Clase 29. Descomposición en Valores Singulares.

Curso 'Linear Algebra' del MIT.

Resumen

El teorema espectral entrega muchas facilidades al trabajar en matemática aplicada, pero no siempre es posible de usar ya que depende de matrices cuadradas y simétricas las que son poco frecuentes. A continuación veremos otra factorización y que es muy usada debido a que se puede aplicar a matrices rectangulares. Se conoce como **Descomposición en Valores Singulares**.

1. Descomposición en valores singulares (DVS).

La clase anterior vimos que trabajar con matrices diagonales hace más fácil aplicarlas en contextos más prácticos. Hasta ahora sabemos que si A es una matriz de $n \times n$ y tiene n eigenvalores distintos, entonces es posible diagonalizarla como:

$$A = S\Lambda S^{-1}$$

con S y Λ siendo las matrices de eigenvectores y eigenvalores de A, respectivamente.

Además, si se añade el hecho de que A también es simétrica, es posible factorizarla a partir del **teorema de descomposición espectral**.

$$A = Q\Lambda Q^T$$

donde Q es una matriz ortogonal cuyas columnas son los eigenvectores de A ortonormales.

Si bien las factorizaciones vistas anteriormente son muy útiles, dependen de matrices cuadradas y simétricas. Éstas son poco recurrentes en contextos prácticos a diferencia de las rectangulares. No obstante, ahora veremos una nueva factorización llamada **Descomposición en valores singulares** y se caracteriza por ser aplicable a matrices de distintas formas.

La descomposición en valores singulares (abreviado como DVS)[1] corresponde a la factorización de una matriz A de $n \times m$ en el siguiente producto matricial:

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

Esta decomposición es aplicable a matrices A de $n \times m$, donde $n \geq m$ o $n \leq m$.

Las matrices que componen la factorización en la DVS se conocen como:

- U: Matriz de vectores singulares izquierdos de A.
- V: Matriz de vectores singulares derechos de A.
- Σ: Matriz de valores singulares de A.

A continuación se resumen las características de estas matrices y sus valores.

1. Las matrices U y V son **ortogonales**, pero $U \neq V$.
2. La matriz Σ es diagonal rectangular y en su **diagonal principal** contiene a los **valores singulares** σ_i.
3. Los σ's son **positivos** (i.e, mayores a cero).

Por lo tanto, si A es de $n \times m$ con $n \neq m$, las dimensiones de las matrices al aplicar la DVS son:

$$A_{n\times m} = U_{n\times n}\ \Sigma_{n\times m}\ V^T_{m\times m}$$

2. La DVS y los subespacios vectoriales fundamentales.

Un hecho que muestra la relevancia de la descomposición en valores singulares, es que entrega **bases (*basis*) ortonormales** para **los cuatro subespacios vectoriales fundamentales**.

2.1 Bases ortonormales para los espacios columna y nulo.

Sea A una matriz de $n \times m$, con $n < m$. La DVS permite factorizarla como:

$$A_{n\times m} = U_{n\times n}\ \Sigma_{n\times m}\ V^T_{m\times m}$$

donde U y V son ortogonales, $U \neq V$ y Σ es diagonal rectangular.

Multipliquemos por V a la derecha de la ecuación de arriba. Como es ortogonal, se cumple

[1] En inglés se abrevia como SVD por *Singular Value Decomposition*. Es bueno considerarlo puesto que al calcularlo usando algún software, su función suele llamarse como dicha sigla.

que $VV^T = V^TV = I$ y nos lleva a que:

$$AV = U\Sigma$$

Las columnas de la matriz obtenida del producto entre dos de ellas, son iguales a la multiplicación entre la matriz de la izquierda con cada columna de la que está a la derecha. En ese sentido, AV se puede representar como:

$$(AV)_{n\times m} = \begin{bmatrix} A\vec{v}_1 & A\vec{v}_2 & \cdots & A\vec{v}_m \end{bmatrix}$$

Veamos que la cantidad de columnas de A y de V es la misma, por lo tanto todos los vectores columna de V estarán en AV. Esto, en cambio, no ocurre con $U\Sigma$.

En primer lugar, si bien cada columna de Σ será multiplicada por U en $U\Sigma$, en la práctica las columnas de este producto serán los vectores $\vec{u}_i$ escalados por los σ_i de la diagonal principal de Σ (i.e, $\sigma_i\vec{u}_i$), puesto que esta última es diagonal.

Por otra parte, la cantidad de columnas de U $(n \times n)$ es **menor** a la de Σ $(n \times m)$. Además, como esta última es diagonal, todo lo anterior concluye en que las n primeras columnas de $U\Sigma$ serán $\sigma\vec{u} \neq \vec{0}$, mientras que las $m - n$ restantes serán $\sigma\vec{u} = \vec{0}$.

$$(U\Sigma)_{n\times m} = \begin{bmatrix} \sigma_1\vec{u}_1 & \sigma_2\vec{u}_2 & \cdots & \sigma_n\vec{u}_n & \vec{0} & \cdots & \vec{0} \end{bmatrix}$$

En otras palabras:

$$(AV)_{n\times m} = (U\Sigma)_{n\times m}$$
$$\begin{bmatrix} A\vec{v}_1 & A\vec{v}_2 & \cdots & A\vec{v}_n & A\vec{v}_{n+1} & \cdots & A\vec{v}_m \end{bmatrix} = \begin{bmatrix} \sigma_1\vec{u}_1 & \sigma_2\vec{u}_2 & \cdots & \sigma_n\vec{u}_n & \vec{0} & \cdots & \vec{0} \end{bmatrix}$$

La igualdad de arriba implica, para $i = 1,\ \ldots,\ m$, que:

$$A\vec{v}_i = \sigma_i\vec{u}_i$$

Como se mencionó anteriormente, en $i = 1,\ \ldots,\ n < m$

$$A\vec{v}_i = \sigma_i\vec{u}_i \neq \vec{0}$$

lo que quiere decir que los primeros n vectores columna $\sigma_i\vec{u}_i$ de $U\Sigma$ provienen de la **combinación lineal** $A\vec{v}_i$.

En la Clase 6 estudiamos que el **espacio columna** de una matriz A, $C(A)$, corresponde al conjunto de vectores $\vec{b}$ que resultan de todas las combinaciones lineales $A\vec{x}$.[2] Por lo tanto, los n vectores $\sigma_i \vec{u}_i$ **ortogonales** generan este subespacio vectorial de A.

$$\text{span}(C(A)) = \{\sigma_1 \vec{u}_1,\ \sigma_2 \vec{u}_2,\ \dots,\ \sigma_n \vec{u}_n\}$$

Además, en la Clase 9 vimos que la dimensión del $C(A)$ es igual al rango (*rank*) r de esta matriz.

$$\dim(C(A)) = \text{rango}(A) = r$$

Supongamos que A es de **rango fila completo** o, en otras palabras, que $r = n < m$. Por lo tanto, los r primeros vectores columna de $U\Sigma$ forman una base (*basis*) para el $C(A)$.

$$\text{base}(C(A)) = \{\sigma_1 \vec{u}_1,\ \sigma_2 \vec{u}_2,\ \dots,\ \sigma_r \vec{u}_r\}$$

Si multiplicamos por $1/\sigma_i$ en $A\vec{v}_i = \sigma_i \vec{u}_i \neq \vec{0}$, entonces obtenemos una **base ortonormal para el espacio columna de** A.

$$\text{base}(C(A)) = \{\vec{u}_1,\ \vec{u}_2,\ \dots,\ \vec{u}_r\}$$

En cuanto a los $m - r$ vectores columna de AV y $U\Sigma$ se puede apreciar que:

$$A\vec{v}_i = \vec{0}$$

A partir de lo estudiado también en la Clase 6, se puede señalar en este caso que el **espacio nulo de** A, $N(A)$, son todos los vectores $\vec{v}_i$ que resuelven la ecuación $A\vec{v}_i = \vec{0}$.

$$\text{span}(N(A)) = \{\vec{v}_{r+1},\ \dots,\ \vec{v}_m\}$$

Por otra parte, en la Clase 9 vimos que la dimensión del $N(A)$ es:

$$\dim(N(A)) = m - r$$

que coincide con la cantidad de vectores que $\vec{v}_i$ donde $A\vec{v}_i = \vec{0}$. Por lo tanto, estas $m - r$ columnas de V conforman una **base ortonormal para el espacio nulo de** A.

$$\text{base}(N(A)) = \{\vec{v}_{r+1},\ \dots,\ \vec{v}_m\}$$

[2] Es decir, el $C(A)$ son todos los vectores $\vec{b} = A\vec{x}$.

2.2 Bases ortonormales para los espacios fila y nulo izquierdo.

Establezcamos que A es una matriz de $n \times m$, donde $n > m$.

Como ya sabemos, a partir de la DVS es posible factorizar a esta matriz A como:

$$A_{n\times m} = U_{n\times n}\ \Sigma_{n\times m}\ V^T_{m\times m}$$

Luego, tomemos la transpuesta de esta igualdad y multipliquemos por U a su derecha[3].

$$A^T = V\Sigma^T U^T$$
$$(A^T U)_{m\times n} = (V\Sigma^T)_{m\times n}$$

Algo a tener en cuenta, es que si bien $\Sigma \neq \Sigma^T$, las entradas en la diagonal principal de la primera también están en la misma ubicación de la segunda. Por lo tanto, la igualdad de arriba podemos expresarla como:

$$\begin{bmatrix} A^T\vec{u}_1 & A^T\vec{u}_2 & \cdots & A^T\vec{u}_m & A^T\vec{u}_{m+1} & \cdots & A^T\vec{u}_n \end{bmatrix} = \begin{bmatrix} \sigma_1\vec{v}_1 & \sigma_2\vec{v}_2 & \cdots & \sigma_m\vec{v}_m & \vec{0} & \cdots & \vec{0}_n \end{bmatrix}$$

Esta igualdad nos muestra que, para $i = 1,\ \ldots,\ n$:

$$A^T\vec{u}_i = \begin{cases} \sigma_i\vec{v}_i \neq \vec{0}, \text{ para } i = 1,\ \ldots,\ m < n \\ \sigma_i\vec{v}_i = \vec{0}, \text{ para } i = m+1,\ \ldots,\ n \end{cases}$$

En $i = 1,\ \ldots,\ m < n$, vemos que:

$$A^T\vec{u}_i = \sigma_i\vec{v}_i$$

Como vimos en la Clase 10, los m vectores $\sigma\vec{v}$ generan el **espacio fila de** A, $C(A^T)$, puesto que corresponden a todas las combinaciones lineales dadas por $A^T\vec{u}$.

$$\text{span}(C(A^T)) = \{\sigma_1\vec{v}_1,\ \sigma_2\vec{v}_2,\ \ldots,\ \sigma_m\vec{v}_m\}$$

Asumamos que A es de **rango columna completo** (i.e, $r = m < n$). En la misma Clase 10 estudiamos que la dimensión del $C(A^T)$ es la misma del $C(A)$.

$$\dim(C(A^T)) = \dim(C(A)) = \text{rango}(A) = r$$

Por lo tanto, los r primeros vectores columna de $V\Sigma^T$ forman una base ortogonal para el

[3] La matriz U es ortogonal, por tanto $UU^T = U^TU = I$.

$C(A^T)$.

$$\text{base}(C(A^T)) = \{\sigma_1\vec{v}_1,\ \sigma_2\vec{v}_2,\ \dots,\ \sigma_r\vec{v}_r\}$$

Si multiplicamos por $1/\sigma_i$ en la ecuación $A^T\vec{u}_i = \sigma_i\vec{v}_i$, obtenemos la **base ortonormal para el espacio fila de** A.

$$\text{base}(C(A^T)) = \{\vec{v}_1,\ \vec{v}_2,\ \dots,\ \vec{v}_r\}$$

Por otra parte, hemos podido observar que en $i = r+1,\ \dots,\ n$ (donde $r = m$):

$$A^T\vec{u}_i = \vec{0}$$

Es decir, los $n - r$ vectores $\vec{u}$ **generan el espacio nulo izquierdo de** A, $N(A^T)$.

$$\text{span}(N(A^T)) = \{\vec{u}_{r+1},\ \dots,\ \vec{u}_n\}$$

Y como coincide que:

$$\dim(N(A^T)) = n - r$$

tal como lo vimos en la Clase 10, entonces podemos concluir que los $n - r$ vectores columna de U forman una **base ortonormal para el espacio nulo izquierdo de** A.

$$\text{base}(N(A^T)) = \{\vec{u}_{r+1},\ \dots,\ \vec{u}_n\}$$

2.3 Conclusiones de la relación entre la DVS y los cuatro subespacios fundamentales.

A lo largo de esta sección se ha podido mostrar, de forma preliminar, que la descomposición en valores singulares de una matriz $A_{n\times m}$ nos entrega bases ortonormales para los cuatro subespacios vectoriales fundamentales de ella, como se resume en la siguiente tabla.

Base ortonormal	subespacio
$\{\vec{u}_1,\ \dots,\ \vec{u}_r\}$	$C(A)$
$\{\vec{u}_{r+1},\ \dots,\ \vec{u}_n\}$	$N(A^T)$
$\{\vec{v}_1,\ \dots,\ \vec{v}_r\}$	$C(A^T)$
$\{\vec{v}_{r+1},\ \dots,\ \vec{v}_m\}$	$N(A)$

A partir de la tabla de arriba podemos señalar que las columnas de $U_{n\times n}$ contienen las bases ortonormales del $C(A)$ y el $N(A^T)$ y las de $V_{m\times m}$ a las del $C(A^T)$ y el $N(A)$. Lo interesante es que coincide con lo estudiado en la Clase 14 sobre los **complementos ortogonales** de estos subespacios, como se aprecia en la tabla de a continuación:

Matriz DSV	Subespacio vectorial	Complemento ortogonal
$U_{n\times n}$	$C(A)$	$N(A^T)$
$V_{m\times m}$	$C(A^T)$	$N(A)$

donde tanto U como V son matrices ortogonales.

Conocer el vínculo entre la DVS y los cuatro subespacios fundamentales permite entender lo que ocurre en esta factorización: Se está **transformando linealmente** mediante una matriz A a vectores ortonormales de sus espacios fila y nulo en vectores ortonormales de sus espacios columna y nulo izquierdo, tal como se aprecia en la siguiente imagen[4].

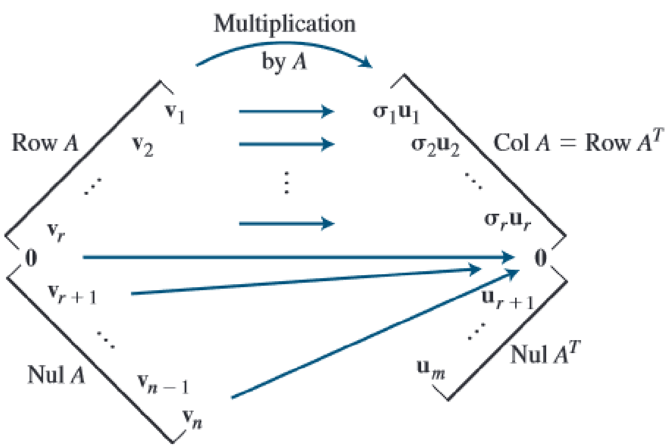


Figura 1: Lay, et al (2021). *Linear Algebra and Its Applications.* Pp. 470.

En la siguiente sección veremos de manera más formalizada lo estudiado en ésta, así como también una manera de obtener las matrices que conforman la descomposición en valores singulares, incluyendo a la matriz Σ que acá no se abordó en detalle.

3. Construyendo la DVS de una matriz A.

A continuación se mostrará cómo obtener las matrices de la DVS, donde lo estudiado sobre los eigenvalores/vectores y sus características en matrices simétricas es fundamental.

[4] En esta figura, la matriz A es de $m \times n$.

Definamos que la descomposición en valores singulares de una matriz $A_{n\times m}$ corresponde a:

$$A_{n\times m} = U_{n\times n}\ \Sigma_{n\times m}\ V^T_{m\times m}$$

Para producir las matrices de la factorización vista arriba, se usará a la matriz A^TA que, como veremos, se caracteriza por ser **positiva semidefinida**.

3.1 Características de la matriz A^TA.

En primer lugar, toda matriz cuadrada o rectangular que es multiplicada por su transpuesta resulta en una nueva matriz cuadrada. Acá A es de $n \times m$ y, por consiguiente, A^T es de $m \times n$. Por lo tanto, A^TA será de $m \times m$.

Por otra parte, se puede demostrar que A^TA es simétrica al calcular su transpuesta y al aplicar las propiedades de esta operación en ella.

$$A^TA = (A^TA)^T = A^T(A^T)^T = A^TA$$

Finalmente, para mostrar que A^TA es positiva semidefinida se evalúa que su forma cuadrática sea no negativa, como lo vimos en la Clase 27.

$$\vec{x}^T(A^TA)\vec{x} = (\vec{x}^TA^T)(A\vec{x}) = (A\vec{x})^T(A\vec{x}) = ||A\vec{x}||^2 \geq 0$$

El resultado obtenido es verdadero, porque todo número elevado al cuadrado siempre será no negativo. Por lo tanto, se demuestra que A^TA es **positiva semidefinida**.

Una de las principales características que vimos en la Clase 27 sobre matrices semidefinidas, es que sus **eigenvalores son mayores o iguales a cero**. Usando el hecho de que $||A\vec{x}||^2 \geq 0$, podemos ver que es cierto para A^TA.

Sea $(A^TA)\vec{x} = \lambda\vec{x}$, donde $\vec{x}$ es el eigenvector de A^TA y λ su correspondiente eigenvalor. Por lo tanto,

$$||A\vec{x}||^2 = (A\vec{x})^T(A\vec{x}) = \vec{x}^T[(A^TA)\vec{x}] = \vec{x}^T(\lambda\vec{x}) = \lambda(\vec{x}^T\vec{x}) = \lambda||\vec{x}||^2$$

Como $||A\vec{x}||^2 = \lambda||\vec{x}||^2$ y $||A\vec{x}||^2 \geq 0$, significa que $\lambda||\vec{x}||^2 \geq 0$. Si despejamos a λ en esta última desigualdad, entonces:

$$\lambda \geq 0$$

Esto nos permite concluir que los eigenvalores λ de A^TA son todos mayores o iguales a cero por ser positiva semidefinida.

3.2 Vectores singulares derechos y valores singulares de A.

Conociendo las características de $A^T A$, factoricémosla en valores singulares:

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

Recordemos que $UU^T = U^T U = I$, ya que la matriz U es ortogonal. Por lo tanto,

$$(A^T A)_{m\times m} = V_{m\times m}\ (\Sigma^T\Sigma)_{m\times m}\ V^T_{m\times m}$$

La matriz $\Sigma^T\Sigma$ es cuadrada y diagonal. En ese sentido, es válido señalar que la igualdad de arriba corresponde al **teorema de descomposición espectral** de $A^T A$, donde:

1. V es la matriz de eigenvectores de $A^T A$.
2. $\Sigma^T\Sigma$ es la matriz diagonal de eigenvalores de $A^T A$.

Como la matriz $A^T A$ es **simétrica**, podemos confirmar que la **matriz** V de los **vectores singulares derechos de** A es **ortogonal**, ya que sus columnas son los **eigenvectores ortonormales** de $A^T A$.

Por otra parte, como Σ es diagonal, las entradas de la diagonal principal de $\Sigma^T\Sigma$ serán $\sigma^2_{i=1,\ \dots,\ m}$, cumpliéndose que:

$$\text{diag}(\Sigma^T\Sigma) = \{\sigma_1^2,\ \dots,\ \sigma_m^2\} = \{\lambda_1,\ \dots,\ \lambda_m\}$$

De este modo, si tomamos la raíz cuadrada en las m igualdades

$$\sigma_i^2 = \lambda_i \qquad (i = 1,\ \dots,\ m)$$

es posible definir a los **valores singulares** de A como las **raíces cuadradas** de los **eigenvalores** $A^T A$.

$$\sigma_i = \sqrt{\lambda_i} \qquad (i = 1,\ \dots,\ m)$$

Recordemos que la matriz $A^T A$ es positiva semidefinida, lo que implica que sus m eigenvalores son $\lambda_i \geq 0$ y, por consiguiente, que $\sqrt{\lambda_i} \geq 0$. Esto nos permite confirmar que los **valores singulares** de A son **no negativos**.

$$\sigma_i \geq 0 \qquad (i = 1,\ \dots,\ m)$$

3.2.1 Número de valores singulares mayores a cero es igual al rango de A.

Que los m valores singulares σ de $A_{n\times m}$ sean no negativos significa que un grupo de ellos pueden ser iguales a cero, pero muchas veces el interés está en saber **cuántos** $\sigma > 0$. En esta ocasión se demostrará que

$$\text{N° de } \sigma_i > 0 \text{ es igual al rango}(A)$$

y se basará en la matriz $(A^TA)_{m\times m}$ puesto que, como veremos,

$$\text{rango}(A^TA) = \text{rango}(\Sigma^T\Sigma)$$

Si conocemos el rango$(\Sigma^T\Sigma)$ podemos saber cuántos $\lambda_i > 0$ tiene A^TA, ya que estarán en las **columnas linealmente independientes** de la primera matriz señalada[5] y debido a que

$$\sigma_i = \sqrt{\lambda_i} \qquad (\text{para } i = 1,\ \ldots,\ m),$$

entonces:

$$\sigma_i = \sqrt{\lambda_i} > 0 \qquad (\text{para } i = 1,\ \ldots,\ \text{rango}(A) \leq m)$$

No obstante, conocer la cantidad de $\sigma_i > 0$ de A a partir de A^TA puede ser tedioso debido al producto matricial, pero también demostraremos que:

$$\text{rango}(A^TA) = \text{rango}(A)$$

permiténdonos concluir que **el número de $\sigma_i > 0$ de A es igual a su rango**.

Las demostraciones se apoyarán en el **Teorema del Rango-Nulidad** que señala que la suma entre las dimensiones del espacio columna y del espacio nulo de una matriz, es igual a su cantidad de columnas. Es decir, si A es de $n \times m$, entonces:

$$\dim(C(A)) + \dim(N(A)) = m$$

Demostración 1. rango$(A^TA) = $ rango$(\Sigma^T\Sigma)$.

Anteriormente vimos que $(A^TA)_{m\times m}$ puede ser factorizada como:

$$A^TA = V(\Sigma^T\Sigma)V^T$$

donde V y $(\Sigma^T\Sigma)$ son las matrices de eigenvectores ortonormales y eigenvalores de (A^TA).

[5] Al ser una matriz diagonal, los vectores columna de $\Sigma^T\Sigma$ linealmente dependientes serán iguales a $\vec{0}$.

Debido a que la matriz V es ortogonal, también es posible escribir la igualdad de arriba como:

$$A^T A = V(\Sigma^T \Sigma)V^{-1}$$

Es decir, como estudiamos en la Clase 28, $A^T A$ y $\Sigma^T \Sigma$ son **matrices semejantes**.

$$A^T A \sim \Sigma^T \Sigma$$

Además de tener los mismos eigenvalores, las matrices semejantes también se caracterizan por tener **el mismo rango (*rank*)**.

Digamos que $\vec{x} \in N(A^T A)$. Es decir,

$$(A^T A)\vec{x} = \vec{0}$$

Debido a que $A^T A = V(\Sigma^T \Sigma)V^{-1}$, entonces:

$$[V(\Sigma^T \Sigma)V^{-1}]\vec{x} = \vec{0}$$

Multipliquemos por V^{-1} a la izquierda de la igualdad de arriba.

$$V^{-1}V(\Sigma^T \Sigma)V^{-1}\vec{x} = V^{-1}\vec{0}$$
$$(\Sigma^T \Sigma)(V^{-1}\vec{x}) = \vec{0}$$

Es decir, el vector $(V^{-1}\vec{x}) \in N(\Sigma^T \Sigma)$. Por lo tanto,

$$N(A^T A) \subseteq N(\Sigma^T \Sigma)$$

Recordemos que la semejanza entre matrices es bidireccional. En otras palabras,

$$\Sigma^T \Sigma = V^{-1}(A^T A)V$$

Establezcamos que $\vec{y} \in N(\Sigma^T \Sigma)$.

$$(\Sigma^T \Sigma)\vec{y} = \vec{0}$$

Esta igualdad también podemos expresarla como:

$$[V^{-1}(A^T A)V]\vec{y} = \vec{0}$$

Al multiplicar por V a la izquierda obtenemos que:

$$V[V^{-1}(A^TA)V]\vec{y} = V\vec{0}$$
$$(A^TA)(V\vec{y}) = \vec{0}$$

En otras palabras, $(V\vec{y}) \in N(A^TA)$, lo que implica que:

$$N(\Sigma^T\Sigma) \subseteq N(A^TA)$$

y por tanto que:

$$N(A^TA) = N(\Sigma^T\Sigma) \Longrightarrow \dim(N(A^TA)) = \dim(N(\Sigma^T\Sigma))$$

Luego, apliquemos el teorema del rango-nulidad en $(\Sigma^T\Sigma)_{m\times m}$.

$$\dim(C(\Sigma^T\Sigma)) + \dim(N(\Sigma^T\Sigma)) = m$$

Debido a que $\dim(C(\Sigma^T\Sigma)) = \text{rango}(\Sigma^T\Sigma)$ y $\dim(N(\Sigma^T\Sigma)) = \dim(N(A^TA))$, entonces:

$$\text{rango}(\Sigma^T\Sigma) + \dim(N(A^TA)) = m$$

Despejemos al rango$(\Sigma^T\Sigma)$.

$$\text{rango}(\Sigma^T\Sigma) = m - \dim(N(A^TA))$$

Como $\dim(N(A^TA)) = m - \text{rango}(A^TA)$, en consecuencia:

$$\text{rango}(\Sigma^T\Sigma) = m - (m - \text{rango}(A^TA)) = \text{rango}(A^TA) \qquad \text{(Q. E. D)}$$

Demostración 2. $\text{rango}(A) = \text{rango}(A^TA)$.

Sea $\vec{x} \in N(A)$. Esto significa que

$$A\vec{x} = \vec{0}$$

Multipliquemos por A^T a la izquierda de esta igualdad.

$$A^TA\vec{x} = A^T\vec{0}$$
$$(A^TA)\vec{x} = \vec{0}$$

Por lo tanto, se puede señalar que $N(A) \subseteq N(A^TA)$.

Luego, definamos que $\vec{x} \in N(A^TA)$. Es decir,

$$A^TA\vec{x} = \vec{0}$$

Al multiplicar por $\vec{x}^T$ a la izquierda de esta igualdad, obtenemos lo siguiente:

$$\begin{aligned} \vec{x}^T(A^TA)\vec{x} &= \vec{x}^T\vec{0} \\ (A\vec{x})^T(A\vec{x}) &= 0 \\ ||A\vec{x}||^2 &= 0 \\ ||A\vec{x}|| &= 0 \end{aligned}$$

Para que el vector $A\vec{x}$ tenga magnitud igual a cero, la única posibilidad es que:

$$A\vec{x} = \vec{0}$$

Como $A = [a_{i,j}] \neq 0$, solo nos queda que el vector $\vec{x} = \vec{0}$. Por lo tanto,

$$N(A^TA) \subseteq N(A)$$

En consecuencia,

$$N(A^TA) = N(A) \Longrightarrow \dim(N(A^TA)) = \dim(N(A))$$

Luego, despejemos la $\dim(C(A^TA))$ en la igualdad del teorema del rango-nulidad de A^TA.

$$\dim(C(A^TA)) = m - \dim(N(A^TA))$$

Puesto que $\dim(C(A^TA)) = \text{rango}(A^TA)$ y que $\dim(N(A^TA)) = \dim(N(A)) = m - \text{rango}(A)$, entonces:

$$\text{rango}(A^TA) = m - (m - \text{rango}(A)) = \text{rango}(A) \qquad \text{(Q. E. D)}$$

3.3 Vectores singulares izquierdos de A.

En esta subsección veremos dos formas de obtener a la matriz $U_{n\times n}$ de los vectores izquierdos de $A_{n\times m}$. La primera es a partir de $(AA^T)_{n\times n}$ y es útil para confirmar que $U \neq V$. En cuanto a la segunda, es una continuación de haber usado a A^TA para obtener a V y a Σ.

3.3.1 Calculando a la matriz U a partir de AA^T.

Sea A una matriz de $n \times m$. Para conocer a $U_{n\times n}$ en su DVS, se usará a la matriz $(AA^T)_{n\times n}$ que se caracteriza por ser **simétrica**:

$$AA^T = (AA^T)^T = (A^T)^T A^T = AA^T$$

y **positiva semidefinida** para todo $\vec{x} \neq \vec{0}$.

$$\vec{x}^T(AA^T)\vec{x} = (\vec{x}^T A)(A^T\vec{x}) = (A^T\vec{x})^T(A^T\vec{x}) = ||A^T\vec{x}||^2 \geq 0$$

Lo anterior implica que **todos los eigenvalores de AA^T son no negativos** como vemos a continuación, donde $(AA^T)\vec{x} = \lambda\vec{x}$ para todo $\vec{x} \neq \vec{0}$.

$$||A^T\vec{x}||^2 = (A^T\vec{x})^T(A^T\vec{x}) = \vec{x}^T(AA^T)\vec{x} = \lambda\vec{x}^T\vec{x} = \lambda||\vec{x}||^2 \geq 0$$
$$\therefore \lambda \geq 0$$

Por la descomposición en valores singulares, podemos señalar que $A = U\Sigma V^T$. Apliquemos esta factorización a AA^T.

$$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T = U\Sigma V^T V\Sigma^T U^T$$

Como $VV^T = V^T V = I$ debido a que V es una matriz ortogonal, entonces:

$$AA^T = U(\Sigma\Sigma^T)U^T$$

donde $U_{n\times n}$ y $(\Sigma\Sigma^T)_{n\times n}$ son las matrices de eigenvectores ortonormales y eigenvalores de AA^T.

Una primera observación que se puede sacar de aplicar la DVS en AA^T, es que se confirma que $U \neq V$ porque V proviene de $A^T A$ y $AA^T \neq A^T A$.

En segundo lugar, debemos ser cuidadosos al usar a las matrices AA^T y $A^T A$ en la DVS de A, ya que los valores singulares se obtienen en dos matrices distintas: $\Sigma^T\Sigma$ y $\Sigma\Sigma^T$.

$$\Sigma^T\Sigma \neq \Sigma\Sigma^T$$

En general, es posible que los valores singulares en ambas matrices sean los mismos, pero esto no siempre es así. El costo es que obtengamos vectores singulares derechos (V) e izquierdos (U) que no nos entreguen de vuelta a la matriz A. Es por ello que veremos una segunda forma de conseguir a los $\vec{u}$ en la que se evita este problema.

3.3.2 Cálculo más sencillo de la matriz U.

Existe una forma más sencilla de obtener a los vectores singulares izquierdo $\vec{u}_i$ que veremos a continuación.

Primero, digamos que tenemos una matriz $A_{n\times m}$ y queremos factorizarla como:

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

Luego, **asumamos que a las matrices V y Σ las obtuvimos a partir de** A^TA como lo hicimos en la Sección 3.2. Esto implica que V es ortogonal, de manera que al multiplicar por esta mariz a la derecha en la ecuación de arriba resulta en:

$$AV = U\Sigma$$

En la Sección 2.1 vimos que esta ecuación es equivalente a:

$$A\vec{v}_i = \sigma_i\vec{u}_i$$

Al multiplicar por $1/\sigma_i$ obtenemos una fórmula para calcular a los vectores singulares izquierdos de A.

$$\vec{u}_i = \frac{1}{\sigma_i}A\vec{v}_i$$

La ventaja de esta fórmula de los $\vec{u}_i$ de A es que resultan de valores y vectores conocidos previamente: $\vec{v}_i$ y σ_i que provienen de A^TA, lo que significa que habrá una mayor concordancia entre ellos a diferencia de lo visto en la Sección 3.3.1.

Los vectores $\vec{u}_i$ son unitarios puesto que son los $A\vec{v}_i$ divididos por la magnitud de estos últimos, σ_i. No obstante, aún necesitamos demostrar que sean **ortogonales** para que la matriz U también lo sea y es lo que veremos a continuación.

Demostración 3. Los vectores $\vec{u}_i$ son ortogonales.

Vamos a demostrar que para $j \neq k$,

$$\vec{u}_j^T\vec{u}_k = 0$$

Como $\vec{u} = (1/\sigma)A\vec{v}$, entonces:

$$\left(\frac{1}{\sigma_j}\vec{v}_j^T A^T\right)\left(\frac{1}{\sigma_k}A\vec{v}_k\right) = 0$$
$$\frac{1}{\sigma_j\sigma_k}(\vec{v}_j^T A^T A\vec{v}_k) = 0$$

Los vectores $\vec{v}_i$ son los eigenvectores ortonormales de A^TA, mientras que σ_i^2 son los eigenvalores de la misma matriz. Esto significa que cada vector $(A^TA)\vec{v}_i = \sigma_i^2\vec{v}_i$ y, por lo tanto,

$$\frac{1}{\sigma_j\sigma_k}(\vec{v}_j^T\sigma_k^2\vec{v}_k) = 0$$
$$\frac{\sigma_k}{\sigma_j}(\vec{v}_j^T\vec{v}_k) = 0$$

Que los vectores $\vec{v}_i$ sean ortonormales implica que $\vec{v}_j^T\vec{v}_k = 0$ para $j \neq k$. En consecuencia,

$$\frac{\sigma_k}{\sigma_j}\cdot 0 = 0$$
$$0 = 0 \qquad \text{(Q. E. D)}$$

Así, hemos verificado que los $\vec{u}_i = (1/\sigma_i)A\vec{v}_i$ son los **vectores singulares izquierdos** de A y que, por consiguiente, conforman las columnas de la **matriz ortogonal** $U_{n\times n}$.

4. Versión reducida de la DVS.

En las secciones 2 y 3 vimos que las matrices $U_{n\times n}$, $V_{m\times m}$ y $\Sigma_{n\times m}$ de la descomposición en valores singulares de $A_{n\times m}$ tendrán vectores columna iguales a $\vec{0}$. Sin embargo, muchas veces el interés está en aquellas que son distintos a ese vector y, por lo tanto, se suele trabajar con una **versión reducida** de esta factorización.

Primero, definamos una práctica común al trabajar con la DVS y que es ordenar de forma descendente a los valores singulares σ_i de A, donde sabemos que los r primeros (con $r =$ rango(A)) son mayores a cero y el resto igual a dicho valor.

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0 \quad \text{y} \quad \sigma_{r+1} = \sigma_{r+2} = \ldots = \sigma_m = 0$$

Si cada valor singular tiene su vector singular (izquierdo y derecho) correspondiente, entonces también estarán ordenados de forma descendente en sus respectivas matrices. Este hecho permite expresarlas a todas, incluido a Σ, como **bloque de matrices** en la DVS.

$$A = U\Sigma V^T = \begin{bmatrix} U_{n\times r} & U_{n\times(n-r)} \end{bmatrix} \begin{bmatrix} \Sigma_{r\times r} & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{bmatrix} \begin{bmatrix} V^T_{m\times r} \\ V^T_{m\times(m-r)} \end{bmatrix}$$

La matriz $\Sigma_{r\times r}$ es cuadrada diagonal, cuyas r entradas de su diagonal principal son los valores singulares $\sigma_i > 0$ de A ordenados en forma descendente.

$$\Sigma_{r\times r} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix}$$

Así, el producto matricial de la DVS de A usando los bloques de matrices resulta en la **versión reducida** de esta factorización:

$$A = U_{n\times r}\ \Sigma_{r\times r}\ V^T_{m\times r}$$

Algo que debemos tener en cuenta es que si bien las columnas de $U_{n\times r}$ y las de $V_{m\times r}$ son ortonormales, dichas matrices **no son ortogonales** puesto que **no son cuadradas**. No obstante, sí se cumplen las siguientes igualdades:

$$(U^TU)_{r\times r} = I_{r\times r} \qquad \text{y} \qquad (V^TV)_{r\times r} = I_{r\times r}$$

pero $(UU^T)_{n\times n} \neq (U^TU)_{r\times r}$ y $(VV^T)_{m\times m} \neq (V^TV)_{r\times r}$.

5. Geometría de la DVS.

Para terminar, revisemos el proceso geométrico de la descomposición en valores singulares.

Supongamos que una matriz $A_{n\times m}$ se puede factorizar como:

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

Luego, digamos que $\vec{x}$ es un vector de $n \times 1$. Al ser multiplicado por A sucede que:

$$A\vec{x} = (U\Sigma V^T)\vec{x}$$

Es decir, $\vec{x}$ es multiplicado por las tres matrices de la DVS de A, cuyo proceso es el siguiente:

$$\begin{tikzcd}
\vec{x} \arrow[r] \arrow[rrr, bend left, dashed, "A"] & V^T\vec{x} \arrow[r] & (\Sigma V^T)\vec{x} \arrow[r] & (U\Sigma V^T)\vec{x}
\end{tikzcd}$$

El producto entre una **matriz ortogonal** y un vector hace que éste sea **rotado** en su punto inicial. Cuando dicha operación se realiza con una **matriz diagonal**, el vector se **estira o achica** según los valores de la diagonal. Dichos efectos los vemos en la DVS de A:

1. $V^T\vec{x} \longrightarrow$ El vector $\vec{x}$ es rotado a partir de V^T.
2. $(\Sigma V^T)\vec{x} \longrightarrow$ El vector $(V^T\vec{x})$ es alargado o achicado por Σ.
3. $(U\Sigma V^T)\vec{x} \longrightarrow$ El vector $(\Sigma V^T\vec{x})$ es vuelto a rotar, pero por U.

En otras palabras, el proceso de ir desde $\vec{x}$ a $A\vec{x}$ podemos entenderlo a partir de los tres puntos señalados arriba y que son dados por la descomposición en valores singulares de A.

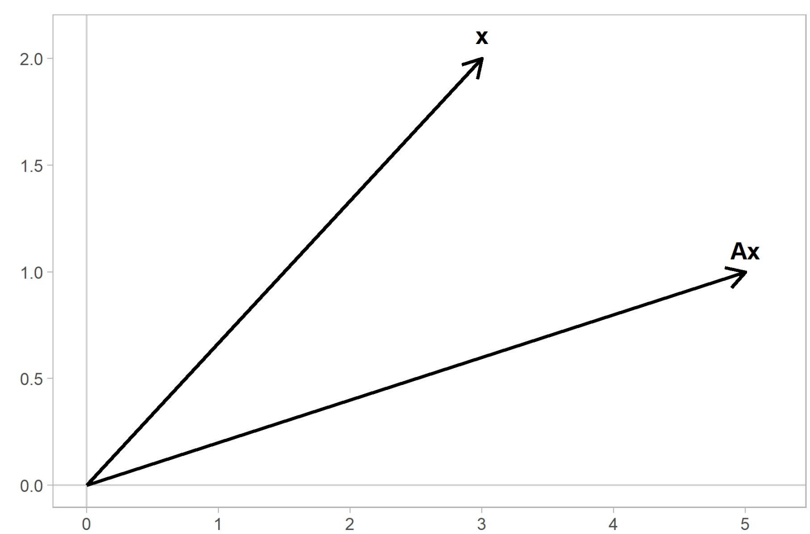
Ejemplo. Sean

$$\vec{x} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad \text{y} \qquad A = \begin{bmatrix} 1 & 1 \\ 3 & -4 \end{bmatrix}$$

Esto implica que:

$$A\vec{x} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

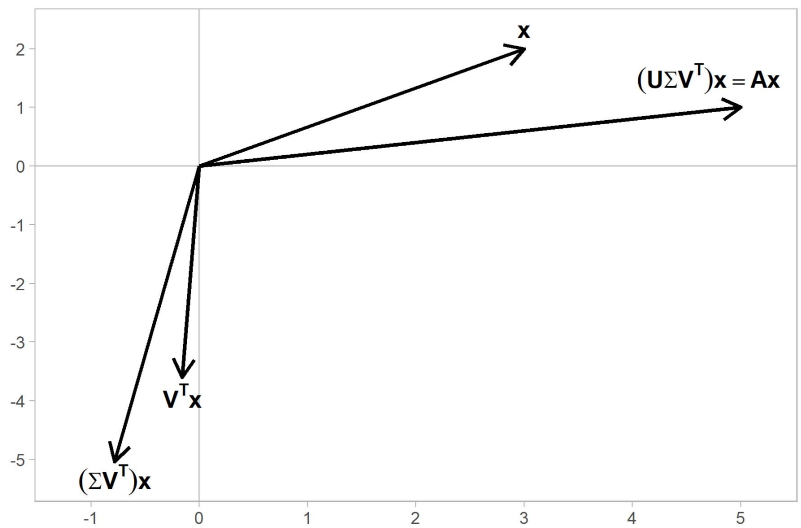
A continuación tenemos la gráfica de $\vec{x}$ y $A\vec{x}$.



Luego, calculemos la DVS de A. Como la idea es ver la geometría de esta factorización, construí las matrices U, Σ y V usando la función `svd()` del lenguaje de programación R. Acá muestro sus entradas redondeadas a 4 dígitos (salvo los ceros de Σ).

$$U = \begin{bmatrix} 0.0434 & -0.9991 \\ -0.9991 & -0.0434 \end{bmatrix}; \quad \Sigma = \begin{bmatrix} 5.0043 & 0 \\ 0 & 1.3988 \end{bmatrix}; \quad V = \begin{bmatrix} -0.5903 & -0.8072 \\ 0.8072 & -0.5903 \end{bmatrix}$$

El producto de $\vec{x}$ con estas matrices nos muestra como se transforma en $A\vec{x}$, tal como se aprecia en la siguiente gráfica.



donde:

$$(U\Sigma V^T)\vec{x} = \begin{bmatrix} 5 \\ 1 \end{bmatrix} = A\vec{x}$$