

Análisis Matemático para Inteligencia Artificial

Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

Clase 4

Autovalores y Autovectores: Motivación

Se acuerdan cuando vimos formas cuadráticas:

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

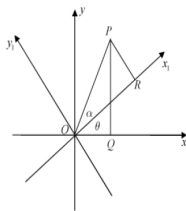
Si rotamos un ángulo θ

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



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

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

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:

$$Ax = \lambda x$$

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 :

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

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 .

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**.

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

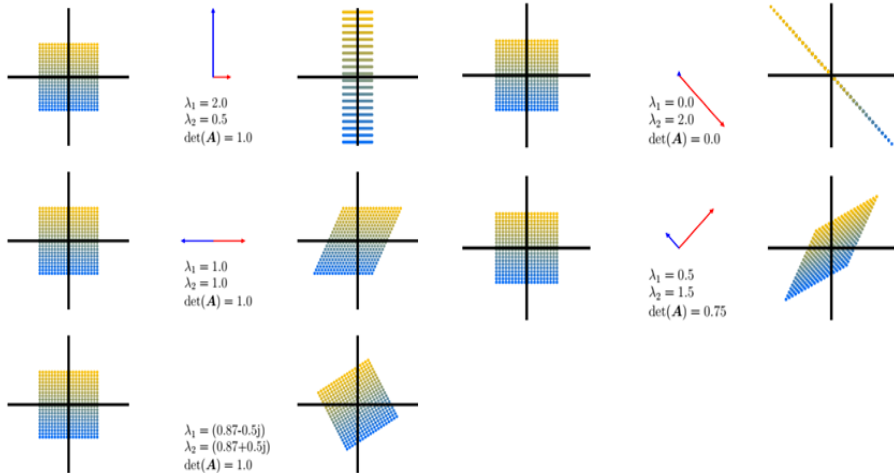
Teorema: Los autovectores x_1, \dots, x_n de una matriz $A \in \mathbb{R}^{n \times n}$ con n autovalores distintos de $\lambda_1, \dots, \lambda_n$ son linealmente independientes. Es decir, forman una base de \mathbb{R}^n .

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 $\lambda_i \in \mathbb{R}$, $\forall i = 1, \dots, n$.

Ejemplo

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

Diagonalización

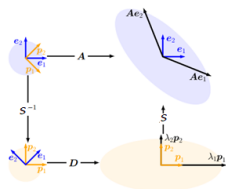
Definición:

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:

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

donde D es una matriz *diagonal*.

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



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. $\Leftrightarrow A$ 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 $\Rightarrow A^n = SDS^{-1}SDS^{-1} \dots = SD^nS^{-1}$

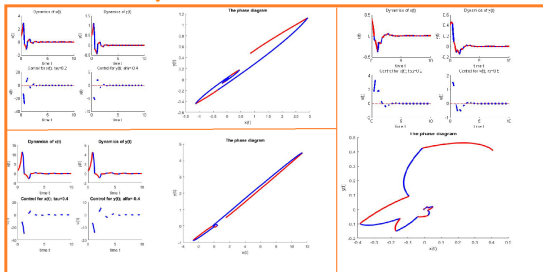
Matrices en bloques de Jordan

Sea $L : \mathbb{V} \rightarrow \mathbb{V}$ 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 ($m < n$).

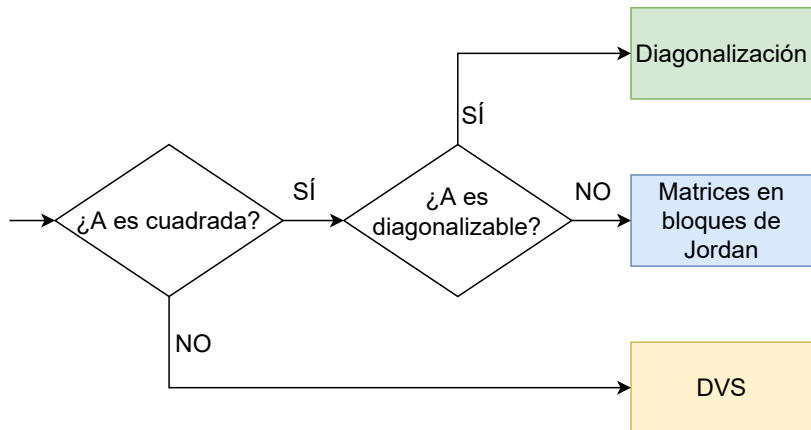
$$J = \begin{bmatrix} A_1 & . & . & . & 0 \\ 0 & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & . & . & . & A_m \end{bmatrix} \in \mathbb{K}^{n \times n}, A_k = \begin{bmatrix} \lambda_k & 1 & 0 & \dots & 0 \\ 0 & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & 1 \\ 0 & . & . &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

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?

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:

$$S \doteq A^H A$$

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

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

DVS: Definición

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:

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

donde:

- $U \in \mathbb{R}^{m \times m}$ es la matriz ortogonal formada por los avas 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 avas de $A^T A$.

$$\begin{array}{c} n \\ \boxed{A} \\ m \end{array} = \begin{array}{c} m \\ \boxed{U} \\ m \end{array} \begin{array}{c} n \\ \boxed{\Sigma} \\ m \end{array} \begin{array}{c} n \\ \boxed{V^T} \\ n \end{array}$$

Ejemplo: Hallar una DVS

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

Descomposición en Valores Singulares (DVS)

DVS Compacta, Reducida y Truncada

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

$$\begin{array}{c} m \times n \\ \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \end{array} = \begin{array}{c} m \times m \\ \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \end{array} \cdot \begin{array}{c} m \times n \\ \begin{array}{|c|c|c|c|} \hline \sigma_1 & 0 & 0 & 0 \\ \hline 0 & \sigma_2 & 0 & 0 \\ \hline 0 & 0 & \sigma_3 & 0 \\ \hline \end{array} \end{array} \cdot \begin{array}{c} n \times n \\ \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \end{array}$$

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

$$\begin{array}{c} p \times n \\ \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \end{array} = \begin{array}{c} p \times p \\ \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \end{array} \cdot \begin{array}{c} p \times p \\ \begin{array}{|c|c|c|c|} \hline \sigma_1 & 0 & 0 & 0 \\ \hline 0 & \sigma_2 & 0 & 0 \\ \hline 0 & 0 & \sigma_3 & 0 \\ \hline \end{array} \end{array} \cdot \begin{array}{c} p \times n \\ \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \end{array}$$

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

$$\begin{array}{c} m \times n \\ \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \end{array} = \begin{array}{c} m \times r \\ \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \end{array} \cdot \begin{array}{c} r \times r \\ \begin{array}{|c|c|c|c|} \hline \sigma_1 & 0 & 0 & 0 \\ \hline 0 & \sigma_2 & 0 & 0 \\ \hline 0 & 0 & \sigma_3 & 0 \\ \hline \end{array} \end{array} \cdot \begin{array}{c} r \times n \\ \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \end{array}$$

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

$$\begin{array}{c} m \times n \\ \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \end{array} \approx \begin{array}{c} m \times k \\ \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \end{array} \cdot \begin{array}{c} k \times k \\ \begin{array}{|c|c|c|c|} \hline \sigma_1 & 0 & 0 & 0 \\ \hline 0 & \sigma_2 & 0 & 0 \\ \hline 0 & 0 & \sigma_3 & 0 \\ \hline \end{array} \end{array} \cdot \begin{array}{c} k \times n \\ \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \end{array}$$

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.

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.