Análisis Matemático para Inteligencia Artificial

Verónica Pastor (vpastor@fi.uba.ar),
Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

19/05/2023

Se acuerdan cuando vimos formas cuadráticas:

$B = \left\{ \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \right\}$

$B' = \left\{ \begin{pmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{pmatrix}, \begin{pmatrix}-1/\sqrt{2}\\1/\sqrt{2}\end{pmatrix} \right\}$

$$\underset{x^T}{\begin{bmatrix}x & y\end{bmatrix}} \underset{A}{\begin{bmatrix}2 & 3\\ 3 & 9\end{bmatrix}} \underset{x}{\begin{bmatrix}x\\ y\end{bmatrix}} = 2x^2 + \underbrace{6xy}_{:(} + 9y^2$$

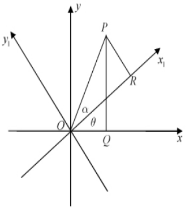
Si rotamos un ángulo θ

la figura dada por: $2x^2 + 6xy + 9y^2 = 1$

a la forma canónica $\lambda_1 x_1^2 + \lambda_2 x_2^2 = 1$

Esto nos lleva a

una matriz de transición $S = \begin{bmatrix} cos(\theta) & -sen(\theta) \\ sen(\theta) & cos(\theta) \end{bmatrix}$



$$x^T A x = (S\tilde{x})^T A (S\tilde{x}) = \tilde{x}^T \underbrace{(S^T A S)}_{A'} \tilde{x} \Rightarrow \underset{\sim}{x}^T A x = \lambda_1 x_1^2 + \lambda_2 x_2^2$$

$\tilde{x} = (x_1, x_2)$

Obs.: S resulta una matriz ortogonal, $S^T A S = S^{-1} A S = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = D$

eigen value

Sea $A \in \mathbb{K}^{n\times n}, \mathbb{R}$ *o* $\mathbb{C}$ diremos que $\lambda \in \mathbb{K}$ es un autovalor (ava) de A y $x \in \mathbb{K}^n, x \neq 0$ es un autovector (ave) asociado a λ si:

eigenvector

$$Ax = \lambda x$$

Interpretación geométrica: A cada vector que se encuentre en la dirección de x, la transformación $T(x) = Ax$ lo contrae (o expende) por un factor λ.

Importante: un autovalor puede ser nulo, un autovector no.

Autovalores y Autovectores: algunas aplicaciones

1. Geometría: curvas planas o superficies.
2. Sistemas dinámicos.
3. Análisis del comportamiento de sistemas de ecuaciones diferenciales.
4. Análisis de estabilidad.
5. Cadenas de Markov.
6. Grafos.
7. Reducción de dimensiones.
8. Cálculo de resonancias del sistema.
9. PageRank.

Hallando los autovalores

Dada $A \in \mathbb{K}^{n\times n}, \mathbb{R}\ o\ \mathbb{C}$, $\lambda \in \mathbb{K}$ es un autovalor de A sii λ es un cero del polinomio característico, $p(\lambda)$, de A:

$$p(\lambda) = det(A - \lambda I)$$

$I \in \mathbb{R}^{n\times n}$

Observación: Puede ocurrir que algún λ sea raíz múltiple de $p(\lambda)$.

Se llama multiplicidad algebraica, m_a, a la cantidad de veces que λ aparece como raíz.

Se llama multiplicidad geométrica, m_g, del autovalor λ, a la cantidad de autovectores l.i. asociados a λ.

$1 \leq m_g \leq m_a$

$A = \begin{pmatrix} 4 & -4 \\ 1 & 0 \end{pmatrix}$ $p(\lambda) = det(A - \lambda I) = \begin{vmatrix} 4-\lambda & -4 \\ 1 & -\lambda \end{vmatrix} = -4\lambda + \lambda^2 + 4 = 0 = \lambda^2 - 4\lambda + 4 = (\lambda - 2)^2$

$\lambda_1 = \lambda_2 = 2$

$\cdot\ m_a(2) = 2$

$A - 2\cdot I = \begin{pmatrix} 2 & -4 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{cases} 2x - 4y = 0 \\ x - 2y = 0 \to x = 2y \end{cases}$ $\underbrace{4y}_{}-4y=0$ $\forall y$ $\Rightarrow \begin{pmatrix} 2y \\ y \end{pmatrix} \forall y$

$= y\begin{pmatrix} 2 \\ 1 \end{pmatrix}$

$\cdot\ m_g(2) = 1$

$ave = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$

Propiedades de los autovectores

Dada $A \in \mathbb{K}^{n\times n}$, el conjunto de todos los autovectores asociados al autovalor λ generan un subespacio de $\mathbb{K}^n$ (autoespacio de A respecto de λ). El conjunto de todos los autovectores de A se llama autoespectro.

$x_1, x_2 \;/\; Ax_1 = \lambda x_1,\; Ax_2 = \lambda x_2$

$A(\overbrace{ax_1 + bx_2}^{x}) = a\cdot A\cdot x_1 + b\cdot A\cdot x_2 = a\lambda x_1 + b\lambda x_2 = \lambda(ax_1 + bx_2) = \lambda\cdot x$

Si λ es un autovalor de $A \in \mathbb{K}^{n\times n}$, el autoespacio E_λ asociado es la solución al sistema homogéneo $(A - \lambda I)x = 0$.

Teorema: Los autovectores $x_1, ..., x_n$ de una matriz $A \in \mathbb{R}^{n\times n}$ con n autovalores distintos de $\lambda_1, ..., \lambda_n$ son linealmente independientes. Es decir, forman una base de $\mathbb{R}^n$.

Teorema: Si $A \in \mathbb{R}^{n\times n}$ es simétrica, entonces existe una base ortonormal correspondiente a un espacio vectorial formado por autovectores de A y además $\lambda_i \in \mathbb{R}$, $\forall i = 1, \ldots, n$.

Ejemplo

$A = \begin{pmatrix} 2 & -3 \\ 2 & -5 \end{pmatrix} \in \mathbb{R}^{2\times 2}$

$p(\lambda) = \det(A - \lambda I) = \det\left(\begin{pmatrix} 2 & -3 \\ 2 & -5 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right) = \begin{vmatrix} 2-\lambda & -3 \\ 2 & -5-\lambda \end{vmatrix} = (2-\lambda)(-5-\lambda) + 6$

$= \lambda^2 + 3\lambda - 4 = 0 = (\lambda - 1)(\lambda + 4) \rightarrow \left.\begin{matrix} \lambda_1 = 1 \\ \lambda_2 = -4 \end{matrix}\right\}$ 2 ave's $\neq$

$\Rightarrow B = \{b_1, b_2\}$ forman matriz

$E_1 : \begin{pmatrix} 1 & -3 \\ 2 & -6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{cases} x - 3y = 0 \Rightarrow x = 3y \\ 2x - 6y = 0 \leftarrow 6y - 6y = 0 \ \forall y \end{cases}$

$\forall y \ \begin{pmatrix} 3y \\ y \end{pmatrix} = y \begin{pmatrix} 3 \\ 1 \end{pmatrix}$

$E_1 = \left\langle \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\rangle$

$E_{-4} : \begin{pmatrix} 6 & -3 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{cases} 6x - 3y = 0 \\ 2x - y = 0 \rightarrow y = 2x \quad \forall x \end{cases}$

$E_{-4} = \left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\rangle = \left\{ \alpha \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \ \alpha \in \mathbb{R} \right\}$

$S = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$

$\quad b_1 \ \ b_2$

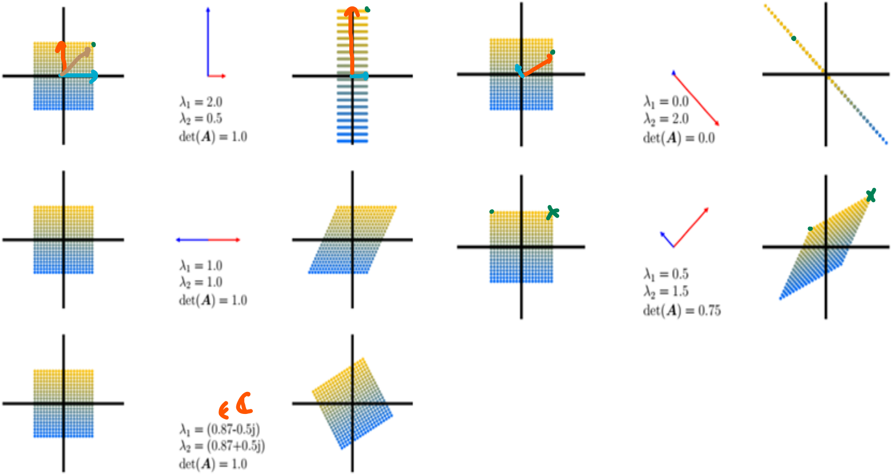
$A = \underbrace{S}_{B \to E} \cdot \underbrace{D}_{\begin{pmatrix} 1 & 0 \\ 0 & -4 \end{pmatrix}} \cdot \underbrace{S^{-1}}_{E \to B}$

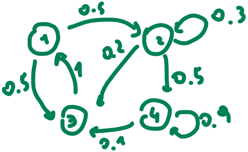
$\overline{S^{-1} \cdot A \cdot S} = D = \begin{pmatrix} 1 & 0 \\ 0 & -4 \end{pmatrix}$

$\sim \mathbb{R}^2$

$S^{-1} = \frac{1}{|S|} \cdot \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ si $S = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

Sea $A \in \mathbb{K}^{n \times n}$: Área: $det(A) = \prod_{i=1}^{n} \lambda_i$ y Perímetro: $Tr(A) = \sum_{i=1}^{n} \lambda_i$





$$P = \begin{pmatrix} 0 & 0.5 & 0.5 & 0 \\ 0 & 0.3 & 0.2 & 0.5 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0.9 \end{pmatrix}$$

$\pi_0 = (p_1, p_2, p_3, p_4)$

Diagonalización

$$\lim_{n\to\infty} \pi_0 \cdot P^n = \pi = (\pi_1, \pi_2, \pi_3, \pi_4)$$

$= (0.1, 0.3, 0.2, 0.4)$

ej.

$\forall \pi_0$

vep asociado a $\lambda = 1$ de P

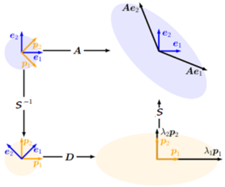
Diagonalización

Definición:

Una matriz $A \in \mathbb{K}^{n\times n}$ es diagonalizable si $\exists S \in \mathbb{K}^{n\times n}$ no singular (es decir, tiene inversa) tal que:

$$S^{-1}AS = D \iff A = S\cdot D\cdot S^{-1}$$

donde D es una matriz *diagonal*.



Teorema: Sea $A \in \mathbb{K}^{n\times n}$ es diagonalizable sii A tiene n autovectores linealmente independientes. $\iff$ $B = \{b_1, \dots, b_n\}$

con asociados $\lambda_1, \dots, \lambda_n$ que NO necesariamente son $\neq$

$\forall v \in \mathbb{K}^n \ \exists! (\alpha_1, \dots, \alpha_n) \in \mathbb{K}^n /$

$v = \alpha_1 \cdot b_1 + \dots + \alpha_n \cdot b_n$

$$\Rightarrow A\cdot v = A\cdot(\alpha_1\cdot b_1 + \dots + \alpha_n \cdot b_n) = \alpha_1 \cdot \underbrace{A\cdot b_1}_{\lambda_1 \cdot b_1} + \dots + \alpha_n \cdot \underbrace{A b_n}_{\alpha_n \cdot b_n} =$$

$$= (\alpha_1 \cdot \lambda_1)\cdot b_1 + \dots + (\alpha_n \cdot \lambda_n)\, b_n = T_B \cdot \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_n \end{pmatrix} \cdot T_B^{-1}$$

con $T_B = \begin{pmatrix} | & & | \\ b_1 & \dots & b_n \\ | & & | \end{pmatrix}$

Conclusiones para $A \in \mathbb{K}^{n\times n}$

1. A es diagonalizable $\Rightarrow$ los vectores columna de la matriz de diagonalización S son los autovectores de A y los elementos de D son los autovalores de A.
2. S no es única, se pueden reordenar columnas, etc $\Rightarrow$ D será distinta.
3. A tiene n autovalores distintos $\Rightarrow$ los autovectores correspondientes son l.i. $\Leftarrow$ A es diagonalizable. $\Rightarrow A \in \mathbb{R}^{n\times n}$
4. A tiene menos de n autovectores l.i. $\Rightarrow$ A no es diagonalizable.
5. Si hay autovalores repetidos y si:
   - $m_a = m_g$ (en los ava repetidos) $\Rightarrow$ los ave son l.i.
   - $m_a > m_g$ (en un ava repetido) $\Rightarrow$ A no es diagonalizable.
6. Si $A \in \mathbb{C}^{n\times n}$ es una matriz hermítica ($A = \overline{A^T} = A^H$) $\Rightarrow$ A es diagonalizable y las columnas forman una BON.

si $A \in \mathbb{R}^{n\times n}$ y simétrica

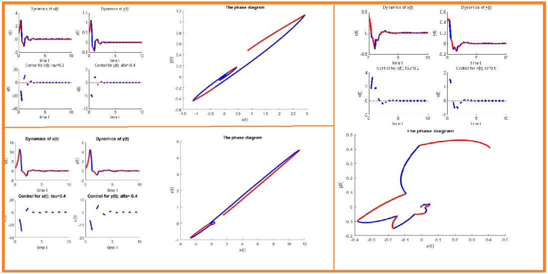
Matrices en bloques de Jordan

Sea $L : \mathbb{V} \to \mathbb{V}$ endomorfismo sobre $\mathbb{K} - e.v.$ con $dim(\mathbb{V}) = n$. Si $p(\lambda)$ se factoriza en $\mathbb{K}$ entonces existe una base que viene dada por la matriz en bloques $(m < n)$.

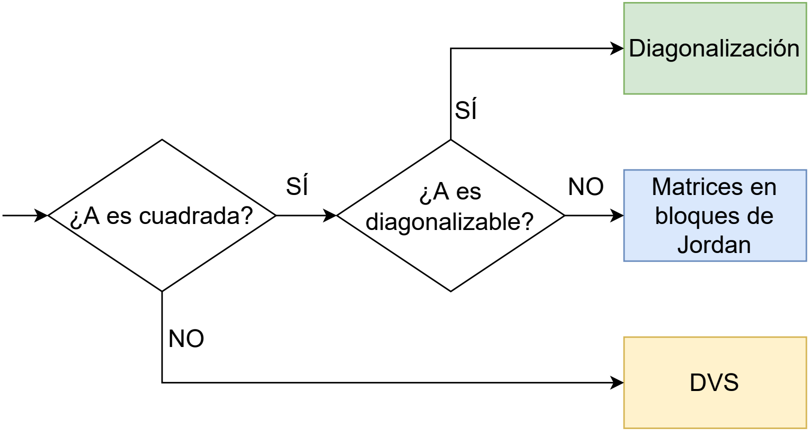
$$J = \left[\begin{array}{ccccc} A_1 & . & . & . & 0 \\ 0 & . & & & . \\ . & & . & & . \\ . & & & . & . \\ 0 & . & . & . & A_m \end{array} \right] \in \mathbb{K}^{n \times n}, \; A_k = \left[\begin{array}{ccccc} \lambda_k & 1 & 0 & \dots & 0 \\ 0 & . & . & & . \\ . & & . & . & . \\ . & & & . & 1 \\ 0 & . & . & . & \lambda_k \end{array} \right], \; \sum_{k=1}^{m} \lambda_k = n.$$

Obs.: Cuando es diagonalizable $m = n$.

Aplicación en Control



¿Qué puede pasar al intentar diagonalizar?



Descomposición en Valores Singulares (DVS)

¿Qué pasa cuando A no es cuadrada?

Teorema: Dada $A \in \mathbb{C}^{m\times n}$, siempre se puede obtener una matriz hermítica, semidefinida positiva $S \in \mathbb{C}^{n\times n}$ como:

$$S \doteq \underbrace{A^H}_{n\times m}\underbrace{A}_{m\times n} \quad \in \mathbb{R}^{n\times n}$$

Obs.: $\tilde{S} \in \mathbb{C}^{m\times m}$ definida como $\tilde{S} \doteq AA^H$ también resulta hermítica y semidefinida positiva.

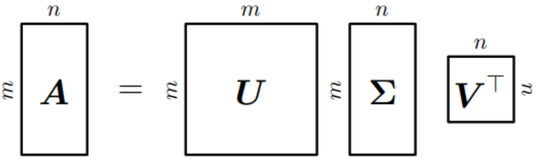
Obs. (II): Como $S, \tilde{S}$ son semidef. positivas, todos sus autovalores son reales **no negativos**.

Teorema: Sea $A \in \mathbb{R}^{m\times n}$ una matriz de rango $r \in \{0, \dots, \overbrace{\min\{m, n\}}^{p}\}$. La **descomposición en valores singulares** (DVS, SVD en inglés) de A se define como:

$$\overbrace{A}^{m\times n} = \overbrace{U}^{m\times m} \cdot \overbrace{\Sigma}^{m\times n} \cdot \overbrace{V^T}^{n\times n}$$

donde:

- $U \in \mathbb{R}^{m\times m}$ es la matriz ortogonal formada por los aves de AA^T.
- $\Sigma \in \mathbb{R}^{m\times n}$ es la matriz "diagonal" de valores singulares $\sigma_i = \sqrt{\lambda_i}$, con λ_i los primeros $p = \min\{m, n\}$ avas de AA^T y $A^T A$.
- $V \in \mathbb{R}^{n\times n}$ es la matriz ortogonal formada por los aves de $A^T A$.



$m=3$, $n=5$

$$\begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{pmatrix}$$

Ejemplo: Hallar una DVS

Sea $A = \overset{2\times 3}{\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}$

$A^T \cdot A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

$tr(A^T A) = 3$

$\det(A^T A) = 0$

$\lambda_1 = 0 \quad \lambda_2 = 1 \quad \lambda_2 = 2$

$\Sigma = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

avas A^TA: $v_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \; v_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \; v_0 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$

$\Rightarrow V = \begin{pmatrix} | & | & | \\ v_2 & v_1 & v_0 \\ | & | & | \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \end{pmatrix}$

$A \cdot A^T = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \rightarrow u_1, u_2 \Rightarrow U = \begin{pmatrix} | & | \\ u_1 & u_2 \\ | & | \end{pmatrix}$

$A = U \cdot \Sigma \cdot V^T \Leftrightarrow A \cdot V = U \cdot \Sigma \Rightarrow u_i = \dfrac{A \cdot v_i}{\sigma_i} = \dfrac{A \cdot v_i}{\sqrt{\lambda_i}}$

x ej $u_2 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix} \cdot \dfrac{1}{\sqrt{2}} = \begin{pmatrix} \sqrt{2} \\ 0 \end{pmatrix} \cdot \dfrac{1}{\sqrt{2}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

en NumPy: np.linalg.svd(A)

Descomposición en Valores Singulares (DVS)

DVS Compacta, Reducida y Truncada

Sea $A \in \mathbb{R}^{m\times n}$ con DVS $A = U\Sigma V^T$:

en el ejemplo

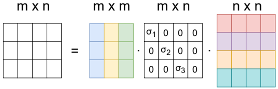
$m < n$

$$\Sigma = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

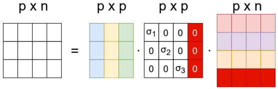
3x3

2x2

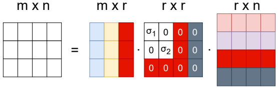
p



- Si $p = min\{m, n\}$, la DVS compacta es $A = U_p\Sigma_p V_p^T$



- Si $\sigma_1, \ldots, \sigma_r > 0, \sigma_{r+1}, \ldots, \sigma_p = 0$, la DVS reducida es $A = U_r\Sigma_r V_r^T$



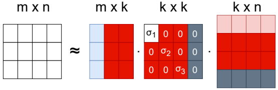
x ej.

$\sigma_1 = 10^4$

$\sigma_2 = 10^2$

$\sigma_3 = 10^{-5} \approx 0$

- Sea $k \in \{1, \ldots, r-1\}$, la DVS truncada es $A \approx U_k\Sigma_k V_k^T$



Pseudoinversa de Moore-Penrose

$D = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \Rightarrow D^{-1} = \begin{pmatrix} 1/a & 0 & 0 \\ 0 & 1/b & 0 \\ 0 & 0 & 1/c \end{pmatrix}$

Sea $A \in \mathbb{R}^{m \times n}$ y el sistema lineal $Ax = y$. La inversa de A no está definida, pero si $m > n$ una posible solución (cuadrados mínimos) era

$A = U \cdot \Sigma \cdot V^T$

$$\hat{x} = (A^T A)^{-1} A^T y$$

$A = U_r \cdot \Sigma_r \cdot V_r^T$

¿Pero qué ocurre si $A^T A$ no es invertible? Se define la pseudoinversa de Moore-Penrose como

$$A^\dagger = V_r \Sigma_r^{-1} U_r^T$$

y $\hat{x} = A^\dagger y$ es (si $m < n$) la solución de mínima norma euclídea o (si $m > n$) la aproximación de mínimo error cuadrático/distancia euclídea.

Obs.: Todo lo anteriormente visto también vale para $A \in \mathbb{C}^{m \times n}$, utilizando $\cdot^H$ en vez de $\cdot^T$. Se utilizaron matrices reales para facilitar la lectura.

Colab

Una aplicación obvia para la DVS es, como vimos recién, la "solución"de sistemas de ecuaciones para los cuales el método de cuadrados mínimos no es factible.

¿Pero es ese el único uso que le podemos dar? Las descomposiciones siempre nos brindan un criterio bajo el cual obtener la *mejor* aproximación de *orden k* para el objeto a descomponer, y eso podría llegar a ser muy útil.