

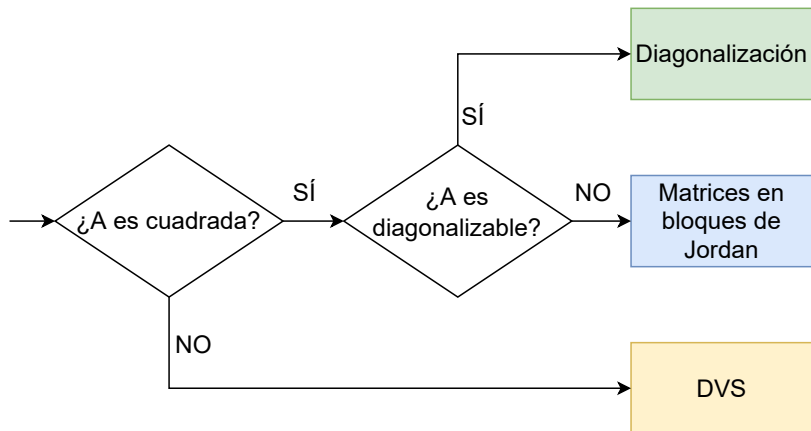
Análisis Matemático para Inteligencia Artificial

Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

Clase 4

¿Qué puede pasar al intentar diagonalizar?



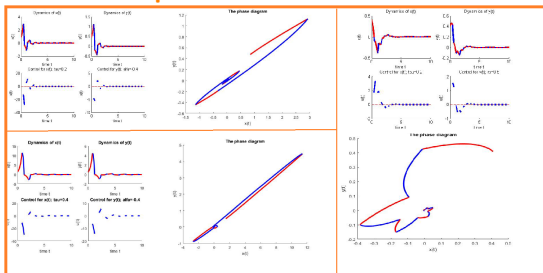
Matrices en bloques de Jordan

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 pasa cuando A no es cuadrada?

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

$$S \doteq A^T A$$

Obs.: $\tilde{S} \in \mathbb{R}^{m \times m}$ definida como $\tilde{S} \doteq A A^T$ también resulta simétrica y semidefinida positiva.

Obs. (II): Como S, \tilde{S} son semidef. positivas, todos sus autovalores son reales **no negativos** (¡incluso si A tomara valores complejos!).

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 $A^T A$.

The diagram shows the matrix equation $A = U \cdot \Sigma \cdot V^T$ with dimensions indicated by subscripts and superscripts on the matrix boxes:

- Matrix A is a vertical rectangle with dimension n written above it and m written to its left.
- Matrix U is a square with dimension m written above it and m written to its left.
- Matrix Σ is a vertical rectangle with dimension n written above it and m written to its left.
- Matrix V^T is a horizontal rectangle with dimension n written above it and m written to its right.

The matrices are connected by an equals sign: $A = U \cdot \Sigma \cdot V^T$.

Ejemplo: Hallar una DVS

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

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.

Si bien la diagonalización y la DVS son las factorizaciones más importantes, existen otras útiles. Algunas de ellas (con aplicaciones en ML) son:

- **LU:** $PA = LU$ con P matriz de permutación de filas, L triangular inferior con diagonal de 1s, U triangular superior. Muy eficiente para resolver problemas del tipo $Ax = b$ cuando b va variando.
- **QR:** $A = QR$ con Q matriz ortogonal y R triangular superior. Algoritmos más simples pero un poco menos eficientes que LU para resolver $Ax = b$, también provee ortogonalización.
- **Cholesky:** $A = LL^T$ para A simétrica (semi)definida positiva, con L triangular inferior. Permite efficientizar muchos cálculos como A^{-1} , updates a A , etc. Muy usado en modelos probabilísticos con matrices de covarianza.

¿Por qué hacer una factorización?

Tomemos por ejemplo la factorización de Cholesky con $A = LL^T$. Si queremos resolver la ecuación $Ax = b$, debemos:

- 1 Invertir A , para obtener A^{-1} : orden $2n^3 + \mathcal{O}(n^2)$
- 2 Computar $x = A^{-1}b$: orden $2n^2$

En total, está dominado por $2n^3$.

Si en cambio usamos la factorización de Cholesky tenemos:

- 1 Computar L tal que $LL^T = A$: orden $\frac{1}{3}n^3$
- 2 Resolver sistema triangular $L^T y = b$: orden n^2
- 3 Resolver sistema triangular $Lx = y$: orden n^2

Si bien el orden de magnitud es el mismo, ahora está dominado por $\frac{1}{3}n^3$.

Nota: hay opciones de pivoteo directo (o vía LU), que son del orden $\frac{2}{3}n^3$.

Vamos a tomar dos modelos clásicos de clasificación multiclase, LDA/QDA y vamos a:

- Implementarlos
- Tensorizarlos
- Explotar su estructura
- Medir las ganancias en tiempo y memoria

Los objetivos del TP son:

- Familiarizar al alumno con el ida y vuelta *cuenta-código*
- Familiarizar y “amigar” al alumno con la paralelización de operaciones matriciales usando tensores de rango 3+
- Concientizar al alumno sobre las ventajas del uso inteligente y eficiente de recursos

Ejemplo: tickets de compra de amigos Ana, Beto y Carlos.

Recordar que está disponible la encuesta de clase! Completarla es cortito y sirve para ir monitoreando el estado del curso.

¿Dónde encontrarla? En la hoja de notas (e.g. "Notas CEIA 10Co2024"), abajo de todo, junto al link de las grabaciones de las clases.