Descomposición en valores singulares de una matriz

Estas notas están dedicadas a demostrar una extensión del teorema espectral conocida como descomposición en valores singulares (SVD en inglés) de gran utilidad en múltiples aplicaciones en estadística (análisis de componente principal), teoría de control, compresión de imágenes, etc.
El teorema espectral establece que cualquier matriz simétrica $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ puede descomponerse como $A = PDP^t$ donde D es diagonal y P ortogonal. Aquí mostraremos que una matriz cualquiera $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ (no necesariamente cuadrada) puede descomponerse como $A = USV^t$ donde S es diagonal y U y V son ortogonales (no necesariamente una inversa de la otra). En lo que sigue V es un espacio vectorial con producto interno sobre un cuerpo $\mathbb{K}$

Definición 1 *Un operador autoadjunto $T : V \to V$ se dice no negativo si*

$$\langle T(v), v\rangle \geq 0, \quad \forall v \in V$$

Si T es no negativo entonces resulta fácil probar que sus valores propios son todos no negativos, en efecto si λ es valor propio de T con vector propio $v \neq \vec{0}$ entonces

$$0 \leq \langle T(v), v\rangle = \lambda\langle v, v\rangle$$

y como en el último término el segundo factor es positivo se deduce que $\lambda \geq 0$.

Definición 2 *Un operador autoadjunto $T : V \to V$ se dice positivo si*

$$\langle T(v), v\rangle > 0, \quad \forall v \in V$$

Y en este caso los valores propios son positivos.

El siguiente lema útil en la demostración ha sido usado también en otras partes de curso.

Lema 1 *Sean V y V espacios vectoriales con producto interno de dimensión finita y sea $T : V \to W$ una transformación lineal entonces T^*T y TT^* son ambos autoadjuntos, no negativos y tienen el mismo rango que T.*

Demostración. Probaremos el resultado solo para T^*T, la prueba para TT^* es análoga y queda a cargo del lector. Veamos primero que T^*T es autoadjunto,

$$(T^*T)^* = T^*(T^*)^* = T^*T.$$

T^*T es no negativo, en efecto,

$$\langle T^*T(v), v\rangle = \langle T(v), T(v)\rangle \geq 0, \quad \forall v \in V.$$

Finalmente observemos que *rango*$(T^*T) =$ *rango*(T). Para esto alcanza con probar que $\ker(T) = \ker(T^*T)$:

Es inmediato verificar que,

$$\ker(T) \subset \ker(T^*T).$$

Además, si $v \in \ker(T^*T)$ entonces

$$0 = \langle T^*T(v), v\rangle = \langle T(v), T(v)\rangle,$$

por lo tanto $T(v) = \vec{0}$ y consecuentemente $v \in \ker(T)$. Como $\ker(T^*T) = \ker(T)$ se deduce inmediatamente del teorema de las dimensiones que

$$\dim(Im(T^*T)) = \dim(Im(T))$$

lo cual concluye la prueba. ■

Teorema 2 *Sean V y W espacios vectoriales con $dim(V) = n$ y $dim(W) = m$ y sea $T : V \to W$ una transformación lineal con* $\text{rango}(T) = r$. *Entonces existe $A = \{v_1, \ldots, v_n\}$ base ortonormal de V, $B = \{w_1, \ldots, w_m\}$ base ortonormal de W y escalares $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ tales que $T(v_i) = \sigma_i w_i$ si $i = 1, \ldots, r$ y $T(v_i) = \vec{0}$ si $i = r+1, \ldots, n$. Es decir:*

$$T(v_i) = \sigma_i w_i \quad con \quad \sigma_i > 0\ \forall i = 1, \ldots, r \quad y \quad \sigma_i = 0\ \forall i = r+1, \ldots, n$$

Demostración. Sea $R = T^*T$, por el lema anterior se sabe que R es un operador lineal en V autoadjunto, no negativo y tiene rango r (pues $r = \text{rango}(T)$).

Por el teorema espectral existe $A = \{v_1, \ldots, v_n\}$ base ortonormal de V tal que $R(v_i) = \lambda_i v_i$ $\forall i = 1, \ldots, n$. Como $\text{rango}(T) = r$ se tiene que $\dim(\ker(T)) = n - r$ por lo tanto la base A se puede elegir de modo que $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0$ y $\lambda_i = 0$ si $i = r+1, \ldots, n$. Para $i = 1, \ldots, r$ definimos

$$w_i = \frac{T(v_i)}{\sigma_i}$$

donde $\sigma_i = \sqrt{\lambda_i}$. Entonces se tiene que $T(v_i) = \sigma_i w_i\ \forall\ i = 1, \ldots, r$. Además vimos que $\ker(T) = \ker(T^*T) = \ker(R)$, entonces $T(v_i) = 0\ \forall\ i = r+1, \ldots, n$ (pues $R(v_i) = 0$) tal como se quería. Falta probar que podemos construir $B = \{w_1, \ldots, w_n\}$ definida de este modo es una base ortonormal de V.

Veamos en primer lugar que $B' = \{w_1, \ldots, w_r\} \subset W$ es un conjunto ortonormal y consecuentemente linealmente independiente. En efecto

$$\langle w_i, w_j\rangle = \left\langle \tfrac{T(v_i)}{\sigma_i}, \tfrac{T(v_j)}{\sigma_j} \right\rangle = \left\langle \tfrac{v_i}{\sigma_i}, \tfrac{T^*(T(v_j))}{\sigma_j} \right\rangle = \frac{1}{\sigma_i\sigma_j}\langle v_i, R(v_j)\rangle = \frac{1}{\sigma_i\sigma_j}\langle v_i, \lambda_j v_j\rangle = \frac{\lambda_j}{\sigma_i\sigma_j}\langle v_i, v_j\rangle$$

Por lo tanto, usando que $A = \{v_1, \ldots, v_n\}$ es una base ortonormal de V, resulta que B' es un conjunto ortonormal. Sea $S = SG(\{w_1, \ldots, w_r\})$, entonces B' es una base ortonormal de S. Además se puede construir $B'' = \{w_{r+1}, \ldots, w_m\}$ base ortonormal de $S^\perp$ y por lo tanto $B = B' \cup B'' = \{w_1, \ldots, w_r, w_{r+1}, \ldots, w_m\}$ es una base ortonormal de W con las propiedades deseadas. ■

Corolario 3 (Descomposición en valores singulares de un matriz) *Sea $M \in \mathcal{M}_{m\times n}(I\!R)$ una matriz cualquiera con* $\text{rango}(M) = r$ *existen $U \in \mathcal{M}_{m\times m}(I\!R)$ y $V \in \mathcal{M}_{n\times n}(I\!R)$ ambas ortogonales y $S \in \mathcal{M}_{m\times n}(I\!R)$ diagonal con* $\text{rango}(S) = r$ *tal que $M = USV^t$*

Demostración. Basta elegir $T : I\!R^n \to I\!R^m$ tal que $T(x) = Mx$ entonces $M =_{\mathcal{C}_m} ((T))_{\mathcal{C}_n}$ donde $\mathcal{C}_i$ es la base canónica de $I\!R^i$. Por el teorema anterior se tiene que existen A y B bases ortonormales de $I\!R^n$ y $I\!R^m$ respectivamente tales que $T(v_i) = \sigma_i w_i$ con $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ y $\sigma_i = 0$ si $i > r$. Por lo tanto,

$$_B((T))_A = \begin{pmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix}$$

Además

$$_{\mathcal{C}_m}((T))_{\mathcal{C}_n} =_{\mathcal{C}_m} ((Id))_{B\,B}((T))_{A\,A}((Id))_{\mathcal{C}_n}.$$

Las matrices de cambio de base $_{\mathcal{C}_m}((Id))_B$ y $_A((Id))_{\mathcal{C}_n}$ son ortogonales pues transforman una base ortonormal en otra. Por lo tanto poniendo

$$U =_{\mathcal{C}_m} ((Id))_B, \quad V =_{\mathcal{C}_n} ((Id))_A \quad \text{y} \quad S =_B ((T))_A$$

se tiene que,

$$M = USV^t$$

y además

$$\text{rango}(S) = r$$

lo cual concluye la prueba. ■

Una consecuencia importante de este resultado es el llamado teorema de descomposición polar. Es sabido que todo número complejo (no nulo) se puede escribir como producto de un complejo de modulo uno y un número real positivo. Es decir, si $z \in \mathbb{C}$, $z \neq 0$ entonces existe $\theta \in [0, 2\pi)$ y $\rho > 0$ tal que $z = \rho e^{i\theta}$. Esta es la llamada descomposición polar de un número complejo. Ya hemos observado que la estructura algebraica de los números complejos con la operación de conjugación tiene una fuerte analogía con la de los operadores lineales donde el papel de la conjugación lo juega la operación de adjunto. El siguiente resultado puede ser interpretado en esta dirección.

Teorema 4 (Descomposición polar) *Sea V un espacio vectorial con producto interno de dimensión finita y sea $T : V \to V$ un operador invertible. Existen R y U operadores lineales tales que $T = R \circ U$ siendo R positivo y U unitario.*

Demostración. En virtud del teorema de descomposición en valores singulares y de la invertibilidad de T existen $A = \{v_1, \ldots, v_n\}$ y $B = \{w_1, \ldots, w_n\}$ bases ortonormales de V y números $\sigma_1 \geq \sigma_2, \geq, \ldots, \geq \sigma_n > 0$ tales que

$$T(v_i) = \sigma w_i, \;\; \forall\, i = 1, \ldots, n$$

Definimos $U : V \to V$ tal que $U(v_i) = w_i \;\forall\, i = 1, \ldots, n$ (como U es lineal basta definir los transformados de una base). U resulta unitaria pues lleva una base ortonormal en una base ortonormal.
Sea ahora $R : V \to V$ tal que $R(w_i) = \sigma_i w_i \; i = 1, \ldots, n$. R es autoadjunto y positivo pues B es una base ortonormal de vectores propios y los valores propios son reales positivos.
Para ver que T y $R \circ U$ coinciden baste ver que lo hacen en una base:

$$R \circ U(v_i) = R(U(v_i)) = R(w_i) = \sigma_i w_i = T(v_i), \;\; \forall\, i = 1, \ldots, n$$

Corolario 5 (Desomposición Polar para Matrices) *Sea $A \in \mathcal{M}_{n\times n}$ invertible. Entonces existen R positiva y U ortogonal tal que $A = R\overline{U}$ (usamos $\overline{U}$ para evitar confusiones con la U de la descomposición en valores singulares).*

Demostración. Sea $T : \mathbb{R}^n \to \mathbb{R}^n$ tal que $T(x) = Ax$. Por el teo. anterior sabemos que existen operadores R positivo y $\overline{U}$ ortogonal tal que:

$$A = {}_{\mathcal{C}}((T))_{\mathcal{C}} = {}_{\mathcal{C}}((R))_{\mathcal{C}\mathcal{C}}((U))_{\mathcal{C}}.$$

Por definición de R y U sabemos que:

$$_{\mathcal{B}'}((R))_{\mathcal{B}'} = S = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix}$$

$$_{\mathcal{B}'}((U))_{\mathcal{B}} = I_{n\times n} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}$$

En este caso como A es invertible, el cero no es valor propio. Finalmente, multiplicando por las matrices de cambio de base se obtiene que:

$$R = {}_{\mathcal{C}}((R))_{\mathcal{C}} = {}_{\mathcal{C}}((Id))_{\mathcal{B}'\mathcal{B}'}((R))_{\mathcal{B}'\mathcal{B}'}((Id))_{\mathcal{C}} = USU^t.$$

$$\overline{U} = {}_{\mathcal{C}}((U))_{\mathcal{C}} = {}_{\mathcal{C}}((Id))_{\mathcal{B}'\mathcal{B}'}((U))_{\mathcal{B}\mathcal{B}}((Id))_{\mathcal{C}} = UIV^t = UV^t.$$

Ejemplo 6 *Sea $A = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -\sqrt{2} & \sqrt{2} \end{pmatrix}$. Vamos a hallar la descomposición en valores singulares y la descomposición polar de A.*

Sea $B = A^tA = \begin{pmatrix} 5/2 & -3/2 \\ -3/2 & 5/2 \end{pmatrix}$. Los valores propios de B son $\lambda_1 = 4$ y $\lambda_2 = 1$ y los respectivos subespacios propios son $S_4 = [(-1,1)]$ y $S_1 = [(1,1)]$.

Por lo tanto los valores singulares son $\sigma_1 = 2$ y $\sigma_2 = 1$. Además, una base ortonormal (con el producto interno usual) formada por vectores propios de B es:

$$\mathcal{B} = \{v_1, v_2\} = \left\{ \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \right\}$$

De esta manera tomamos $\mathcal{B}' = \{w_1, w_2\}$ donde:

$$w_1 = \frac{T(v_1)}{\sigma_1} = \frac{Av_1}{2} = \frac{(0,2)}{2} = (0,1)$$
$$w_2 = \frac{T(v_2)}{\sigma_2} = \frac{Av_2}{2} = \frac{(1,0)}{1} = (1,0)$$

Entonces, la desomposición en valores singulares está dada por $A = USV^t$ donde

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

$$U = {}_{\mathcal{C}}((Id))_{\mathcal{B}'} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$V = {}_{\mathcal{B}}((Id))_{\mathcal{C}} = ({}_{\mathcal{C}}((Id))_{\mathcal{B}})^t = \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

La descomposición polar $A = R\overline{U}$ (usamos $\overline{U}$ para evitar confusiones) también se obtiene de las bases $\mathcal{B}$ y $\mathcal{B}'$ calculadas antes. Usando las matrices U y V halladas para la descomosición en valores singulares se obtiene que las matrices R y U están dadas por:

$$R = USU^t \quad \text{y} \quad \overline{U} = UV^t$$

En este caso, resulta que:

$$R = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{y} \quad U = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}.$$