Descomposición en valores singulares

Notas para los cursos 21 y 22 (J.L. Mancilla Aguilar)

1. Valores Singulares

Tanto los valores singulares como la descomposición en valores singulares de una matriz son conceptos de gran utilidad en las aplicaciones del álgebra lineal a diversos problemas prácticos y teóricos.

A continuación definiremos el concepto de valor singular de una matriz.

Dada una matriz $A \in \mathbb{R}^{m\times n}$, la matriz A^TA es simética, pues $(A^TA)^T = A^T(A^T)^T = A^TA$, y semidefinida positiva, ya que

$$x^TA^TAx = (Ax)^T(Ax) = \|Ax\|^2 \geq 0 \quad \forall x \in \mathbb{R}^n.$$

Por lo tanto los autovalores de A^TA son reales y no negativos.

Definición 1 *Sea $A \in \mathbb{R}^{m\times n}$. Sean $\lambda_1, \lambda_2, \dots, \lambda_n$ los autovalores de A^TA ordenados en forma decreciente, es decir,*

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0.$$

Entonces $\sigma_i = \sqrt{\lambda_i}$ es el i-ésimo valor singular de A.

El primer y el último valor singular de una matriz proporcionan la siguiente información.

Proposición 1 *Sea $A \in \mathbb{R}^{m\times n}$. Entonces, si σ_1 y σ_n son, respectivamente, el mayor y el menor valor singular de A, se tiene que*

$$\max_{\|x\|=1} \|Ax\| = \sigma_1 \qquad y \qquad \min_{\|x\|=1} \|Ax\| = \sigma_n.$$

Demostración. Sean λ_1 y λ_n el máximo y el mínimo autovalor de A^TA, respectivamente. Entonces, por Rayleigh,

$$\max_{\|x\|=1} x^T(A^TA)x = \sigma_1 \qquad \text{y} \qquad \min_{\|x\|=1} x^T(A^TA)x = \sigma_n.$$

Luego

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{\max_{\|x\|=1} x^T(A^TA)x} = \max_{\|x\|=1} \sqrt{\|Ax\|^2} = \max_{\|x\|=1} \|Ax\|,$$

valiendo la primera igualdad por la definición de valor singular, la segunda por Rayleigh, la tercera por la igualdad $x^T(A^TA)x = \|Ax\|^2$ y por ser la función $\eta(t) = \sqrt{t}$ monótona creciente en $[0, \infty)$ (si η es una función escalar monótona creciente sobre la imagen de $f(x)$ entonces $\eta(\max f(x)) = \max \eta(f(x))$ y $\eta(\min f(x)) = \min \eta(f(x))$).

En forma similar se prueba que $\min_{\|x\|=1} \|Ax\| = \sigma_n$. ∎

Como corolario de la Proposición 1 resultan las desigualdades

$$\sigma_n\|x\| \leq \|Ax\| \leq \sigma_1\|x\| \quad \forall x \in \mathbb{R}^n,$$

que acotan superior e inferiormente la magnitud de Ax en función de la magnitud de x.

En efecto, si $x = 0$ las desigualdades son obviamente ciertas. Si $x \neq 0$ tenemos que $v = x/\|x\|$ es unitario y por lo tanto $\sigma_n \leq \|Av\| \leq \sigma_1$. Pero $\|Av\| = \|Ax\|/\|x\|$, con lo cual

$$\sigma_n \leq \frac{\|Ax\|}{\|x\|} \leq \sigma_1 \quad \Longrightarrow \quad \sigma_n\|x\| \leq \|Ax\| \leq \sigma_1\|x\|.$$

El siguiente resultado será clave para la construcción de la descomposición en valores singulares de una matriz. Recordemos que toda matriz simétrica $n \times n$ con coeficientes reales es diagonalizable ortogonalmente, o, lo que es equivalente, existe una b.o.n. de $\mathbb{R}^n$ compuesta por autovectores de ella.

Teorema 1 *Sea $A \in \mathbb{R}^{m\times n}$. Supongamos que $\lambda_1, \lambda_2, \ldots, \lambda_n$ son los autovalores de $A^T A$ y que*

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

en otras palabras, los autovalores de $A^T A$ están ordenados en forma decreciente y el número de autovalores no nulos es r.

Sea $\{v_1, \ldots, v_n\}$ una b.o.n. de $\mathbb{R}^n$ tal que $A^T Av_i = \lambda_i v_i$.

Entonces

1. *$\{Av_1, \ldots, Av_n\}$ es un conjunto ortogonal y $\|Av_i\| = \sqrt{\lambda_i} = \sigma_i$ para todo $i = 1, \ldots, n$;*
2. *$\{\frac{Av_1}{\sigma_1}, \ldots, \frac{Av_r}{\sigma_r}\}$ es b.o.n. de $\mathrm{col}(A)$;*
3. *$\{v_{r+1}, \ldots, v_n\}$ es b.o.n. de $\mathrm{Nul}(A)$;*
4. $\mathrm{rango}(A) = r =$ *número de v.s. no nulos de A.*

Demostración. Respecto del primer punto, como

$$(Av_i, Av_j) = (Av_i)^T(Av_j) = v_i^T(A^T Av_j) = v_i^T(\lambda_j v_j) = \lambda_j(v_i^T v_j) = \begin{cases} \lambda_j & i = j \\ 0 & i \neq j \end{cases},$$

entonces $Av_i \perp Av_j$ si $i \neq j$ y $\|Av_i\| = \sqrt{\lambda_i} = \sigma_i$ para $i = 1, \ldots, n$.

Respecto del punto 2., como por el punto 1. el conjunto $\{\frac{Av_1}{\sigma_1}, \ldots, \frac{Av_r}{\sigma_r}\}$ es ortonormal, basta probar que éste genera el espacio columna de A. Para ello alcanza con ver que

$$\mathrm{gen}\{Av_1, \ldots, Av_r\} = \mathrm{col}(A).$$

Consideremos la transformación lineal $T : \mathbb{R}^n \to \mathbb{R}^m$, $T(x) = Ax$. Tenemos por un lado que $\mathrm{Im}(T) = \mathrm{col}(A)$. Por otra parte, como $\{v_1, \ldots, v_n\}$ es base de $\mathbb{R}^n$ y $T(v_i) = Av_i$,

$$\mathrm{col}(A) = \mathrm{Im}(T) = \mathrm{gen}\{Av_1, \ldots, Av_n\} = \mathrm{gen}\{Av_1, \ldots, Av_r\},$$

la última igualdad debida al punto 1., ya que $\|Av_i\| = \sqrt{\lambda_i} = 0$ si $i \geq r + 1$.

El punto 4. es inmediato del 2. ya que el rango de A es la dimensión del espacio columna de la matriz y ésta es r, que por otra parte es el número de autovalores no nulos de $A^T A$, el cual coincide con el número de v.s. no nulos de A.

Finalmente el punto 3. resulta del hecho que $\{v_{r+1}, \ldots, v_n\}$ es un conjunto ortonormal, que $\|Av_i\| = 0$ para todo $i \geq r + 1$ y que $\dim(\mathrm{Nul}(A)) = n - r$.∎

Ejemplo 1 Consideremos la matriz

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}.$$

Como

$$A^T A = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix},$$

y sus autovalores son $\lambda_1 = 360$, $\lambda_2 = 90$ y $\lambda_3 = 0$, tenemos que los v.s. de A son

$$\sigma_1 = \sqrt{360} = 6\sqrt{10}, \qquad \sigma_2 = \sqrt{90} = 3\sqrt{10} \qquad \sigma_3 = 0.$$

Notamos que el rango de A es 2 y que hay exactamente 2 v.s. no nulos como informa el Teorema 1.

Busquemos ahora una b.o.n. de $\mathbb{R}^3$ compuesta por autovectores de $A^T A$. Como

$$\mathcal{S}_{\lambda_1} = \text{gen}\left\{ \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \right\}, \qquad \mathcal{S}_{\lambda_2} = \text{gen}\left\{ \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix} \right\} \quad \text{y} \quad \mathcal{S}_{\lambda_3} = \text{gen}\left\{ \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} \right\},$$

$B = \{v_1, v_2, v_3\}$, con

$$v_1 = \frac{1}{3}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \qquad v_2 = \frac{1}{3}\begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix} \quad \text{y} \quad v_3 = \frac{1}{3}\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix},$$

es una de tales bases.

Calculemos Av_i para $i = 1, 2, 3$,

$$Av_1 = \begin{bmatrix} 18 \\ 6 \end{bmatrix}, \qquad Av_2 = \begin{bmatrix} 3 \\ -9 \end{bmatrix} \quad \text{y} \quad Av_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Notamos que $Av_i \perp Av_j$ si $i \neq j$ y que $\|Av_i\| = \sigma_i$ para $i = 1, 2, 3$.

En particular $\{v_3\}$ es b.o.n. de $\text{Nul}(A)$ y $\{\frac{Av_1}{\sigma_1}, \frac{Av_2}{\sigma_2}\}$ es b.o.n. de $\text{col}(A)$, que en este caso coincide con $\mathbb{R}^2$.

2. Descomposición en valores singulares

En lo que sigue definiremos lo que se conoce como descomposición en valores singulares (DVS) de una matriz.

Definición 2 *Sea $A \in \mathbb{R}^{m\times n}$. Una descomposición en valores singulares de A es una factorización*

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

con $U \in \mathbb{R}^{m\times m}$ y $V \in \mathbb{R}^{n\times n}$ ortogonales y $\Sigma \in \mathbb{R}^{m\times n}$ con

$$\Sigma = \begin{bmatrix} D & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \qquad y \quad D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \qquad \text{con } \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$$

$0_{k\times l}$ representa la matriz $k \times l$ cuyos coeficientes son nulos.

Notar que si $A = U\,\Sigma\,V^T$ es una DVS de A, con Σ como en la Definición 2, y v_i y u_i son las i-ésimas columnas de V y U respectivamente, entonces es fácil ver que A puede ser escrita en la forma

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T,$$

donde cada una de las matrices $u_i v_i^T$ es de rango 1.

Definición 3 Si $A = U\,\Sigma\,V^T$ es una DVS de A, a los vectores que aparecen como columnas de la matriz V se los denomina vectores singulares derechos de A mientras que a los que aparecen como columnas de U se los denomina vectores singulares izquierdos de A.

El siguiente teorema nos dice que toda matriz admite una DVS.

Teorema 2 Sea $A \in \mathbb{R}^{m\times n}$. Entonces existe un descomposición en valores singulares de A.

Demostración. Sean $\lambda_1, \ldots, \lambda_n$ los autovalores de $A^T A$. Supongamos que están ordenados en forma decreciente y que exactamente r de ellos son no nulos (r podría ser n). Sea $\{v_1, v_2, \ldots, v_n\}$ una b.o.n. de $\mathbb{R}^n$ tal que $A^T A v_i = \lambda_i v_i$ para todo $i = 1, \ldots, n$.

Definimos $V = [v_1\ v_2 \cdots v_n]$, que es ortogonal, y

$$\Sigma = \begin{bmatrix} D & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \qquad \text{con} \quad D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \quad \text{donde } \sigma_i = \sqrt{\lambda_i}.$$

Con el objeto de definir U, consideremos los vectores

$$u_1 = \frac{Av_1}{\sigma_1},\ u_2 = \frac{Av_1}{\sigma_2}, \cdots, u_r = \frac{Av_r}{\sigma_r}.$$

Entonces, de acuerdo con el Teorema 1, el conjunto $\{u_1, \ldots, u_r\}$ es ortonormal. Si $r < m$ buscamos vectores $u_{r+1}, \ldots, u_m$ de modo tal que $\{u_1, \ldots, u_m\}$ sea b.o.n. de $\mathbb{R}^m$. Luego definimos $U = [u_1\ u_2 \cdots u_m]$ que es una matriz ortogonal. Veamos ahora que $A = U\,\Sigma\,V^T$.

Por un lado

$$AV = [Av_1 \cdots Av_n] = [\sigma_1 u_1 \ldots \sigma_r u_r\ 0 \cdots 0]$$

ya que $Av_i = \sigma_i u_i$ para $i = 1, \ldots, r$ por la definición de los vectores u_i y $Av_i = 0$ para $i = r+1, \ldots, n$ por el Teorema 1.

Por otro lado

$$U\,\Sigma = [u_1\ u_2 \cdots u_m] \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} = [\sigma_1 u_1 \ldots \sigma_r u_r\ 0 \cdots 0].$$

Por lo tanto, $AV = U\Sigma$. Teniendo en cuenta que $V^{-1} = V^T$, llegamos a la igualdad $A = U\Sigma V^T$. ∎

Observación 1 La demostración del Teorema 2 nos da un método para construir una DVS de una matriz A. Método que podemos resumir de la siguiente manera.

Sea $A \in \mathbb{R}^{m\times n}$.

1. Calcular los autovalores de $A^T A$ y ordenarlos de mayor a menor: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.
2. Hallar una b.o.n. $\{v_1, \ldots, v_n\}$ de $\mathbb{R}^n$ tal que $A^T A v_i = \lambda_i v_i$, $i = 1, \ldots, n$.
3. Calcular los v.s. de A, $\sigma_i = \sqrt{\lambda_i}$.
4. Si r es el número de v.s. de A no nulos, es decir, $\sigma_r > 0$ y $\sigma_i = 0$ para todo $i \geq r+1$, definir
$$u_1 = \frac{Av_1}{\sigma_1},\ u_2 = \frac{Av_1}{\sigma_2}, \cdots, u_r = \frac{Av_r}{\sigma_r}.$$
5. Si $r < m$, hallar $u_{r+1}, \ldots, u_m$ tales que $\{u_1, \ldots, u_m\}$ sea b.o.n. de $\mathbb{R}^m$.
6. Definir las matrices $V = [v_1\ v_2 \cdots v_n]$, $U = [u_1\ u_2 \cdots u_n]$ y
$$\Sigma = \begin{bmatrix} D & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \qquad \text{con} \quad D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix};$$
7. Entonces $A = U\,\Sigma\,V^T$ es una DVS de A.

Ejemplo 2 Consideremos la matriz A del Ejemplo 1. Los autovalores de $A^T A$ son $\lambda_1 = 360$, $\lambda_2 = 90$ y $\lambda_3 = 0$ y los v.s.,
$$\sigma_1 = \sqrt{360} = 6\sqrt{10}, \qquad \sigma_2 = \sqrt{90} = 3\sqrt{10} \qquad \sigma_3 = 0.$$
$B = \{v_1, v_2, v_3\}$ con
$$v_1 = \frac{1}{3}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \qquad v_2 = \frac{1}{3}\begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix} \quad \text{y} \quad v_3 = \frac{1}{3}\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix},$$
es b.o.n. de $\mathbb{R}^3$ y $A^T A v_i = \lambda_i v_i$. Como hay dos v.s. no nulos, calculamos
$$u_1 = \frac{Av_1}{\sigma_1} = \begin{bmatrix} \frac{3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{bmatrix} \qquad \text{y} \qquad u_2 = \frac{Av_2}{\sigma_2} = \begin{bmatrix} \frac{1}{\sqrt{10}} \\ -\frac{3}{\sqrt{10}} \end{bmatrix}.$$
Como $\{u_1, u_2\}$ es b.o.n. de $\mathbb{R}^2$, no hace falta completar el conjunto. Ahora definimos
$$V = \frac{1}{3}\begin{bmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{bmatrix}, \quad U = \frac{1}{\sqrt{10}}\begin{bmatrix} 3 & 1 \\ 1 & -3 \end{bmatrix} \quad \text{y} \quad \Sigma = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix},$$
y tenemos que $A = U\,\Sigma\,V^T$ es una DVS de A.

Notamos que en la construcción de la DVS de A que hicimos en la demostración del Teorema 2, los elementos no nulos que aparecen en la diagonal principal de Σ son los v.s. no nulos de A y que las columnas de V, es decir, los vectores singulares derechos que utilizamos, son autovectores de $A^T A$. El siguiente resultado dice que lo anterior siempre ocurre, independientemente de la forma en que la DVS haya sido obtenida.

Teorema 3 Sea $A \in \mathbb{R}^{m\times n}$ y sea $A = U\,\Sigma\,V^T$ una DVS de A.

Supongamos que

$$\Sigma = \begin{bmatrix} D & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \quad \text{y} \quad D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \quad \text{con } \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$$

Entonces

1. $\sigma_1, \ldots, \sigma_r$ son los v.s. no nulos de A;
2. Si $V = [v_1 \cdots v_n]$, $\{v_1, \ldots, v_n\}$ es una b.o.n. de $\mathbb{R}^n$ tal que $A^T A v_i = \sigma_i^2 v_i$ si $i = 1, \ldots, r$ y $A^T A v_i = 0$ si $i = r+1, \ldots, n$;
3. Si $U = [u_1 \cdots u_m]$, $A v_i = \sigma_i u_i$ para $i = 1, \ldots, r$ y $A v_i = 0$ para $i = r+1, \ldots, n$.

Demostración. Como $A = U\,\Sigma\,V^T$ tenemos que

$$A^T A = (U\,\Sigma\,V^T)^T (U\,\Sigma\,V^T) = V(\Sigma^T \Sigma) V^T.$$

Pero $\Sigma^T \Sigma$ es la matriz $n \times n$

$$\Sigma^T\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Por lo tanto $A = V(\Sigma^T \Sigma)V^T$ es una diagonalización ortogonal de $A^T A$. En consecuencia, $\lambda_1 = \sigma_1^2, \lambda_2 = \sigma_2^2, \ldots, \lambda_r = \sigma_r^2$, son los autovalores no nulos de $A^T A$ ordenados en forma decreciente. Por lo tanto $\sigma_1, \ldots, \sigma_r$ son los v.s. no nulos de A.

También tenemos que las columnas de V conforman una b.o.n de $\mathbb{R}^n$ y son autovectores de $A^T A$, más precisamente, $A^T A v_i = \sigma_i^2 v_i$ para $i = 1, \ldots r$, y $A^T A v_i = 0$ para $i = r+1, \ldots, n$, ya que $v_{r+1}, \ldots, v_n$ son autovectores asociados al autovalor nulo.

Finalmente el punto 3. resulta inmediatamente de la igualdad $A\,V = U\,\Sigma$. ∎

Notemos que del Teorema 3 se desprende que los vectores singulares derechos son siempre autovectores de $A^T A$. En lo que sigue veremos que los vectores singulares izquierdos son necesariamente autovectores de la matriz AA^T.

Para ello es útil el siguiente resultado.

Proposición 2 Sea $A = U\,\Sigma\,V^T$ una DVS de $A \in \mathbb{R}^{m\times n}$. Entonces

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

es DVS de A^T. En particular A y A^T tienen los mismos v.s. no nulos.

Demostración. Es inmediato a partir de la definición de DVS que $A^T = V\,\Sigma^T\,U^T$ es DVS de A^T. Que A y A^T tienen los mismos v.s. no nulos se deduce del hecho que en Σ aparecen los v.s. no nulos de A y en Σ^T los v.s. no nulos de A^T.∎

De la Proposición 2 y del Teorema 3 deducimos que los vectores singulares izquierdos de una matriz A son a la vez vectores singulares derechos de A^T y por lo tanto autovectores de AA^T.

Ejemplo 3 Hallar una DVS de

$$A = \begin{bmatrix} 1 & -2 & 2 \\ -1 & 2 & -2 \end{bmatrix}.$$

Como los v.s. no nulos de A y de A^T coinciden, calculamos los v.s. de A^T, que son los autovalores de AA^T. Como

$$AA^T = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix},$$

sus autovalores son $\lambda_1 = 18$ y $\lambda_2 = 0$. Luego $\sigma_1 = \sqrt{18} = 3\sqrt{2}$ es el único v.s. no nulo tanto de A^T como de A. En particular, los restantes v.s. de A son $\sigma_2 = \sigma_3 = 0$.

Para hallar una DVS de A, lo que hacemos primero es hallar una DVS de A^T. Calculando obtenemos la siguiente b.o.n de $\mathbb{R}^2$ compuesta por autovectores de AA^T (ordenados según el orden de los v.s.) $\left\{[\frac{1}{\sqrt{2}} \;\; -\frac{1}{\sqrt{2}}]^T, [\frac{1}{\sqrt{2}} \;\; \frac{1}{\sqrt{2}}]^T\right\}$. Designemos por u_1 y u_2 a los vectores de esa base. Ambos son vectores singulares derechos de A^T (u_1 corresponde a σ_1 y u_2 a $\sigma_2 = 0$) y por lo tanto vectores singulares izquierdos de A.

Para obtener los vectores singulares izquierdos de A^T que nos permitan luego construir una DVS de A^T, teniendo en cuenta que hay un solo valor singular no nulo, definimos

$$v_1 = \frac{A^T u_1}{\sigma_1} = \frac{1}{3}\begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}.$$

Para hallar los restantes vectores singulares izquierdos de A^T debemos encontrar v_2, v_3 tales que $\{v_1, v_2, v_3\}$ sea b.o.n. de $\mathbb{R}^3$. Pero como v_2 y v_3 son a su vez vectores singulares derechos de A y corresponden a valores singulares nulos de A, necesariamente $Av_2 = 0$ y $Av_3 = 0$. Luego bastará con que hallemos una b.o.n. de $\mathrm{Nul}(A)$. Una posible b.o.n. es $\{v_2, v_3\}$ con

$$v_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad \text{y} \quad v_3 = \frac{1}{3\sqrt{2}}\begin{bmatrix} 4 \\ 1 \\ -1 \end{bmatrix}.$$

Note que $\{v_1, v_2, v_3\}$ es b.o.n. de $\mathbb{R}^3$. (Otra forma de computar v_2, v_3 consiste en elegir v_2^* y v_3^* tales que $\{v_1, v_2^*, v_3^*\}$ sea l.i. y ortonormalizar el conjunto mediante Gram-Schmidt.)

Entonces

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

con $V = [v_1\ v_2\ v_3]$, $U = [u_1\ u_2]$ y

$$\Sigma^* = \begin{bmatrix} 2\sqrt{3} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

es DVS de A^T y, por lo tanto, $A = U\,\Sigma\,V^T$ con $\Sigma = \Sigma^{*T}$ es DVS de A.

Como ejercicio, se propone al lector que calcule una DVS de A en forma directa.

3. DVS y los subespacios fundamentales de una matriz

En esta sección veremos que las matrices U y V de una DVS de A están compuestas por bases ortonormales de los cuatro subespacios fundamentales de A.

Supongamos que $A \in \mathbb{R}^{n\times m}$ es una matriz de rango r y que $A = U\,\Sigma\,V^T$ es una DVS de A. Entonces, si $V = [v_1 \cdots v_n]$ y $U = [u_1 \cdots u_m]$ y definimos las matrices $V_r = [v_1 \cdots v_r]$, $V_{n-r} = [v_{r+1} \cdots v_n]$, $U_r = [u_1 \cdots u_r]$ y $U_{m-r} = [u_{r+1} \cdots u_m]$, se tiene que

1. $\{v_1, \cdots, v_r\}$ es b.o.n. de fil(A), $V_r^T V_r = I$ y $V_r V_r^T = P_{\mathrm{fil}(A)}$.
2. $\{v_{r+1}, \cdots, v_n\}$ es b.o.n. de Nul(A), $V_{n-r}^T V_{n-r} = I$ y $V_{n-r} V_{n-r}^T = P_{\mathrm{Nul}(A)}$.
3. $\{u_1, \cdots, u_r\}$ es b.o.n. de col(A), $U_r^T U_r = I$ y $U_r U_r^T = P_{\mathrm{col}(A)}$.
4. $\{u_{r+1}, \cdots, u_m\}$ es b.o.n. de Nul(A^T), $U_{m-r}^T U_{m-r} = I$ y $U_{m-r} U_{m-r}^T = P_{\mathrm{Nul}(A^T)}$.

La justificación de este resultado es la siguiente. Como el rango de A es r, A tiene r valores singulares no nulos, y por lo tanto A^TA tiene r autovalores no nulos. Por el Teorema 3, los vectores $v_1, \ldots, v_n$ son autovectores de A^TA, estando los primeros r de ellos asociados a los autovalores no nulos de A^TA y los restantes al autovalor nulo de A^TA. Por lo tanto, por el Teorema 1, $\{v_{r+1}, \cdots, v_n\}$ es b.o.n. de Nul(A) y, dado que $\{v_1, \cdots, v_n\}$ es b.o.n. de $\mathbb{R}^n$, necesariamente $\{v_1, \cdots, v_r\}$ es b.o.n. de Nul$(A)^\perp =$ fil(A). Con esto, y la forma en que se construyen las matrices de proyección, quedan demostrados 1. y 2.

Respecto de 3. y 4., combinando el Teorema 1 con el Teorema 3, tenemos que $\{u_1, \cdots, u_r\}$ es b.o.n. de col(A) y, por lo tanto, dado que $\{u_1, \cdots, u_m\}$ es b.o.n. de $\mathbb{R}^m$, $\{u_{r+1}, \cdots, u_m\}$ es b.o.n. de col$(A)^\perp =$ Nul(A^T). De esto último y la forma en que se construyen las matrices de proyección, se deducen las restantes afirmaciones.

Ejemplo 4 Dada

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix},$$

hallar los valores singulares de A, bases de sus cuatro subespacios fundamentales y calcular sus matrices de proyección.

La descomposición que tenemos de A es "casi" una DVS, salvo por el hecho de que la matriz U^* que aparece a la izquierda, si bien tiene columnas mutuamente ortogonales, éstas no son de norma 1. Procedemos entonces a normalizar las columnas de U^* dividiendo cada una de ellas por su norma. Para que el producto siga siendo A, es necesario multiplicar la fila i de la matriz central, por el número por el cual dividimos la columna i de U^*. Queda entonces la factorización

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Esta factorización no es aún una DVS, porque los elementos de la diagonal de la matriz central no están ordenados de mayor a menor. Para ello lo que hacemos es permutar las dos primeras columnas de la matriz de la izquierda y las dos primeras filas de la matriz de la derecha. Entonces obtenemos

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix},$$

que ahora sí es una DVS de A, ya que $A = U\,\Sigma\,V^T$ con

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{y} \quad V = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Ahora podemos hallar lo solicitado. Los valores singulares de A son $\sigma_1 = 6, \sigma_2 = 2$ y $\sigma_3 = 0$. Las primeras dos columnas de V son una b.o.n. de fil(A), mientras que la última columna de V es b.o.n. de Nul(A). Respecto de col(A), las dos primeras columnas de U son una b.o.n. de ese subespacio, y la última columna de U es b.o.n. de Nul(A^T). Respecto de las matrices de proyección, éstas son:

$$P_{\text{fil}(A)} = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} \\ 1 & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix},$$

$$P_{\text{Nul}(A)} = I - P_{\text{fil}(A)} = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix},$$

$$P_{\text{Nul}(A^T)} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$P_{\text{col}(A)} = I - P_{\text{Nul}(A^T)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

4. DVS reducida

Veremos ahora cómo a partir de una DVS, $A = U\Sigma V^T$, podemos obtener una descomposición de A que emplea matrices de tamaño “reducido”. Empleando la notación de la sección anterior, escribimos $U = [U_r\ U_{m-r}]$ y $V = [V_r\ V_{n-r}]$, donde r es el rango de A.

Entonces

$$A = U\,\Sigma\,V^T = [U_r\ U_{m-r}] \begin{bmatrix} D & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix} = U_r D V_r^T.$$

A la factorización $A = U_r D V_r^T$ se la denomina *DVS reducida de A*. Notar que la matriz D es inversible, pues D es diagonal, y los elementos que aparecen en la diagonal principal son los v.s. no nulos de A.

Ejemplo 5 Para la matriz A del Ejemplo 4,

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix},$$

es una DVS reducida.

Notamos que es más simple calcular una DVS reducida que una DVS, ya que solo se necesitan los v.s. no nulos y conjuntos de vectores singulares derechos e izquierdos correspondientes a esos valores singulares.

5. Solución por cuadrados mínimos de norma mínima. Seudoinversa de Moore-Penrose

Si $A \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^n$, la ecuación $Ax = b$ siempre admite soluciones por cuadrados mínimos (c.m.). Si $\mathrm{Nul}(A) = \{0\}$, o, equivalentemente, rango(A) $= n$, la solución por c.m. es única; en caso contrario hay infinitas soluciones por c.m., más aún, si $\hat{x}_p$ es una solución por c.m., entonces todas las demás son de la forma: $\hat{x} = \hat{x}_p + x_n$ con $x_n \in \mathrm{Nul}(A)$.

Es importante en el caso en que hay infinitas soluciones por c.m. disponer de un criterio que permita seleccionar una de estas infinitas soluciones. Un posible criterio es el siguiente.

Definición 4 *x^* es una solución por c.m. de norma mínima de la ecuación $Ax = b$ si x^* es solución por c.m. y, además, $\|x^*\| \leq \|\hat{x}\|$ para todo $\hat{x}$ que sea solución por c.m. de $Ax = b$.*

En otras palabras, x^* es, de todas las soluciones por c.m. de $Ax = b$, una que tiene la menor norma (longitud) posible.

A continuación veremos que existe una única solución por c.m. de norma mínima y daremos una caracterización de ella que será útil para calcularla.

Proposición 3 *Consideremos la ecuación $Ax = b$ con $A \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^n$. Entonces*

1. *Existe una única solución x^* por c.m. de la ecuación $Ax = b$ que pertenece a $\mathrm{fil}(A)$.*

2. *x^* es la única solución por c.m. de norma mínima de la ecuación $Ax = b$.*

Demostración. Primero probamos que existe una solución por c.m. que pertenece a $\text{fil}(A)$. Sea $\hat{x}_p$ una solución por c.m. de la ecuación $Ax = b$. Sea $x^* = P_{\text{fil}(A)}\hat{x}_p$. Veamos que x^* también es solución por c.m. de la ecuación. Como $\text{fil}(A) = \text{Nul}(A)^\perp$,

$$\hat{x}_p = P_{\text{fil}(A)}\hat{x}_p + P_{\text{Nul}(A)}\hat{x}_p = x^* + P_{\text{Nul}(A)}\hat{x}_p.$$

Luego, dado que $\hat{x}_p$ una solución por c.m. de la ecuación $Ax = b$,

$$P_{\text{col}(A)}b = A\hat{x}_p = A(x^* + P_{\text{Nul}(A)}\hat{x}_p) = Ax^*,$$

y por lo tanto x^* es solución por c.m. de la ecuación $Ax = b$ y pertenece a $\text{fil}(A)$.

Ahora veamos que x^* es la única solución por c.m. que pertenece a $\text{fil}(A)$. Supongamos que x' es una solución por c.m. que pertenece a $\text{fil}(A)$. Entonces $Ax^* = Ax' = P_{\text{col}(A)}b$. Luego $x^* - x' \in \text{Nul}(A)$. Como $x^* - x' \in \text{fil}(A)$ y $\text{fil}(A) = \text{Nul}(A)^\perp$, resulta $x^* - x' = 0$, y, por lo tanto, $x^* = x'$.

Hasta aquí hemos probado el punto 1. de la proposición. Ahora probamos que x^* es la única solución por c.m. de norma mínima de la ecuación $Ax = b$.

Primero vemos que x^* es solución por c.m. de norma mínima. Sea $\tilde{x}$ una solución por c.m. de la ecuación. Entonces $\tilde{x} - x^* \in \text{Nul}(A)$. Luego

$$\|\tilde{x}\|^2 = \|x^* + (\tilde{x} - x^*)\|^2 = \|x^*\|^2 + \|\tilde{x} - x^*\|^2$$

pues $\text{fil}(A) = \text{Nul}(A)^\perp$. Entonces $\|\tilde{x}\|^2 \geq \|x^*\|^2$ y x^* es de norma mínima.

Finalizamos la demostración viendo que x^* es la única solución de norma mínima. Supongamos que y también es solución de norma mínima. Entonces necesariamente $\|y\| = \|x^*\|$. Por otro lado, dado que $y - x^* \in \text{Nul}(A)$ y que $x^* \in \text{fil}(A)$,

$$\|y\|^2 = \|x^* + (y - x^*)\|^2 = \|x^*\|^2 + \|y - x^*\|^2.$$

Entonces, necesariamente $\|y - x^*\| = 0$, y, por lo tanto, $y = x^*$. ∎

De acuerdo con la Proposición 3, para hallar la solución por c.m. de norma mínima, debemos buscar entre las soluciones por c.m. de la ecuación la solución x^* que pertenece a $\text{fil}(A)$. Para hallar tal solución podemos proceder como sigue.

Supongamos que $A = U_r\, D\, V_r^T$ es una DVS reducida de A. Entonces, por lo expuesto en la Sección 3, $U_r U_r^T = P_{\text{col}(A)}$, $U_r^T U_r = I$, las columnas de V_r son b.o.n. de $\text{fil}(A)$ y D es inversible.

Luego, como x^* pertenece a $\text{fil}(A)$, existe $\alpha \in \mathbb{R}^r$ tal que $x^* = V_r\alpha$. Como además x^* es solución por c.m. de $Ax = b$ entonces

$$P_{\text{col}(A)}b = Ax^* = AV_r\alpha.$$

Luego, usando que $U_r U_r^T = P_{\text{col}(A)}$, $V_r^T V_r = I$ y $A = U_r\, D\, V_r^T$, tenemos que

$$U_r U_r^T b = U_r\, D\, V_r^T V_r \alpha = U_r D\alpha.$$

Multiplicando por izquierda por U_r y usando que $U_r^T U_r = I$, obtenemos

$$D\alpha = U_r^T b \quad \Longrightarrow \quad \alpha = D^{-1}U_r^T b,$$

y, por lo tanto,

$$x^* = V_r\alpha = V_r D^{-1} U_r^T b.$$

Notemos que hemos probado que la solución por c.m. de norma mínima de la ecuación $Ax = b$ se obtiene multiplicando b por la matriz $V_r D^{-1} U_r^T$.

Definición 5 Sea $A = U_r\, D\, V_r^T$ es una DVS reducida de A. La matriz $A^+ = V_r D^{-1} U_r^T$ es la matriz seudoinversa de Moore-Penrose de A.

Notamos que del hecho que para todo $b \in \mathbb{R}^n$, $x^* = A^+ b$ es la única solución por c.m. de longitud mínima de la ecuación $Ax = b$, se deduce que A^+ no depende de la DVS reducida empleada para calcularla. En efecto, si $A = U'_r D {V'_r}^T$ es otra DVS reducida de A, tenemos la igualdad

$$(V_r D^{-1} U_r^T) b = x^* = (V'_r D^{-1} {U'_r}^T) b \quad \forall b \in \mathbb{R}^n.$$

Pero entonces, necesariamente

$$V_r D^{-1} U_r^T = V'_r D^{-1} {U'_r}^T.$$

Teorema 4 *Sea $A \in \mathbb{R}^{m\times n}$ y sea A^+ la matriz seudoinversa de Moore-Penrose de A. Entonces*

1. *Para todo $b \in \mathbb{R}^n$, $x^* = A^+ b$ es la única solución por c.m. de longitud mínima de la ecuación $Ax = b$.*
2. *$A^+ A = P_{\mathrm{fil}(A)}$ y $AA^+ = P_{\mathrm{col}(A)}$*

Demostración. El punto 1. ya fue probado, el punto 2. se deduce inmediatamente del hecho que $A^+ A = V_r V_r^T$ y que $AA^+ = U_r U_r^T$. ∎

Observación 2 Del punto 2. del Teorema 4 deducimos inmediatamente que

- $A^+ A = I \iff \mathrm{fil}(A) = \mathbb{R}^n \iff \mathrm{rango}(A) = n$.
- $AA^+ = I \iff \mathrm{col}(A) = \mathbb{R}^m \iff \mathrm{rango}(A) = m$.
- Si A es inversible, entonces $A^{-1} = A^+$.

Cuando $\mathrm{rango}(A) = n$, la ecuación $Ax = b$ tiene una única solución por cuadrados mínimos $\hat{x}$, que está dada por la fórmula $\hat{x} = A^\sharp b$, con $A^\sharp = (A^T A)^{-1} A^T$. Como en el caso en que la solución por c.m. es única, ésta es necesariamente la de norma mínima, también tenemos que $\hat{x} = A^+ b$. Luego

$$A^\sharp b = A^+ b \quad \forall b \in \mathbb{R}^n \qquad \Longrightarrow \qquad A^\sharp = A^+.$$

La siguiente expresión de A^+ a partir de una DVS de A es útil en algunas circunstancias. Supongamos que $A = U\, \Sigma\, V^T$ es una DVS de $A \in \mathbb{R}^{m\times n}$, y que A es de rango r. Definamos

$$\Sigma^+ = \begin{bmatrix} D^{-1} & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{bmatrix} \in \mathbb{R}^{n\times m}.$$

Entonces $A^+ = V\, \Sigma^+\, U^T$, como puede comprobar fácilmente el lector, usando las expresiones $V = [V_r\ V_{n-r}^T]$ y $U = [U_r\ U_{m-r}^T]$ y efectuando el producto.

Ejemplo 6 Hallar la seudoinversa de Moore-Penrose de

$$A = \begin{bmatrix} 2 & 4 \\ 6 & 12 \end{bmatrix}.$$

Notemos que el rango de A es 1, y por lo tanto A carece de inversa. Vamos a buscar entonces una DVS reducida de A.

Como

$$A^TA = \begin{bmatrix} 40 & 80 \\ 80 & 160 \end{bmatrix},$$

sus autovalores son $\lambda_1 = 200$ y $\lambda_2 = 0$. Luego, A tiene un único v.s. no nulo, $\sigma_1 = \sqrt{200} = 10\sqrt{2}$. Para construir una DVS reducida debemos hallar un vector singular derecho v_1 de A asociado a σ_1, o, lo que es lo mismo, un autovector unitario de A^TA asociado a λ_1, por ejemplo, $v_1 = \frac{1}{\sqrt{5}}[1\ \ 2]^T$. Ahora, a partir de v_1 definimos el vector singular izquierdo

$$u_1 = \frac{Av_1}{\sigma_1} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

Entonces

$$A = \begin{bmatrix} \frac{1}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} \end{bmatrix} [10\sqrt{2}] \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}$$

es una DVS reducida de A y

$$A^+ = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{1}{10\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix} = \frac{1}{100} \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}.$$

6. Imagen de la esfera unitaria

En esta sección veremos cómo una DVS de una matriz $A \in \mathbb{R}^{m\times n}$ nos permite determinar fácilmente cual es la imagen de la esfera unitaria $S^{n-1} = \{x \in \mathbb{R}^n :\ \|x\| = 1\}$ a través de la transformación lineal $T(x) = Ax$.

Supongamos que $A = U\ \Sigma\ V^T$ es una DVS de A y que rango$(A) = r$. Con el objeto de determinar qué clase de conjunto es

$$T(S^{n-1}) = \{z \in \mathbb{R}^m :\ z = Ax \text{ con } x \in S^{n-1}\},$$

consideremos el cambio de variable $x = Vy$. Notamos que $\|x\| = \|y\|$ por ser V ortogonal.

Entonces

$$z \in T(S^{n-1}) \iff z = Ax \text{ con } \|x\| = 1 \iff z = AVy \ \text{ con } \ \|y\| = 1 \iff z = U\Sigma y \ \text{ con } \ \|y\| = 1.$$

Llamando $U = [u_1 \cdots u_m]$ y teniendo en cuenta que A tiene r v.s. no nulos $\sigma_1, \ldots, \sigma_r$, tenemos que $U\Sigma y = \sigma_1 y_1 u_1 + \sigma_2 y_2 u_2 + \cdots + \sigma_r y_r u_r$.

Entonces

$$z \in T(S^{n-1}) \iff z = \sigma_1 y_1 u_1 + \sigma_2 y_2 u_2 + \cdots + \sigma_r y_r u_r \quad \text{con} \quad \|y\| = 1.$$

Si consideramos la base ortonormal $B = \{u_1, \ldots, u_m\}$ de $\mathbb{R}^m$, y $[z]_B = [w_1 \cdots w_m]^T$ es el vector de coordenadas de z en la base B, tenemos entonces que

$$z \in T(S^{n-1}) \iff \left\{ \begin{array}{ccc} w_1 & = & \sigma_1 y_1 \\ w_2 & = & \sigma_2 y_2 \\ \vdots & \vdots & \vdots \\ w_r & = & \sigma_r y_r \\ w_{r+1} & = & 0 \\ \vdots & \vdots & \vdots \\ w_m & = & 0 \end{array} \right. \quad \text{con} \quad \|y\| = 1.$$

Como $y_i = w_i/\sigma_i$ para $i = 1, \ldots, r$ y $\sum_{i=1}^{n} y_i^2 = 1$, finalmente llegamos a la siguiente conclusión:

1. Si $r = n$

$$z \in T(S^{n-1}) \iff \frac{w_1^2}{\sigma_1^2} + \cdots + \frac{w_n^2}{\sigma_n^2} = 1 \quad \wedge \quad w_{n+1} = w_{n+2} = \cdots = w_m = 0.$$

Con lo cual $T(S^{n-1})$ resulta ser un elipsoide n-dimensional (Si $n = 1$ es un par de puntos y cuando $n = 2$ es una elipse) contenido en el subespacio generado por $\{u_1, \ldots, u_n\}$, que es col(A), y tiene por ejes a las rectas generadas por los vectores $u_1, u_2, \ldots, u_n$.

2. Si $r < n$,

$$\frac{w_1^2}{\sigma_1^2} + \cdots + \frac{w_r^2}{\sigma_r^2} = \sum_{i=1}^{r} y_i^2 \leq 1,$$

y, por lo tanto,

$$z \in T(S^{n-1}) \iff \frac{w_1^2}{\sigma_1^2} + \cdots + \frac{w_r^2}{\sigma_r^2} \leq 1 \quad \wedge \quad w_{r+1} = w_{r+2} = \cdots = w_m = 0.$$

Luego $T(S^{n-1})$ resulta ser un elipsoide r-dimensional sólido (si $r = 1$ es un segmento, si $r = 2$ es una elipse junto con su interior), contenido en el subespacio generado por $\{u_1, \ldots, u_n\}$ (col(A)), y tiene por ejes a las rectas generadas por los vectores $u_1, u_2, \ldots, u_r$.

Ejemplo 7 Supongamos que $A \in \mathbb{R}^{2\times 2}$ y que $A = U\Sigma V^T$ es DVS de A. Sean u_1 y u_2 la primera y segunda columna de U, y σ_1 y σ_2 el primer y segundo v.s. de A. Entonces, la imagen de la circunsferencia unitaria $S^1 \subset \mathbb{R}^2$ a través de la transformación $T(x) = Ax$ depende del rango de A de la siguiente forma:

1. si rango$(A) = 0$, A es la matriz nula y $T(S^1) = \{0\}$.

2. Si rango$(A) = 1$, $\sigma_1 > 0$, $\sigma_2 = 0$ y

$$T(S^1) = \left\{ z \in \mathbb{R}^2 : \ z = w_1 u_1 + w_2 u_2, \ \frac{w_1^2}{\sigma_1^2} \leq 1, \ w_2 = 0 \right\} = \left\{ z \in \mathbb{R}^2 : \ z = t u_1, \ -\sigma_1 \leq t \leq \sigma_1 \right\},$$

es el segmento de extremos $P_1 = -\sigma_1 u_1$, $P_2 = \sigma u_1$.

3. Si rango$(A) = 2$, $\sigma_1 \geq \sigma_2 > 0$, y

$$T(S^1) = \left\{ z \in \mathbb{R}^2 : \ z = w_1 u_1 + w_2 u_2, \ \frac{w_1^2}{\sigma_1^2} + \frac{w_2^2}{\sigma_2^2} = 1 \right\},$$

resulta una elipse con centro en el origen, eje mayor de longitud $2\sigma_1$ contenido en la recta generada por u_1 y eje menor de longitud $2\sigma_2$ contenido en la recta u_2 (Figura 1).

Ejemplo 8 Supongamos ahora que $A \in \mathbb{R}^{3\times 3}$ y que $A = U\Sigma V^T$ es DVS de A. Sea $U = [u_1 \ u_2 \ u_3]$ y $\sigma_1, \sigma_2, \sigma_3$ los v.s. de A. Entonces tenemos las siguientes posibilidades para la imagen de la esfera unitaria $S^2 \subset \mathbb{R}^3$ a través de la transformación $T(x) = Ax$:

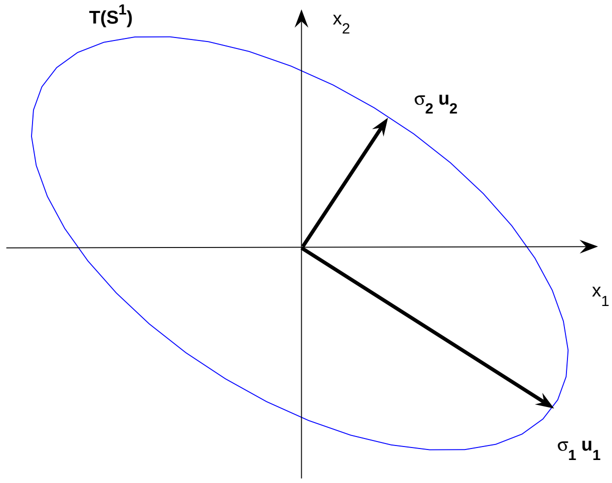

Figura 1: $T(S^1)$ caso rango 2

1. si rango$(A) = 0$, A es la matriz nula y $T(S^2) = \{0\}$.
2. Si rango$(A) = 1$, $\sigma_1 > 0$, $\sigma_2 = \sigma_3 = 0$ y la imagen de es el conjunto

$$\begin{aligned} T(S^2) &= \left\{ z \in \mathbb{R}^3 : \; z = w_1u_1 + w_2u_2 + w_3u_3, \; \frac{w_1^2}{\sigma_1^2} \le 1, \; w_2 = w_3 = 0 \right\} \\ &= \left\{ z \in \mathbb{R}^3 : \; z = tu_1, \; -\sigma_1 \le t \le \sigma_1 \right\}, \end{aligned}$$

que es el segmento de extremos $P_1 = -\sigma_1 u_1$, $P_2 = \sigma u_1$.

3. Si rango$(A) = 2$, $\sigma_1 \ge \sigma_2 > 0$, $\sigma_3 = 0$ y

$$T(S^2) = \left\{ z \in \mathbb{R}^3 : \; z = w_1u_1 + w_2u_2 + w_3u_3, \; \frac{w_1^2}{\sigma_1^2} + \frac{w_2^2}{\sigma_2^2} \le 1; \; w_3 = 0 \right\},$$

es la superficie contenida en el plano generado por u_1 y u_2 (que es col(A)), que contiene al origen y está limitada por la elipse centrada en el origen, cuyo eje mayor tiene longitud $2\sigma_1$ y está contenido en la recta generada por u_1 y su eje menor, contenido en la recta u_2, tiene longitud $2\sigma_2$ (Figura 2.)

4. Finalmente, si rango$(A) = 3$, $\sigma_1 \ge \sigma_2 \ge \sigma_3 > 0$ y

$$T(S^2) = \left\{ z \in \mathbb{R}^3 : \; z = w_1u_1 + w_2u_2 + w_3u_3, \; \frac{w_1^2}{\sigma_1^2} + \frac{w_2^2}{\sigma_2^2} + \frac{w_3^2}{\sigma_3^2} = 1 \right\},$$

resulta un elipsoide centrado en el origen, con ejes de longitudes $2\sigma_1$, $2\sigma_2$ y $2\sigma_3$ contenidos en la rectas generadas por u_1, u_2 y u_3, respectivamente (Figura 3.).

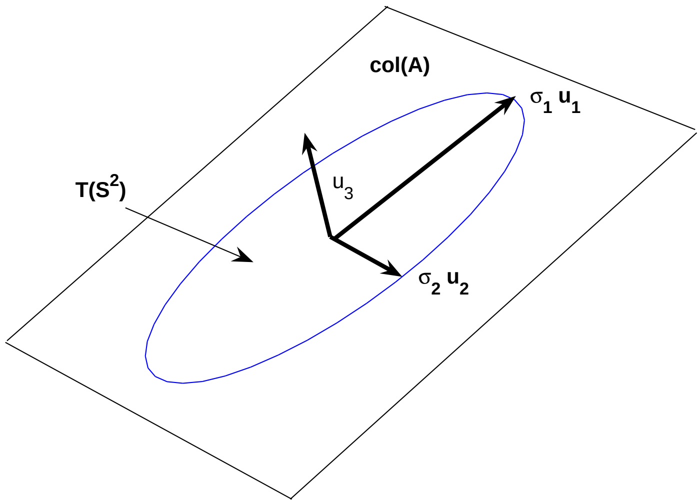

Figura 2: $T(S^2)$ caso rango 2

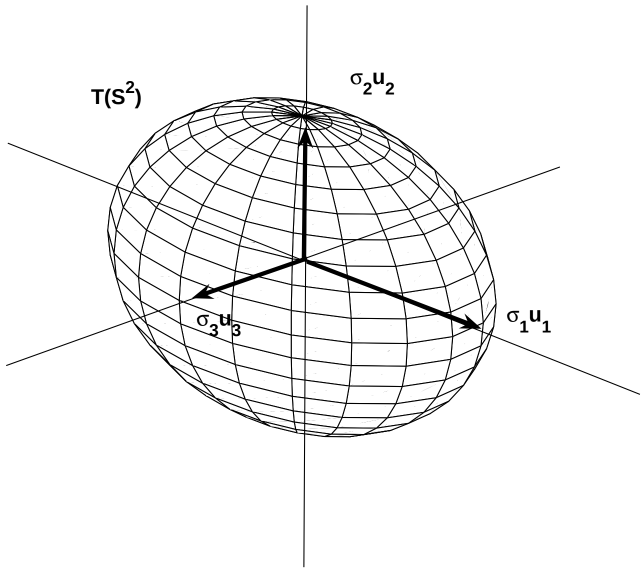

Figura 3: $T(S^2)$ caso rango 3