Descomposicón en valores singulares de una transformación lineal entre espacios euclídeos

Eugenio Hernández, Maria-Angeles Zurro

2 de noviembre de 2020

Índice

1. El teorema espectral

Fijaremos en esta sección un espacio vectorial euclídeo $\mathbf{E} = (E, \langle , \rangle)$ con E un espacio vectorial sobre el cuerpo $\mathbb{R}$ de dimensión finita $n > 0$. Consideremos una transformación lineal $A : E \to E$. Esta transformación admite una factorización según los siguientes resultados.

Teorema 1.1 (Teorema espectral). *1. Dada una transformación ortogonal O en $O(n;\mathbb{R})$, existen una transformación ortogonal $C \in O(n;\mathbb{R})$ y una transfor-*

mación del tipo

$$J = \begin{pmatrix} I_p & 0 & 0 & \cdots & 0 \\ 0 & -I_q & 0 & \cdots & 0 \\ 0 & 0 & G_{\alpha_1} & \cdots & 0 \\ 0 & 0 & \cdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & G_{\alpha_k} \end{pmatrix} \ , \ \textit{donde } G_{\alpha_i} \textit{ es el giro de ángulo } \alpha_i \ , \tag{1}$$

tales que se da la factorización:

$$O = CJC^t \ . \tag{2}$$

2. *Dada una transformación autoadjunta S en E, entonces existen Una transformación ortogonal $C \in O(n;\mathbb{R})$, y una matriz diagonal $D \in \mathbb{M}_{n,n}(\mathbb{R})$ tal que*

$$S = CDC^t \ . \tag{3}$$

Demostración. Véase [2], página 397. □

Teorema 1.2. *Dada una transformación invertible $A \in GL(n;\mathbb{R})$ se tiene que:*

1. *Existen una transformación ortogonal s $O \in O(n;\mathbb{R})$, y una transformación autoadjunta S tal que se da la factorización:*

$$A = OS \ . \tag{4}$$

2. *Existen dos transformaciones ortogonales $O_1, O_2 \in O(n;\mathbb{R})$, y una transformación diagonal D tal que se da la factorización:*

$$A = O_1 D O_2 \ . \tag{5}$$

Demostración. La demostración de la fórmula (4) se encuentra en [2], página 400. Para demostrar (5), basta observar que, por (4), $A = OS$, y como, por (3), $S = CDC^t$, concluímos que:

$$A = OS = (OC)D(C^t), \tag{6}$$

luego, basta definir $O_1 = OC$ y $O_2 = C^t$ para obtener la fórmula requerida (5). □

2. Factorización ortogonal de una transformación lineal

La generalización de la fórmula (5) para matrices $m \times n$, no necesariamente cuadradas, se conoce con el nombre de **descomposición en valores singulares** de la matriz (o abreviadamente por sus siglas en inglés como *SVD decomposition*). Tiene muchas aplicaciones prácticas (véase por ejemplo `https://en.wikipedia.org/wiki/Singular_value_decomposition`), por esta razón en la sección 5 aprenderemos cómo calcularla para cualquier matriz real.

Teorema 2.1 (Descomposición en valores singulares)**.** *Dada una matriz no nula $A \in \mathbb{M}_{m.n}(\mathbb{R})$, existen matrices ortogonales $U \in O(m;\mathbb{R})$ y $V \in O(n;\mathbb{R})$ tales que*

$$A = U\Sigma V^t \ , \tag{7}$$

donde Σ es una matriz maximal diagonal, es decir,

$$\Sigma = (D \ \mathbf{0}_{n-m}) \quad si \ n \geq m, \ \grave{o} \quad \Sigma = \begin{pmatrix} D \\ \mathbf{0}_{m-n} \end{pmatrix} \quad en \ caso \ contrario \ , \tag{8}$$

para una cierta matriz diagonal D.

Definición 2.1. *Dada una transformación lineal $A : \mathbb{R}^n \to \mathbb{R}^m$, la factorización dada en el teorema 2.1 se llama la* **descomposición en valores singulares** *de A. Los elementos de la diagonal principal de la matriz diagonal D en (8) se llaman los* **valores singulares** *de A.*

A continuación procederemos a dar una demostración detallada del teorema 2.1.

Demostración. Comencemos fijando bases ortonormales, $\beta_1 = \{e_1, \ldots, e_n\}$, $\gamma_1 = \{w_1, \ldots, w_m\}$ en $\mathbb{R}^n$ y $\mathbb{R}^m$ respectivamente. Sea A la matriz de la transformación dada en estas bases, $A : \mathbb{R}^n_{\beta_1} \to \mathbb{R}^m_{\gamma_1}$.

A continuación procederemos a buscar bases apropiadas, β_2 de $\mathbb{R}^n$ y γ_2 de $\mathbb{R}^m$, tales que la transformación A tenga una matriz en estas bases de la forma (8). Para esto seguiremos los siguientes pasos.

1. Definimos la matriz simétrica $S = A^t A \in \mathbb{M}_{n,n}(\mathbb{R})$. Por la fórmula (3) se tiene que $S = V \Delta V^t$ donde Δ es una matriz diagonal no nula , y las columnas de V, $\vec{v}_i$, forman una base ortonormal $\mathbb{R}^n$ formada por autovectores de S. Pongamos

$\beta_2 = \{\vec{v}_1, \ldots, \vec{v}_n\}$. Además los autovalores no nulos de S son no negativos ya que:

$$0 \leq ||A(\vec{v}_i)||^2 = \langle \vec{v}_i, S(\vec{v}_i) \rangle = \langle \vec{v}_i, \lambda_i \vec{v}_i \rangle = \lambda_i ||\vec{v}_i||^2 = \lambda_i \ . \tag{9}$$

En consecuencia $S(\vec{v}_i) = \lambda_i \vec{v}_i$, y podemos asumir la reordenación de β_2 dada por el criterio

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0 \ . \tag{10}$$

2. Definimos los siguientes escalares

$$\sigma_i = \sqrt{\lambda_i}, \quad i = 1, \ldots, r, \tag{11}$$

y los siguientes vectores:

$$\vec{u}_i = \frac{1}{\sigma_i} A(\vec{v}_i), \quad i = 1, \ldots, r. \tag{12}$$

Observa que $\mathcal{C} = \{\vec{u}_1, \ldots, \vec{u}_r\}$ es un sistema libre ortonormal en $\mathbb{R}^m$ ya que

- $||\vec{u}_i||^2 = \langle \vec{u}_i, \vec{u}_i \rangle = \frac{1}{\sigma_i^2} ||A(\vec{v}_i)||^2 = \frac{\lambda_i}{\sigma_i^2} = 1.$
- $\langle \vec{u}_i, \vec{u}_j \rangle = \frac{1}{\sigma_i \sigma_j} \langle A(\vec{v}_i), A(\vec{v}_j) \rangle = \frac{1}{\sigma_i \sigma_j} \langle \vec{v}_i, S(\vec{v}_j) \rangle = \frac{\lambda_j}{\sigma_i \sigma_j} \langle \vec{v}_i, \vec{v}_j \rangle = 0.$

3. Tomamos vectores unitarios $\vec{u}_{r+1}, \ldots, \vec{u}_m$ tales que $\gamma_2 = \mathcal{C} \cup \{\vec{u}_{r+1}, \ldots, \vec{u}_m\}$ forme una base ortonormal de $\mathbb{R}^m$.

4. Definimos la matriz $m \times n$

$$\Sigma := U^t A V \ . \tag{13}$$

La matriz $\Sigma = (\Sigma_{ij})$ satisface la identidad (8) para la matriz diagonal D dada por

$$D = \begin{pmatrix} \sigma_1 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \ldots & \ddots & \ldots & \vdots & \ldots & \vdots \\ 0 & \ldots & 0 & \sigma_r & 0 & \ldots & 0 \\ 0 & \ldots & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \ldots & \ldots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & \ldots & 0 & 0 & \ldots & 0 \end{pmatrix} , \tag{14}$$

y cuyo tamaño es $\nu \times \nu$ para $\nu = \text{mín}\{n, m\}$. En efecto, si $\Sigma = (\Sigma_{ij})$, entonces

$$\Sigma_{ij} = \langle \vec{u}_i, A(\vec{v}_j) \rangle = \begin{cases} \langle \vec{u}_i, \vec{0} \rangle = 0 & \text{si } j > r \\ \\ \langle \vec{u}_i, \sigma_j \vec{u}_j \rangle = \begin{cases} \sigma_i & i = j \\ 0 & i \neq j \end{cases} & \text{si } j \leq r \end{cases} , \tag{15}$$

lo que da la fórmula (8) para D la matriz diagonal definida en (14).

□

De la demostración anterior se deducen los siguientes corolarios.

Corolario 2.1. *Si hay r valores singulares de A no nulos, entonces el rango de A es r.*

Corolario 2.2. *Con la notaciones del teorem 2.1, se tiene que*

$$Traza(A^t A) = \sigma_1^2 + \cdots + \sigma_r^2 \ . \tag{16}$$

3. Descomposición SVD y aproximación matricial

Dada una matriz no nula $A \in \mathbb{M}_{m.n}(\mathbb{R})$, con las notaciones del teorema 2.1, podemos tener una escritura de A como

$$A = \sigma_1 \vec{u_1} \vec{v_1}^t + \cdots + \sigma_r \vec{u_r} \vec{v_r}^t \ , \tag{17}$$

donde $\vec{u_1}, \ldots, \vec{u_r}$ y $\vec{v_1}, \ldots, \vec{v_r}$ son las r primeras columnas de las matrices U y V respectivamente, y σ_i es el i-ésimo valor singular no nulo de la matriz A.

Consideraremos también, para cada $k < r$, la siguiente matriz $A_k \in \mathbb{M}_{m.n}(\mathbb{R})$,

$$A_k = \sigma_1 \vec{u_1} \vec{v_1}^t + \cdots + \sigma_k \vec{u_k} \vec{v_k}^t \ , \tag{18}$$

que se obtiene tomando sólo los k valores singulares más grandes de la matriz A. Un resultado importante relativo a estas matrices $A_1, A_2, \ldots, A_k$ es el teorema de aproximación (Teorema 3.1), que pone de manifiesto su carácter de mejor aproximación de A por matrices $\mathbb{M}_{m.n}(\mathbb{R})$ de rango $1, 2, \ldots, k$ respectivamente. El concepto de aproximación implica necesariamente la idea de medir, o al menos acotar, el error cometido. Para esto definimos normas en el espacio de matrices.

Definición 3.1. *Sea $A \in \mathbb{M}_{m.n}(\mathbb{R})$ una matriz no nula. La norma espectral de A viene dada por*

$$||A||_2 = \sigma_1 \ . \tag{19}$$

La norma de Frobenius de A viene dada por

$$||A||_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} \ . \tag{20}$$

Nota 3.1. *Es fácil comprobar las siguientes igualdades:*

- $||A - A_k||_2 = \sigma_{k+1}$.
- $||A - A_k||_F = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2}$

Teorema 3.1 (Teorema de aproximación)**.** *Supongamos que, a partir del teorema 2.1, tenemos A expresada como*

$$A = \sigma_1 \vec{u_1} \vec{v_1}^t + \cdots + \sigma_r \vec{u_r} \vec{v_r}^t \ . \tag{21}$$

Sea k un natural no nulo $k < r$. Entonces, la mejor aproximación de A, en la norma espectral, mediante matrices de su mismo tamaño y de rango $\leq k$, es la matriz $A_k = \sigma_1 \vec{u_1} \vec{v_1}^t + \cdots + \sigma_k \vec{u_k} \vec{v_k}^t$.

La demostración de este teorema se puede ver [1]; no la incluiremos ya que excede los objetivos de este curso. El teorema de aproximación fue publicado en 1936 por Eckart y Young. Estos autores utilizaron la norma de Frobenius, y señalan en su trabajo la cantidad de trabajo numérico que involucra su aplicación en casos realistas. Los avances recientes en la construcción de algoritmos de cálculo de la SVD se incorporan en la librería LAPACK de Álgebra Lineal Numérica (`http://netlib.org/lapack/`) y está integrada en MATLAB.

Una aplicación del teorema 3.1 es el uso de las matrices aproximantes para la recuraración de la información. Algunas otras pueden verse en el interesante trabajo [3], publicado en 2005 en La Gaceta de la Real Sociedad Matemática Española (RS-ME). La sección 4 muestra cómo estas técnicas de aproximación se pueden aplicar. La sección 5 incluye dos ejemplos. El ejemplo 5.1 está calculado con Sagemath [4], y en él se calculan explícitamente las aproximantes A_k.

4. Solución aproximada de sistemas lineales

Consideramos una matriz A de tamaño $m \times n$, y un vector $\vec{b} \in \mathbb{R}^m$. Supongamos que el siguiente sistema no tiene solución:

$$A\vec{x} = \vec{b}. \tag{22}$$

La matriz A define una transformación lineal $A : \mathbb{R}^n \to \mathbb{R}^m$ cuyo subespacio imagen será denotado por $W = Im(A)$. Como el sistema (22) no tiene solución, se tiene que $\vec{b} \notin W$. Es el objetivo de esta sección mostrar cómo podemos dar una solución aproximada a dicho sistema. El vector $\vec{x}_0$, que hace mínima la diferencia entre $A\vec{x}_0 - \vec{b}$ se llama *solución de mínimos cuadrados del sistema de* (22). A continuación detallaremos cómo construir una tal aproximación óptima. En concreto, se tiene que verificar que

$$||\vec{b} - A\vec{x}_0|| = \text{mín}\left\{||\vec{b} - A\vec{x}|| \ : \vec{x} \in \mathbb{R}^n\right\} \tag{23}$$

Luego, buscamos $\vec{x}_0 \in \mathbb{R}^n$ talque la proyección de $\vec{b}$ sobre W sea $A\vec{x}_0$, $p_W(\vec{b}) = A\vec{x}_0$. La solución a este problema la da el siguiente resultado.

Teorema 4.1. *Sea Σ^+ la matriz $m \times n$ cuyos elementos diagonales son los $1/\sigma_i$, es decir los inversos de los valores singulares no nulos de A. Sea $A^+ = V\Sigma^+U^t$, con U y V como en el teorema 2.1. Entonces,*

$$x^+ = A^+\vec{b} \tag{24}$$

es una solución al problema de mínimos cuadrados para el sistema (22)*.*

Demostración. Observamos en primer lugar que, por el teorema 2.1, se tiene $U \in O(m; \mathbb{R})$ y también $V \in O(n; \mathbb{R})$. En consecuencia tenemos la igualdad:

$$||\vec{b} - A\vec{x}|| = ||UU^t\vec{b} - U\Sigma V^t\vec{x}|| = ||U\left(U^t\vec{b} - \Sigma V^t\vec{x}\right)|| = ||U^t\vec{b} - \Sigma\vec{y}|| \tag{25}$$

donde $\vec{y} := V^t\vec{x} = (y_1, \ldots, y_n)^t$. En consecuencia, minimizar $\vec{b} - A\vec{x}$ es equivalente a minimizar $U^t\vec{b} - \Sigma\vec{y}$.

Consideremos el vector constante $(c_1, \ldots, c_m)^t = U^t\vec{b}$. Entonces, con las notaciones del teorema 2.1, la ecuación (25) se reescribe como

$$||U^t\vec{b} - \Sigma\vec{y}||^2 = \sum_{i=1}^{r}(\sigma_i y_i - c_i)^2 + C,$$

donde $C = \sum_{j=r+1}^{m} c_j$ es siempre un número real ≥ 0. Observamos además que

$$\begin{array}{ll} \frac{\partial}{\partial y_j}\left(\sum_{i=1}^{r}(\sigma_i y_i - c_i)^2 + C\right) & = 2\sigma_j(\sigma_j y_j - c_j)\ , \\ \frac{\partial^2}{\partial x_i \partial y_j}\left(\sum_{i=1}^{r}(\sigma_i y_i - c_i)^2 + C\right) & = \delta_{ij} 2\sigma_j^2 \geq 0 \end{array}$$

Por lo que el mínimo buscado se alcanza para $y_i = c_i/\sigma_i$, $1 \leq i \leq r$, y valores reales arbitrarios y_j, para $j > r$. □

Nota 4.1. *Es de señalar que si $A^t A$ no es una matriz invertible, la solución al problema de mínimos cuadrados no es única, como se deduce de la demostración del teorema 4.1.*

5. Ejemplos

Ejemplo 5.1. *Considera la matriz*

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{26}$$

Usando Sagemath podemos calcula fácilmente la SVD de A mediante el comando

```
U,Sigma,V=A.SVD()
```

obteniéndose:

$$U = \begin{pmatrix} 0{,}0 & 1{,}0 & 0{,}0 & 0{,}0 \\ -0{,}7071067811865472 & 0{,}0 & 0{,}0 & -0{,}7071067811865475 \\ -0{,}7071067811865475 & 0{,}0 & 0{,}0 & 0{,}7071067811865475 \\ 0{,}0 & 0{,}0 & 1{,}0 & 0{,}0 \end{pmatrix},$$

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

$$V = \begin{pmatrix} -0{,}0 & 1{,}0 & 0{,}0 & 0{,}0 \\ -0{,}7071067811865476 & 0{,}0 & 0{,}0 & -0{,}7071067811865475 \\ -0{,}7071067811865475 & 0{,}0 & 0{,}0 & 0{,}7071067811865476 \\ -0{,}0 & 0{,}0 & 1{,}0 & 0{,}0 \end{pmatrix},$$

Ahora bien, usando el siguiente comando:

```
(A - U*Sigma*V.transpose()).norm(p='frob')
```

obtenemos que el error de la aproximación es $\epsilon = 5{,}978733960281817e - 16$.

- *La primera aproximante daría:*

$$A_1 = \begin{pmatrix} 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\ 0{,}0 & 0{,}9999999999999997 & 0{,}9999999999999996 & 0{,}0 \\ 0{,}0 & 1{,}0 & 0{,}9999999999999998 & 0{,}0 \\ 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \end{pmatrix}$$

  con un error de $1{,}4142135623730951$.

- *La segunda aproximante daría*

$$A_2 = \begin{pmatrix} 1{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\ 0{,}0 & 0{,}9999999999999997 & 0{,}9999999999999996 & 0{,}0 \\ 0{,}0 & 1{,}0 & 0{,}9999999999999998 & 0{,}0 \\ 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \end{pmatrix}$$

  con un error de $1{,}0$.

- *La tercera aproximante daría*

$$A_3 = \begin{pmatrix} 1{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\ 0{,}0 & 0{,}9999999999999997 & 0{,}9999999999999996 & 0{,}0 \\ 0{,}0 & 1{,}0 & 0{,}9999999999999998 & 0{,}0 \\ 0{,}0 & 0{,}0 & 0{,}0 & 1{,}0 \end{pmatrix}$$

  con un error de $5{,}978733960281817e - 16$.

Ejemplo 5.2. *Resuelve de manera óptima para la norma* $|| \; || = || \; ||_2$ *el problema lineal* $A\vec{x} = \vec{b}$ *dado por:*

$$\begin{pmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{27}$$

Es de señalar que el sistema es incompatible, luego no tiene solución en el sentido estricto de la palabra. Buscaremos la solución que proporciona el método de mínimos cuadrados.

Calcularemos la descomposición SVD de A para obtener

$$V = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \ , \ \Sigma^+ = \begin{pmatrix} 1/\sqrt{2} & 0 & 0 \\ 0 & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \ , \ U = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{28}$$

Ahora consideramos la matriz $A^* = V\Sigma^+U^t$, *que resulta tener la expresión*

$$A^+ = \begin{pmatrix} 1/2 & -1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ ,$$

lo que da una solución óptima calculada mediante la fórmula $x^+ = A^+\vec{b} = (1,1,0)^t$.

Observa que $A(1,1,0)^t = (2,0,0)$, *pero este vector es el punto de la imagen de* A *que está más cerca (en la norma* $||\ ||_2$ *al dato dado* $\vec{b} = (1,1,1)^t$.

Referencias

[1] Golub,G. H., Van Loan, C., 1996. *Matrix Computations*, 3rd edition, Johns Hopkins University Press, Baltimore.

[2] Hernández, E., Vázquez, M. J., Zurro, M. A., 2012. *Álgebra Lineal y Geometría.* Pearson.

[3] Marínez de las Heras, J. J., 2005. *La descomposición en valores singulares (SVD) y algunas de sus aplicaciones*. LA GACETA DE LA RSME, Vol. 8.3 (2005), Pags. 795–810.

[4] *SageMath Documentation*, 2020, `https://doc.sagemath.org/`, `urldate 2020/11/02`.

También te pueden ser útiles los siguientes vídeos.

`https://www.youtube.com/watch?v=DG7YTlGnCEo`

`https://www.youtube.com/watch?v=9vJDjkx825k`
