Análisis Matemático para Inteligencia Artificial

Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

Clase 4

Se acuerdan cuando vimos formas cuadráticas:

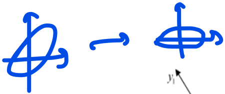
$x^TAx \in \mathbb{R}$

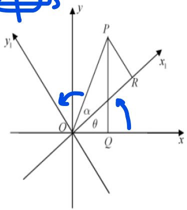
$x \in \mathbb{R}^n$

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

$$\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 2 & 3 \\ 3 & 9 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 2x^2 + 6xy + 9y^2$$

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





$\underset{(x,y)}{x} = S \cdot \underset{(x_1,x_2)}{\tilde{x}}$

$$x^TAx = (S\tilde{x})^TA(S\tilde{x}) = \tilde{x}^T(S^TAS)\tilde{x} \Rightarrow x^TAx = \lambda_1 x_1^2 + \lambda_2 x_2^2$$

$(AB)^T = B^TA^T$

$x \underset{S}{\overset{S^{-1}}{\rightleftarrows}} \tilde{x}$

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

Autovalores y Autovectores: definición

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:

(λ, x) de A
ava ave

$$Ax = \overset{\mathbb{K}}{\lambda} x$$

NO ALTERA DIR. DE 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.

si x es ave asociado a λ ava de A, entonces $k \cdot x$ $\forall k \in \mathbb{K}$ también lo es

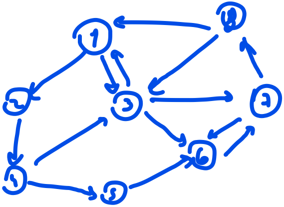
$y = k \cdot x$ con $k \in \mathbb{K}$

$$\Rightarrow \boxed{A \cdot y} = A \cdot k \cdot x \underset{k \text{ escalar}}{=} k \cdot A \cdot x \underset{x \text{ def.}}{=} k \cdot \lambda \cdot x \underset{k, \lambda \text{ escalares}}{=} \lambda \cdot k \cdot x \underset{y \text{ def.}}{=} \boxed{\lambda \cdot y}$$

$\Rightarrow$ de x sólo me importa su dirección eigenvalues eigenvectors

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.

$A_{1:} = (0, 0.5, 0.5, 0, 0, 0, 0)$



$A_{ij} = p_{i \to j}$

$$A = \begin{pmatrix} p_{1\to 1} & \cdots & p_{1\to 8} \\ \vdots & & \\ p_{8\to 1} & \cdots & p_{8\to 8} \end{pmatrix}$$

$p_0 = \left(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \dots, \frac{1}{8}\right)$

$\lambda = 1$, p_∞

$\lambda = 0$, ...

$\lim_{n\to\infty} A^n p_0 = p_\infty$

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:

$\Leftrightarrow \exists \lambda, x \neq 0 \,/\, A \cdot x = \lambda x$

$$p(\lambda) = det(A - \lambda I) \quad \nexists (A-\lambda I)^{-1} \Leftrightarrow |A-\lambda I| = 0$$

$Ax - \lambda x = 0 \Leftrightarrow (A - \lambda I)x = 0$

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.

$1 \leq m_g \leq m_a$

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

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

$\hookrightarrow \lambda_1 = \lambda_2 = 2$

$m_a(2) = 2$

$Ax = 2x \Leftrightarrow (A - 2I)x = 0 \Rightarrow \begin{cases} 2a - 4b = 0 \\ 1a - 2b = 0 \end{cases} \rightsquigarrow^{\forall b}$ $ave = \begin{pmatrix} 2b \\ b \end{pmatrix} = b \begin{pmatrix} 2 \\ 1 \end{pmatrix}$

$\hookrightarrow \begin{pmatrix} 2 & -4 \\ 1 & -2 \end{pmatrix}$

$\hookrightarrow a = 2b$

$\hookrightarrow x = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \to m_g(2) = 1$

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.

λ: x, y son ave asociados a λ ~ $A \Rightarrow z = ax + by \quad \forall a, b \in \mathbb{K}$

$\boxed{Az} = A(ax+by) = Aax + Aby = aAx + bAy = a\lambda x + b\lambda y = \boxed{\lambda(ax+by) = \lambda z}$

z tmb es ave asoc. a λ

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

$\dim(E_\lambda) = m_g(\lambda)$

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 $\underbrace{\text{base}}_{B}$ de $\mathbb{R}^n$.

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

Ejemplo

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

$p(\lambda) = \begin{vmatrix} 2-\lambda & -3 \\ 2 & -5-\lambda \end{vmatrix} = (2-\lambda)(-5-\lambda)+6 = -10+3\lambda+\lambda^2+6$

$= \lambda^2+3\lambda-4 = 0$

$= (\lambda-1)(\lambda+4) = 0$

$\hookrightarrow \lambda_1 = 1, \ \lambda_2 = -4$) $n=2$, $\lambda_1 \neq \lambda_2$

$\Downarrow$

$m_g(1) = m_g(-4) = 1$

x_1 l.i. con x_2

$B = \{x_1, x_2\}$

$\rightarrow$ base de $\mathbb{R}^2$

$\underline{E_1}: (A-1I)x = 0$

$\begin{pmatrix} 1 & -3 \\ 2 & -6 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0 \iff \begin{cases} x-3y=0 \\ 2x-6y=0 \end{cases} \checkmark \ \forall y \rightarrow x = 3y$

$\vec{x} = \begin{pmatrix} 3y \\ y \end{pmatrix} = y\begin{pmatrix} 3 \\ 1 \end{pmatrix}$

$E_1 = \left\langle \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\rangle \hookrightarrow x_1$

$\underline{E_{-4}}: (A+4I)x = 0$

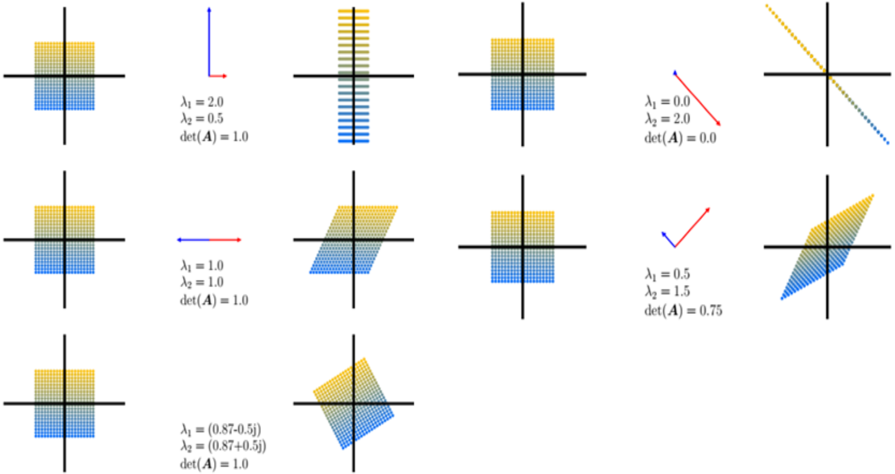
$\begin{cases} 6x-3y=0 \\ 2x-1y=0 \end{cases} \checkmark \ \forall x, \quad y = 2x$

$\vec{x} = \begin{pmatrix} x \\ 2x \end{pmatrix} = x\begin{pmatrix} 1 \\ 2 \end{pmatrix}$

$E_{-4} = \left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\rangle \hookrightarrow x_2$

Interpretación gráfica (Mathematics for Machine Learning)

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



Diagonalización

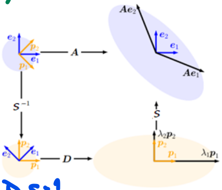
$\begin{pmatrix}1&0&0\\0&2&0\\0&0&3\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix}=\begin{pmatrix}1\cdot1\\2\cdot1\\3\cdot1\end{pmatrix}$

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:

$$S^{-1}AS = D$$

donde D es una matriz *diagonal*.

$\exists S, D \in \mathbb{K}^{n\times n}$ S invertible, D diagonal $/ A = SDS^{-1}$



Teorema: Sea $A \in \mathbb{K}^{n\times n}$ es diagonalizable sii A tiene n autovectores linealmente independientes.

x

$b_1 \quad b_2 \ldots \quad b_n$

$\alpha_1 \quad \alpha_2 \quad \alpha_n$

$d_1 \quad d_2 \quad d_n$

Ax

$\Leftarrow$: $B = \{b_1, \ldots, b_n\}$ con avec asociados $\lambda_1, \ldots, \lambda_n$

$\forall x \in \mathbb{K}^n \quad Ax = S\cdot D\cdot S^{-1}\cdot x$ ① $S^{-1}x = \tilde{x} = (\alpha_1, \ldots, \alpha_n) /$

$x = \alpha_1 b_1 + \alpha_2 b_2 + \ldots + \alpha_n b_n$

A diag

② $D\tilde{x} = (d_1\alpha_1, \ldots, d_n\alpha_n)$ $\quad D = \begin{pmatrix}d_1 & & 0\\ & \ddots & \\ 0 & & d_n\end{pmatrix}$

③ $S(D\tilde{x}) = \begin{pmatrix}| & & |\\ b_1 & \cdots & b_n\\ | & & |\end{pmatrix}\begin{pmatrix}d_1\alpha_1\\ \vdots\\ d_n\alpha_n\end{pmatrix} = d_1\alpha_1\begin{pmatrix}|\\b_1\\|\end{pmatrix} + \ldots + d_n\alpha_n\begin{pmatrix}|\\b_n\\|\end{pmatrix}$

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

$S = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$, $D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$

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. $\Leftrightarrow$ A es diagonalizable.
4. A tiene menos de n autovectores l.i. $\Leftrightarrow$ 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.
7. A es diagonalizable $\Rightarrow$ $A^n = SDS^{-1}SDS^{-1}\cdots = SD^nS^{-1}$

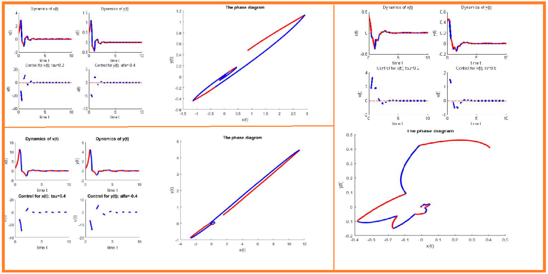
$D^n = \begin{pmatrix} \lambda_1^n & & 0 \\ & \ddots & \\ 0 & & \lambda_n^n \end{pmatrix}$

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 $(m < n)$.

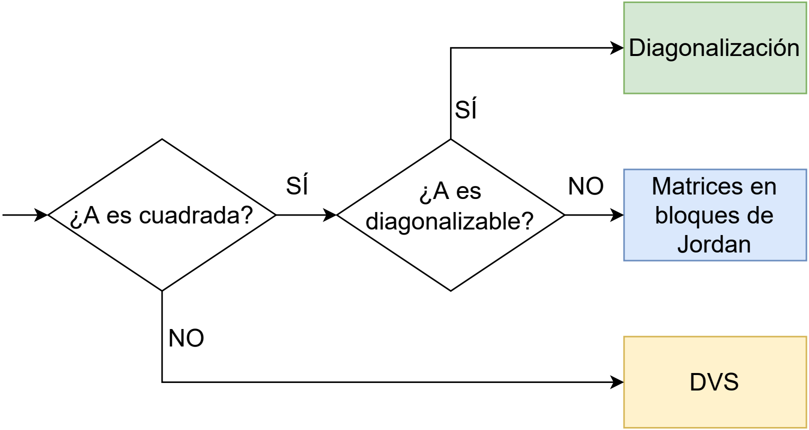
$$J = \begin{bmatrix} A_1 & . & . & . & 0 \\ 0 & . & & & . \\ . & & . & & . \\ . & & & . & . \\ 0 & . & . & . & A_m \end{bmatrix} \in \mathbb{K}^{n\times n}, \; A_k = \begin{bmatrix} \lambda_k & 1 & 0 & \dots & 0 \\ 0 & . & . & & . \\ . & & . & . & . \\ . & & & . & 1 \\ 0 & . & . & . & \lambda_k \end{bmatrix},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

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?

$S^H \doteq \bar{S}^T = S$

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:

$x^T S x \geq 0 \quad \forall x \neq 0$

$$S \doteq A^H A$$

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**.

$p = \min(m,n)$

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

$B_{B_m \to E_m}$ "D"$_{m\times n}$ $B_{E_n \to B_n}$

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

$m\times n$ $\quad$ $m\times m$ $\quad$ $m\times n$ $\quad$ $n\times n$

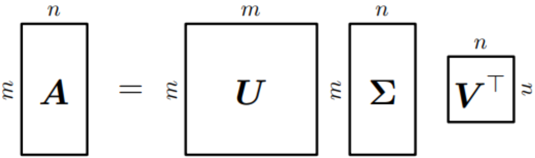
$S \cdot D \cdot S^{-1}$

$$\Sigma = \left(\begin{array}{cc|cc} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \end{array}\right)$$

si $A \in \mathbb{R}^{2\times 4}$

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^TA.
- $V \in \mathbb{R}^{n\times n}$ es la matriz ortogonal formada por los aves de A^TA.



si X es ortogonal

$X^{-1} = X^T$

Ejemplo: Hallar una DVS

Sea $A = \begin{pmatrix}1 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$

1) $A^TA = \begin{pmatrix}1&0\\1&0\\0&1\end{pmatrix}\begin{pmatrix}1&1&0\\0&0&1\end{pmatrix} = \begin{pmatrix}1&1&0\\1&1&0\\0&0&1\end{pmatrix} \in \mathbb{R}^{3\times 3}$ $\quad \lambda_1 = 2,\ \lambda_2 = 1,\ \lambda_3 = 0$

$p = 2 \qquad \in \mathbb{R}^{2\times 3}$

$AA^T = \begin{pmatrix}1&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&0\\1&0\\0&1\end{pmatrix} = \begin{pmatrix}2&0\\0&1\end{pmatrix} \in \mathbb{R}^{2\times 2}$ $\quad \lambda_1 = 2,\ \lambda_2 = 1$

$\sigma_1 = \sqrt{\lambda_1} = \sqrt{2}$

$\sigma_2 = \sqrt{\lambda_2} = 1$

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

avs de A^TA: $b_1 = \begin{pmatrix}1\\1\\0\end{pmatrix},\ b_2 = \begin{pmatrix}0\\0\\1\end{pmatrix},\ b_3 = \begin{pmatrix}1\\-1\\0\end{pmatrix}$

3) $V = \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix} = \begin{pmatrix}1/\sqrt{2} & 0 & 1/\sqrt{2}\\ 1/\sqrt{2} & 0 & -1/\sqrt{2}\\ 0 & 1 & 0\end{pmatrix}$

4) avs de $AA^T \to$ opc 1) buscar como con A^TA

opc 2) despejo $u_i = \dfrac{A \cdot v_i}{\sigma_i}$

$u_1 = \dfrac{\begin{pmatrix}1&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}1/\sqrt{2}\\1/\sqrt{2}\\0\end{pmatrix}}{\sqrt{2}} = \dfrac{\begin{pmatrix}2/\sqrt{2}\\0\end{pmatrix}}{\sqrt{2}} = \begin{pmatrix}1\\0\end{pmatrix}$

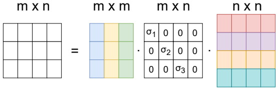
$u_2 = \dfrac{\begin{pmatrix}1&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}0\\0\\1\end{pmatrix}}{1} = \dfrac{\begin{pmatrix}0\\1\end{pmatrix}}{1} = \begin{pmatrix}0\\1\end{pmatrix}$ $\Rightarrow$ $U = \begin{pmatrix}1&0\\0&1\end{pmatrix}$

Descomposición en Valores Singulares (DVS)

DVS Compacta, Reducida y Truncada

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

$$\Sigma = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad 3\times 5$$



compacta 3×3

$p=3$

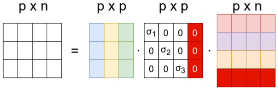
- Si $p = min\{m,n\}$, la DVS compacta es $A = U_p\Sigma_p V_p^T$

reducida 2×2

$r=2$

3, 2, 1, 0,1, 0,05

2, 2, 0,0, 0, 0



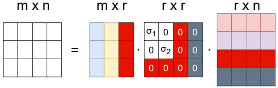
truncada 1×1

$k=1$

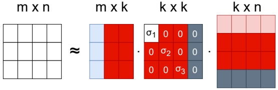
- Si $\sigma_1,\ldots,\sigma_r > 0, \sigma_{r+1},\ldots,\sigma_p = 0$, la DVS reducida es $A = U_r\Sigma_r V_r^T$

100, 10, 1, 0, 0

$\underbrace{10^{10}, 10^{8}, 10^{7},}\ \underbrace{10^{-1}, 10^{-2}}$



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



Pseudoinversa de Moore-Penrose

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

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

¿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.