+1} can be time consuming. An attractive alternative can be obtained by applying the Matrix Inversion Lemma:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B\left(DA^{-1}B + C^{-1}\right)^{-1} DA^{-1}$$

Now, following the substitution rule as shown below,

$$A \leftrightarrow M_k \quad B \leftrightarrow C_{k+1}^\top \quad C \leftrightarrow S_{k+1} \quad D \leftrightarrow C_{k+1},$$

we can obtain that

$$\begin{aligned} M_{k+1}^{-1} &= \left(M_k + C_k^\top S_{k+1} C_{k+1}\right)^{-1} \\ &= M_k^{-1} - M_k^{-1} C_{k+1}^\top \left[C_{k+1} M_k^{-1} C_{k+1}^\top + S_{k+1}^{-1}\right]^{-1} C_{k+1} M_k^{-1}, \end{aligned}$$

which is a recursion for M_k^{-1}!

Upon defining

$$P_k = M_k^{-1},$$

we have

$$P_{k+1} = P_k - P_k C_{k+1}^\top \left[C_{k+1} P_k C_{k+1}^\top + S_{k+1}^{-1}\right]^{-1} C_{k+1} P_k$$

We note that we are now inverting a matrix that is $m \times m$, instead of one that is $n \times n$. Typically, $n > m$, sometimes by a lot!

In Kalman filter, we assume the state 'x' is evolving in time according to some linear dynamics equation:

$x_{k+1} = A_k x_k + G_k w_k$ → process noise

$y_k = C_k x_k + v_k$ → measurement noise

measurement equation

w_k, v_k are zero mean Gaussian random variables

RLS is a special case:

$G_k = 0 \quad A_k = I$

$x_{k+1} = x_k$

(constant state)
(so, we can drop k in x_k)

$y_k = C_k x + v_k$

$v_k = e_k$

assumed to be deterministic

The Kalman Filter

Definition of Terms:

estimate of $\hat{x}_k$ at time k

$$\widehat{x}_{k|k} := \mathcal{E}\{x_k | y_0, \cdots, y_k\}$$

$$P_{k|k} := \mathcal{E}\{(x_k - \widehat{x}_{k|k})(x_k - \widehat{x}_{k|k})^\top | y_0, \cdots, y_k\}$$

estimate of $\hat{x}_{k+1}$ at time k

$$\widehat{x}_{k+1|k} := \mathcal{E}\{x_{k+1} | y_0, \cdots, y_k\}$$

$$P_{k+1|k} := \mathcal{E}\{(x_{k+1} - \widehat{x}_{k+1|k})(x_{k+1} - \widehat{x}_{k+1|k})^\top | y_0, \cdots, y_k\}$$

Initial Conditions:

$$\widehat{x}_{0|-1} := \bar{x}_0 = \mathcal{E}\{x_0\}, \quad \text{and} \quad P_{0|-1} := P_0 = \text{cov}(x_0)$$

For $k \geq 0$

M_k^{-1} in RLS

Kalman gain

S_k in RLS

Measurement Update Step:

$$K_k = P_{k|k-1} C_k^\top \boxed{\left(C_k P_{k|k-1} C_k^\top + Q_k\right)^{-1}}$$

(Kalman Gain)

$$\widehat{x}_{k|k} = \widehat{x}_{k|k-1} + K_k \left(y_k - C_k \widehat{x}_{k|k-1}\right)$$

innovation term

$$P_{k|k} = P_{k|k-1} - K_k C_k P_{k|k-1}$$

Time Update or Prediction Step:

$$\widehat{x}_{k+1|k} = A_k \widehat{x}_{k|k}$$

$$P_{k+1|k} = A_k P_{k|k} A_k^\top + G_k R_k G_k^\top$$

End of For Loop (Just stated this way to emphasize the recursive nature of the filter)

First probability review

$f: \mathbb{R} \to \mathbb{R}$ is a probability density for a continuous (real-valued) random variable X if:

a) $\forall x \in X$, $f(x) \geq 0$ ✓

b) $\int_{-\infty}^{\infty} f(x)\,dx = 1$ ✓

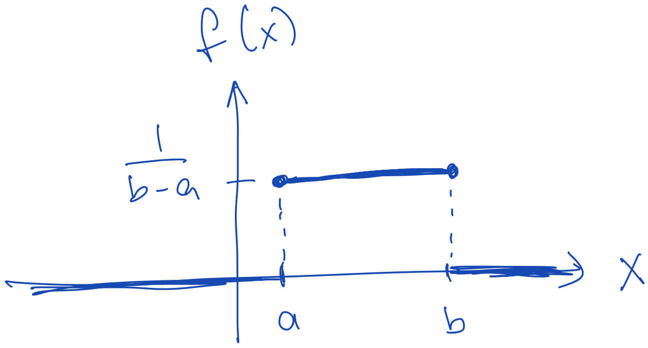
c) $\forall x_1 < x_2$, the probability X takes values between x_1 and x_2 is

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)\,dx$$

* Probability density functions (PDF) represent continuous random variables (R.V.)

Ex: ① Uniform R.V. over $[a,b]$, $a<b$:

$$U[a,b] \sim f(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a}, & a \leq x \leq b \\ 0 & x > b \end{cases}$$



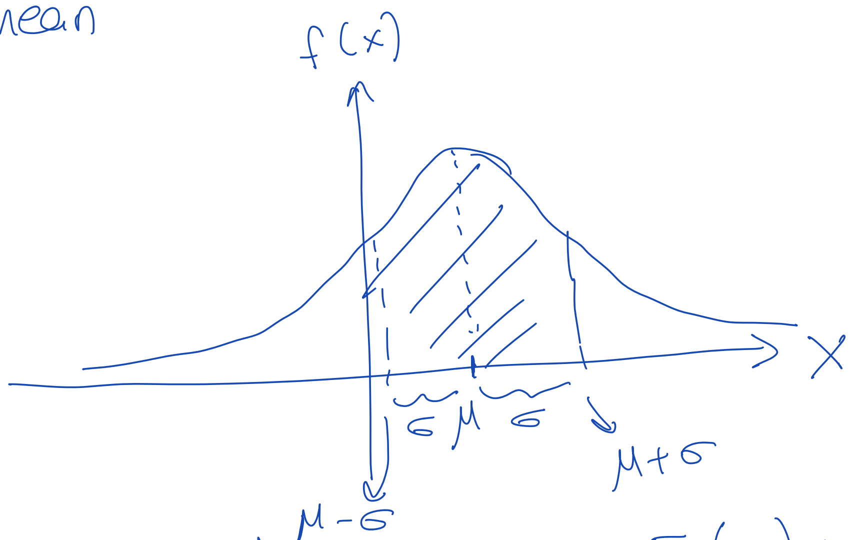
$$\int_{-\infty}^{\infty} f(x)\,dx = \int_a^b f(x)\,dx = \int_a^b \frac{1}{b-a}\,dx = \frac{b-a}{b-a} = 1$$

② Normal or Gaussian R.V.

$$\underbrace{N(\mu, \sigma^2)}_{} \sim f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]}$$

μ → mean

σ^2 → variance



<u>Mean value</u>: $m := E(x) := \int_{-\infty}^{\infty} x\, f(x)\,dx$

$E(x)$ → expected value

- also known as average value, expected value, or expectation

- In general, let $g: \mathbb{R} \to \mathbb{R}$ be a function,

$$\underbrace{E\{g(x)\}}_{\text{expected value}} := \int_{-\infty}^{\infty} g(x)\, f(x)\, dx$$

Ex: (1) $f(x) \sim U[a,b]$

$$m = E(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$= \int_{a}^{b} x \frac{1}{b-a}\, dx = \frac{b+a}{2}$$

(2) $f(x) \sim \mathcal{N}(\mu, \sigma^2)$

$$m = E(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \mu$$

Variance of R.V.

$$Var(X) = E\left((X - \underset{\text{mean}}{m})^2\right) = \int_{-\infty}^{\infty} (x-m)^2 f(x)\, dx$$

Ex: (1) $U[a, b]$, where $m = \frac{b+a}{2}$

$$Var(X) = E\left(\left(X - \frac{b+a}{2}\right)^2\right)$$

$$= \int_a^b \left(x - \frac{b+a}{2}\right)^2 \underbrace{\frac{1}{b-a}}_{f(x)}\, dx$$

$$= \frac{1}{12}(b-a)^2$$

(2) $N(\mu, \sigma^2)$

σ is the standard deviation

$$Var(X) = E\left((X-\mu)^2\right) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

$$= \underset{\text{variance}}{\sigma^2}$$

Random Vectors

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}, \text{ where } X_i \text{ is a R.V.}$$

Its density $f_X(x_1, x_2, \dots, x_n) \geq 0$

and $\underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n \text{ integrals}} f_X(x_1, x_2, \dots, x_n)\, dx_1 \dots dx_n = 1$

Mean Vector

$$\mathbf{m} := E(X) := \begin{bmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_n) \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix}$$

$$E(x_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i f_X(x_1, \dots, x_n)\, dx_1 dx_2 \dots dx_n$$

Covariance Matrix

$$\underbrace{Q}_{n \times n} := E\left(\underbrace{(x-\mu)}_{n \times 1} \underbrace{(x-\mu)^T}_{1 \times n} \right) = Q^T$$

$\Downarrow$

Q is symmetric

$$[Q]_{ij} = E\left((x_i - m_i)(x_j - m_j) \right)$$

Fact: $Q \geq 0$, but for most of our work, we will have $Q > 0$.

(norm of) Q is big $\Rightarrow$ lots of uncertainty

$\Rightarrow$ Q^{-1} is small, less information

Q^{-1} is called the information matrix

Ex: $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$

X_1 $\to$ high uncertainty sensor (MEMS IMU)

X_2 $\to$ low uncertainty sensor (fiberoptic IMU)

1) $Q = \begin{bmatrix} 10 & 0 \\ 0 & 0.1 \end{bmatrix} \Rightarrow Q^{-1} = \begin{bmatrix} \frac{1}{10} & 0 \\ 0 & 10 \end{bmatrix}$

2) $Q = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \theta^T \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \theta$

$\lambda_1 + \lambda_2 = 4$ $\quad \lambda_1 = 3$ $\quad \theta = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

$\lambda_1 \lambda_2 = 3$ $\quad \lambda_2 = 1$

Define a new random vector by

$Y = \theta \cdot X = \frac{1}{\sqrt{2}} \begin{bmatrix} X_1 + X_2 \\ X_1 - X_2 \end{bmatrix}$

Assume $m = E(X) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$E(Y) = \frac{1}{\sqrt{2}} \begin{bmatrix} E(X_1) + E(X_2) \\ E(X_1) - E(X_2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$Cov(Y) =$

OFFICE HOURS

$$M = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

$$\begin{bmatrix} A & B \\ 0 & C \end{bmatrix} \begin{bmatrix} 0 \\ v_2 \end{bmatrix} \begin{bmatrix} Bv_2 \\ Cv_2 \end{bmatrix}$$

M has e-value λ then M^T also has e-value λ.

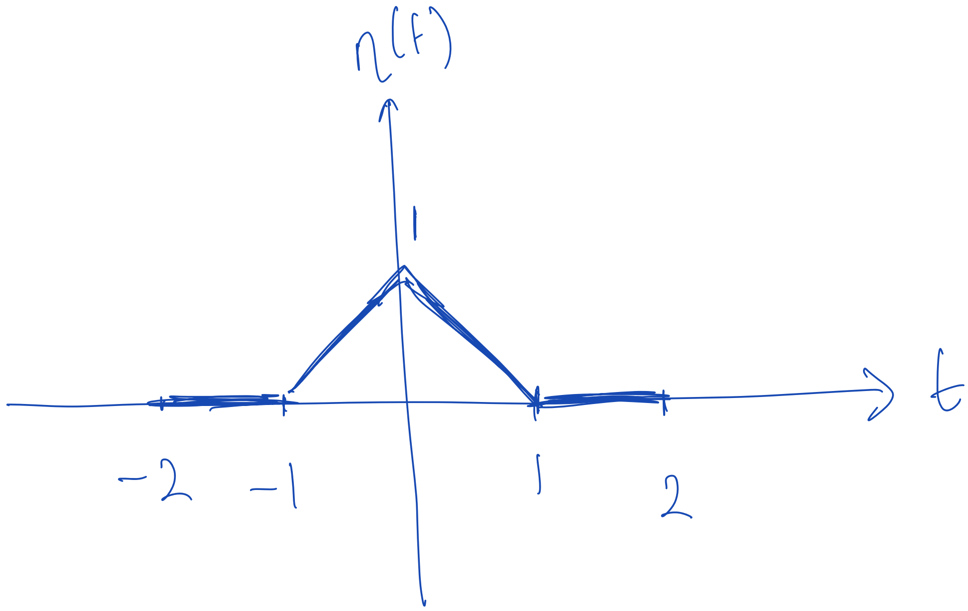
Let v' be an eigen-vector of A, with eigen-value λ'. Then $\exists$ an eigen-value of M that is λ'.

$v' \neq 0 \qquad Av' = \lambda' v'$

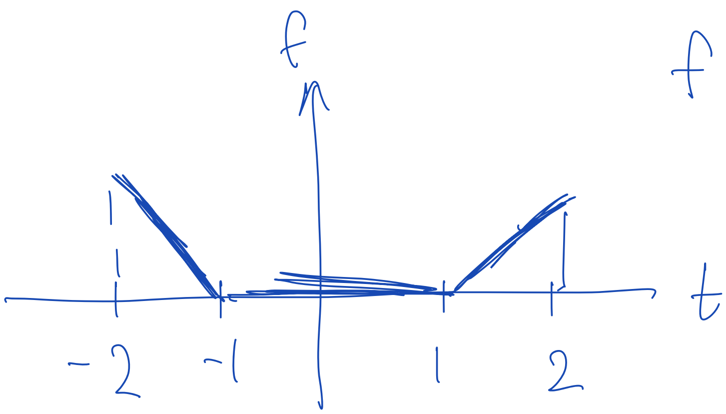
$$\underbrace{\overbrace{\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}}^{M} \overbrace{\begin{bmatrix} v' \\ 0 \end{bmatrix}}^{\bar{v}'}}_{M\bar{v}'} = \begin{bmatrix} Av' \\ 0 \end{bmatrix} = \begin{bmatrix} \lambda' v' \\ 0 \end{bmatrix}$$

$$= \lambda' \begin{bmatrix} v' \\ 0 \end{bmatrix}$$

$$M\bar{v}' = \lambda' \bar{v}'$$



$$\langle f, f \rangle > 0 \quad \forall f \neq 0$$



$f \neq 0$

$\langle f, f \rangle = 0$

$P(0)$ ✓

$\boxed{P(0)} \Rightarrow \boxed{P(1)}$ ✓

$P(1) \Rightarrow \boxed{P(2)}$

$\vdots$

$P(k) \Rightarrow P(k+1)$

$L : \mathbb{P}^3 \to \mathbb{P}^2$

$4 \text{ dim} \to 3 \text{ dim}$

$[L(p)]_{\mathbb{P}^2} = A [p]_{\mathbb{P}^3}$

$\in \mathbb{R}^3 \qquad \in \mathbb{R}^4$

$y = A x$, $y \in \mathbb{R}^3$, $x \in \mathbb{R}^4$

$A = \begin{bmatrix} & \end{bmatrix}_{3\times 4}$

$C_{k\times 3}\, x \to \in \mathbb{R}^3$

$x \in M \iff \begin{bmatrix} a & b & 0 \end{bmatrix} x = 0$

$\Updownarrow$

$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \iff \begin{cases} a x_1 + b x_2 = 0 \\ x_1 = \dfrac{-b}{a} x_2 \end{cases}$

$x = \begin{bmatrix} \frac{-b}{a} x_2 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{-b}{a} \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} x_3$

$x_2 = \dfrac{-a}{b} x_1$

$x = \begin{bmatrix} x_1 \\ -\frac{a}{b} x_1 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -a/b \\ 0 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} x_3$

$$+ \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix} x_4 \qquad \tilde{\alpha} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\tilde{\alpha} = 5x_4 + x_3$$

$$\mathcal{L} : \mathbb{P}^2 \longrightarrow \mathbb{P}^3$$

$p \in \mathbb{P}^2$

$\mathcal{L}(p) \in \mathbb{P}^3$

$$[\mathcal{L}(p)]_{\mathbb{P}^3} = A\,[p]_{\mathbb{P}^2}$$

$$[\mathcal{L}(u^1)]_{V_1} = A\,[u^1]_{U_1}$$

$$[\mathcal{L}(p)]_{V_1} = A\,[p]_{U_1}$$

$$\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = A_1$$

$U_1 = \{u^1, u^2, u^3\}$

$1, x, x^2$

$$[\mathcal{L}(u^i)]_{V_1} = A\,[u^i]_{U_1}$$

$= A_i$

$$\mathcal{L} : X \to X$$

$$\mathcal{L} : X \to Y$$

If V and W are subspaces

and $V \cap W = \{0\}$.

Then $x \in V \oplus W$

$\Rightarrow \exists$ a unique $x_v \in V$, $x_w \in W$ s.t.

$x = x_v + x_w$

If $V \cap W \neq \{0\} \Leftrightarrow \dim(V \cap W) \geq 1$,

then $x \in V + W$ can be written

as $x = \underbrace{\tilde{x}_v}_{\in V} + \underbrace{\tilde{x}_w}_{\in W}$ but $\tilde{x}_v$ and $\tilde{x}_w$ can have infinitely many values

$y \in V \cap W$

$$x = \underbrace{(\tilde{x}_V + y)}_{\in V} + (\tilde{x}_W - y) \in$$

Say $V \cap W = \{0\}$.

$\dim(V) = k_1$ $\quad\quad$ $\dim(V \oplus W) = k_1 + k_2$

$\dim(W) = k_2$

$V \oplus W = \text{span}\{V, W\}$

If I have V and I compute $V^\perp$

$\dim(V) = k_1 \implies \dim(V^\perp) = n - k_1$

$V \oplus V^\perp = X$

$\dim\left(N\left(\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\right)\right) = 2$

$\to N(A - \lambda_1 I)$

$\dim(N(A - 0 \cdot I)) =$

$$A_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \rightarrow \text{diagonalizable}$$
but $\text{rank}(A) = 0$

$\lambda_1 = \lambda_2 = 0$

$$A_2 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \rightarrow \text{not diagonalizable}$$
but $\text{rank}(A) = 2$

$\lambda_1 = \lambda_2 = 0$

any $x \in \mathbb{R}^2$ $x \neq 0$ is an eigevector of A_1 with eigenvalue 0.

$$A_1 \cdot x = 0 \cdot x$$

$\Rightarrow$ I can find v^1, v^2 that are e-vectors of A_1 that are linearly independent

$M = [v^1 \ v^2]$

$AM =$
$A[v_1 \ v_2] =$
$[Av_1 \ Av_2]$
$= [\lambda_1 v_1 \ \lambda_2 v_2]$
$$= \underbrace{[v_1 \ v_2]}_{M} \underbrace{\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}}_{S}$$

for A_2:

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 1 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$x_1 + 2x_2 = x_1 \rightarrow x_2 = 0$

$x_2 = x_2$

$AM = M\Sigma$

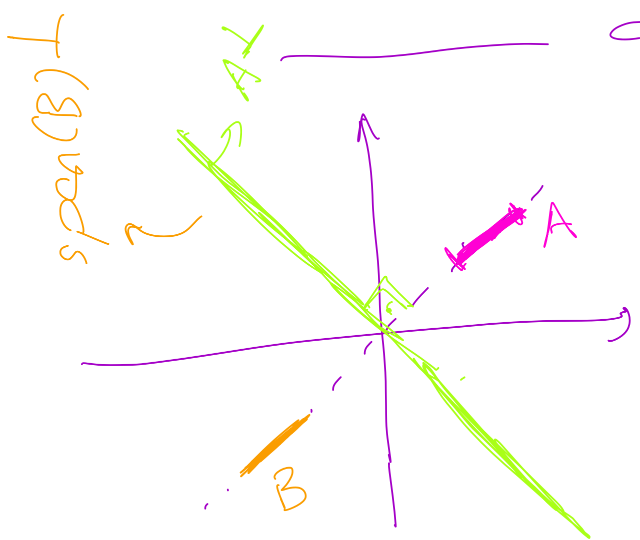
$\Sigma = M^{-1}AM$

eigenvectors of A_2 are of the form $\begin{bmatrix} x_1 \\ 0 \end{bmatrix}$ $\rightarrow$ all e-vectors are linearly dependent

There is no "M" that is invertible.

$$N(A_2 - 1I) = N\left(\begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}\right)$$

$$\dim\left(N\begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}\right) = 1$$



$A^\perp = (\text{span}(A))^\perp$

my A and B satisfy span$(B)^\perp = A^\perp$

$$A^\perp = \mathrm{span}(B)^\perp \iff \mathrm{span}(A) = \mathrm{span}(B)$$

$$\Updownarrow$$

$$(A^\perp)^\perp = \underbrace{\left(\mathrm{span}(B)^\perp\right)^\perp}_{\mathrm{span}(B)}$$

$$\| \qquad$$

$$\left(\mathrm{span}(A)^\perp\right)^\perp$$

$$\|$$

$$\mathrm{span}(A) = \mathrm{span}(A)$$

$$A = B \implies A^\perp = \mathrm{span}(B)^\perp$$

$$e_i^t P e_i = \underline{p_{ii}} > 0 \qquad \text{because } e_i \neq 0$$

$$(0\ 1\ 0)\, P \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$N = \{ x \in \mathbb{R}^3 \mid Cx = 0 \}$

Take $C = [a \ \ b \ \ c]$

$$a x_1 + b x_2 + c x_3 = 0 \qquad x_1 = \frac{-b x_2 - c x_3}{a}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{-b x_2 - c x_3}{a} \\ x_2 \\ x_3 \end{bmatrix}$$

$x \in N$

$$\Longleftrightarrow x = \begin{bmatrix} -\frac{b}{a} \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} -\frac{c}{a} \\ 0 \\ 1 \end{bmatrix} x_3$$

symmetric $\Rightarrow$ e-vectors are orthogonal

$\Rightarrow$ e-vectors are lin indep

$\Rightarrow$ orthogonally diagonalizable.

options:

not symmetric

- not diagonalizable
- diagonalizable but not with orthogonal matrices
- ? diagonalizable ~~with orthogonal matrices~~ not possible.

let's assume A is not symmetric but diagonalizable by an orthogonal matrix: $\exists\ \Sigma, Q$ s.t.

$$\Sigma = Q^T A Q$$

diagonal

orthogonal

$$\Sigma^T = (Q^T A Q)^T$$

$$\Sigma = \underbrace{Q^T A^T Q = Q^T A Q}$$

$$A^T Q = A Q$$

$$A^T = A$$

contradiction