Coloquio de matemática aplicada, 2019

Fundamentos teóricos de la descomposición en valores singulares de una matriz y sus aplicaciones

Prof. Juan José Fallas

TEC | Tecnológico de Costa Rica

Matrices por bloques

La idea es descomponer una matriz en submatrices para operar más eficientemente. Se busca organizar una matriz $A_{m\times n}$ de la forma:

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1r} \\ \vdots & & & \vdots \\ A_{p1} & A_{p2} & \dots & A_{pr} \end{pmatrix},$$

de tal manera que ahora se tiene p **bloques-fila** y r **bloques columna**, y por ende A fue organizada como una "nueva" matriz de "tamaño" $p \times r$. Se debe satisfacer que para i **fijo** $(1 \leq i \leq p)$ todas las submatrices A_{ij} $(1 \leq j \leq r)$ tienen que tener la misma cantidad de filas (no así con el número de columnas). Recíprocamente, para j **fijo** todas las submatrices A_{ij} tienen que tener la misma cantidad de columnas (no así con el número de filas). Evidentemente, la representación por bloques de una matriz **no es única**.

Ejemplo

$$A = \begin{pmatrix} -1 & 0 & 2 & 3 & 4 & 1 \\ 0 & 3 & -2 & 4 & 1 & 1 \\ -4 & -2 & 1 & 0 & 5 & 3 \\ 1 & 3 & 3 & 2 & 2 & 0 \end{pmatrix}$$

$$A = \left(\begin{array}{cc|c|cc|c} -1 & 0 & 2 & 3 & 4 & 1 \\ \hline 0 & 3 & -2 & 4 & 1 & 1 \\ -4 & -2 & 1 & 0 & 5 & 3 \\ \hline 1 & 3 & 3 & 2 & 2 & 0 \end{array} \right) = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \end{pmatrix}$$

$$A = \left(\begin{array}{cc|ccc|c} -1 & 0 & 2 & 3 & 4 & 1 \\ 0 & 3 & -2 & 4 & 1 & 1 \\ \hline -4 & -2 & 1 & 0 & 5 & 3 \\ 1 & 3 & 3 & 2 & 2 & 0 \end{array} \right) = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{pmatrix}$$

Definición (Multiplicación de matrices por bloques)

Sean $A_{m\times n}$ y $B_{n\times s}$ matrices organizadas por bloques, de tal manera que:

- A tiene p bloques-fila y r bloques columna.
- B tiene r bloques-fila y t bloques columna.

El producto AB **por bloques** de dichas matrices se define como la matriz C que tendrá p bloques-fila y t bloques columna y tal que su bloque ij ($1 \leq i \leq p$, $1 \leq j \leq t$) está definido por:

$$C_{ij} = \sum_{k=1}^{r} A_{ik}B_{kj},$$

siempre que el producto de bloques $A_{ik}B_{kj}$ esté bien definido **para todo** k (esto es, el número de columnas del pimer bloque en dicho producto debe coincidir con el número de filas del segundo bloque).

Ejemplo

$$A = \begin{pmatrix} -1 & 0 & 2 & 3 & 4 & 1 \\ 0 & 3 & -2 & 4 & 1 & 1 \\ -4 & -2 & 1 & 0 & 5 & 3 \\ 1 & 3 & 3 & 2 & 2 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 4 & 0 & 1 \\ 2 & -1 & 3 \\ 4 & 1 & 5 \\ 3 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & -4 & 1 \end{pmatrix}$$

$$A = \left(\begin{array}{cc|cc|c|c} -1 & 0 & 2 & 3 & 4 & 1 \\ \hline 0 & 3 & -2 & 4 & 1 & 1 \\ -4 & -2 & 1 & 0 & 5 & 3 \\ \hline 1 & 3 & 3 & 2 & 2 & 0 \end{array} \right)_{3\times 4} \qquad B = \left(\begin{array}{cc|c} 4 & 0 & 1 \\ 2 & -1 & 3 \\ \hline 4 & 1 & 5 \\ \hline 3 & 2 & -1 \\ 2 & 2 & 2 \\ \hline -1 & -4 & 1 \end{array} \right)_{4\times 2}$$

Ejemplo

$$A = \left(\begin{array}{cc|c|cc|c} -1 & 0 & 2 & 3 & 4 & 1 \\ \hline 0 & 3 & -2 & 4 & 1 & 1 \\ -4 & -2 & 1 & 0 & 5 & 3 \\ \hline 1 & 3 & 3 & 2 & 2 & 0 \end{array}\right)_{3\times 4} \quad B = \left(\begin{array}{cc|c} 4 & 0 & 1 \\ 2 & -1 & 3 \\ \hline 4 & 1 & 5 \\ \hline 3 & 2 & -1 \\ 2 & 2 & 2 \\ \hline -1 & -4 & 1 \end{array}\right)_{4\times 2}$$

$$C = \left(\begin{array}{cc|c} 20 & 12 & 15 \\ \hline 11 & 1 & -2 \\ -9 & 1 & 8 \\ \hline 32 & 8 & 27 \end{array}\right)_{3\times 2}$$

Rango de una matriz

Sea $A \in \mathrm{Mat}(\mathbb{C}, m, n)$. El **rango columna** de A es el número máximo de columnas que son linealmente independientes. Análogamente se define el **rango fila**.

Se demuestra que el rango fila y el rango columna siempre son iguales, y a ese valor se le denomina el rango de la matriz y se denota $\mathrm{r}(A)$. Además,

$$0 \leq \mathrm{r}(A) \leq \mathrm{mín}\{m, n\}$$

Si $\mathrm{r}(A) = \mathrm{mín}\{m, n\}$, entonces se dice que A es de **rango completo**.

Propiedades del rango de una matriz

1. Sea $A \in \mathrm{Mat}(\mathbb{C}, m, n)$ y $B \in \mathrm{Mat}(\mathbb{C}, n, p)$, entonces $\mathrm{r}(AB) \leq \mathrm{mín}\{\mathrm{r}(A), \mathrm{r}(B)\}$.
2. Sea $A \in \mathrm{Mat}(\mathbb{C}, n)$. A es invertible si y solo si $\mathrm{r}(A) = n$, esto es, A es invertible si y solo si es de rango completo.
3. Si $A \in \mathrm{Mat}(\mathbb{C}, m, n)$ y $B \in \mathrm{Mat}(\mathbb{C}, m)$, tal que B **es invertible**, entonces: $\mathrm{r}(BA) = \mathrm{r}(A)$.
4. Sea $\Sigma \in \mathrm{Mat}(\mathbb{C}, m, n)$ tal que $\sigma_{ij} = 0$, cuando $i \neq j$, esto es Σ es una **matriz diagonal generalizada**. Entonces el rango de Σ es el número de entradas diferentes de cero que tenga Σ.

Valor propio y vector propio de una matriz

Sea $A \in \mathrm{Mat}(\mathbb{C}, n)$. Un vector $v \in \mathbb{C}^n$, $v \neq \mathbf{0}$, se le llama *vector propio* de A si existe un escalar λ tal que

$$Av = \lambda v \iff \lambda v - Av = \mathbf{0} \iff (\lambda I_n - A)\, v = \mathbf{0}$$

A λ se le denomina *valor propio* de A, asociado al vector propio v.

Note que λ da lugar a una ecuación matricial $(\lambda I_n - A)\, X = \mathbf{0}$, donde $X_{n\times 1}$ es la matriz de las incógnitas. Así, λ es un valor propio de A si esta ecuación tiene alguna solución $X \neq \mathbf{0}$. Se busca entonces que dicha ecuación tenga soluciones no triviales, las cuales darán origen a los vectores propios asociados a λ.

Teorema de Rouché-Frobënius

Un sistema de ecuaciones lineales $AX = B$ es compatible si y solo si $\mathrm{r}(A) = \mathrm{r}\left(A \vdots B\right)$.

Para el caso del sistema $(\lambda I_n - A)\, X = \mathbf{0}$, si $\mathrm{r}(\lambda I_n - A) = n$, entonces el sistema tiene solución única y por ende no existe $X \neq \mathbf{0}$ que satisfaga el sistema. Por ende, λ es valor propio de A si y solo si $\mathrm{r}(\lambda I_n - A) < n$, y por lo tanto λ es valor propio de A si y solo si $|\lambda I_n - A| = 0$.

Ecuación y polinomio característico

La ecuación $|\lambda I_n - A| = 0$ se le llama *ecuación característica* de A y al determinante $P(\lambda) = |\lambda I_n - A|$ se le denomina *polinomio característico* de A. Los valores propios de A son las raíces de su polinomio característico.

Matriz diagonal

La matriz $D \in \mathrm{Mat}(\mathbb{C}, n)$ se le llama diagonal si $D_{ij} = 0$ para $i \neq j$.

Matriz diagonalizable

Sea $A \in \mathrm{Mat}(\mathbb{C}, n)$. Se dice que A es diagonalizable si existe una matriz Q no singular tal que

$$A = QDQ^{-1} \iff Q^{-1}AQ = D,$$

donde D es una matriz diagonal. Esto es, A debe ser *semejante* a una matriz diagonal.

Preliminares

Teorema

La matriz $A \in \mathrm{Mat}(\mathbb{C}, n)$ es diagonalizable si y solo si existe una base para $\mathbb{C}^n$ formada por vectores propios de A.

Teorema

La matriz $A \in \mathrm{Mat}(\mathbb{C}, n)$ es diagonalizable si y solo si su polinomio característico tiene solo raíces reales y para cada valor propio λ, de multiplicidad $k \geq 2$, deben existir k vectores propios linealmente independientes.

En caso que A sea diagonalizable, las columnas de Q están formadas por tales vectores propios y D contiene los valores propios de A, ordenados de manera respectiva como los vectores propios correspondientes fueron colocados como columnas de Q.

Ejemplo diagnonalizable

Considere la matriz $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$. Su polinomio característico es $P(\lambda) = (\lambda - 2)(\lambda - 4)$. Para $\lambda = 2$ resolvemos:

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

que genera $-x - y = 0 \iff y = -x$. Por ende, podemos seleccionar el vector propio $(1, -1)^{\mathbf{T}}$. Luego, con $\lambda = 4$ resolvemos:

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

que genera $x - y = 0 \iff y = x$. Luego, tomamos el vector propio $(1, 1)^{\mathbf{T}}$. Así,

$$Q = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad \text{y} \quad D = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$$

Ejemplo NO diagonalizable

Considere $A = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 0 & 1 \\ 0 & -2 & 3 \end{pmatrix}$, cuyo polinomio característico es $P(\lambda) = (\lambda - 1)^2(\lambda - 2)$. Analizamos el valor propio $\lambda = 1$. Se debe resolver el sistema:

$$\begin{pmatrix} 0 & -2 & 1 \\ 0 & 1 & -1 \\ 0 & 2 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \iff \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

El espacio propio asociado a $\lambda = 1$ está dado por $\{t \cdot (0,0,1) : t \in \mathbb{C}\}$ y por ende no es posible seleccionar dos vectores propios linealmente independientes. Finalmente, A no es diagonalizable.

Desventajas de la diagonalización

El proceso de expresar a A, en caso que sea posible, como QDQ^{-1} en general tiene algunas desventajas:

- La matriz A tiene que ser cuadrada.
- Los vectores propios de A usualmente no son ortogonales.
- Usualmente no hay "suficientes" vectores propios.

La descomposición de A en **valores singulares** evita dichos problemas, comenzando porque dicha descomposición existe para cualquier $A \in \mathrm{Mat}(\mathbb{C}, m, n)$.

Definición de la transpuesta conjugada de una matriz

Sea $m, n \in \mathbb{N}$ y $U \in \mathrm{Mat}(\mathbb{C}, m, n)$. La **transpuesta conjugada** de U se denotará U^*, y satisface que:

$$u_{ij}^* = \overline{u_{ji}}$$

Si U solo tiene entradas reales, entonces $U^* = U^{\mathbf{T}}$.

Definición de matriz hermitiana

Una matriz $U \in \mathrm{Mat}(\mathbb{C}, n)$ se le llama **hermitiana** o **autoadjunta** si $U = \overline{U^{\mathbf{T}}} = U^*$ (si es igual a su transpuesta conjugada).

Producto interno en $\mathbb{C}$

Sea V un espacio vectorial sobre $\mathbb{C}$, un producto interno en $V \times V$ es una función

$$\langle \cdot, \cdot \rangle \ : V \times V \to F$$

que satisface para todo $u, v, z \in V$ y para todo $\alpha \in \mathbb{C}$ lo siguiente:

1. $\langle x, y \rangle \ = \overline{\langle y, x \rangle}$
2. $\langle \alpha x, y \rangle \ = \alpha \langle x, y \rangle$
3. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$

1. $\langle x, x \rangle \ \geq 0$
2. $\langle x, x \rangle \ = 0 \iff x = 0$

De la definición anterior se ve deduce fácilmente que:

$$\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle \quad \text{y} \quad \langle x, \alpha y \rangle \ = \overline{\alpha} \langle x, y \rangle$$

Forma general de un producto interno en $\mathbb{C}^n$

La forma general de un producto interno en $\mathbb{C}^n$ se conoce como la forma hermitiana y es tal que para $u, v \in \mathbb{C}^n$:

$$\langle u, v \rangle = v^* M u = \overline{u^* M v},$$

donde M es cualquier matriz hermitiana definida positiva y v^* es el conjugado transpuesto de v. Es común seleccionar $M = I_n$ y, por ende, usualmente se esribe: $\langle u, v \rangle = v^* u$.

Nota: La matriz M hermitiana se dice definida positiva si $u^* M u > 0$, para todo $u \in \mathbb{C}^n$. Otra forma equivalente de definirla es que todos los autovalores de M son positivos.

Conjuntos y bases ortogonales

Teorema

Sea V un espacio vectorial de dimensión n sobre $\mathbb{C}$, con producto interno. Si $S = \{u_1, \ldots, u_n\}$ es un conjunto ortogonal de vectores no nulos de V, entonces S es un conjunto linealmente independiente. De hecho, S es una base de V (denominada **base ortogonal** de V y si, además, los vectores de S son unitarios, entonces S es una **base ortonormal** de V).

Prueba: Tome $i \in \{1, \ldots, n\}$. La independencia lineal se tiene del hecho que:

$$\mathbf{0} = \sum_{k=1}^{n} c_k \cdot u_k \Rightarrow u_i^* \cdot \mathbf{0} = c_i \cdot (u_i^* \cdot u_i) \Rightarrow \mathbf{0} = c_i \cdot \langle u_i, u_i \rangle \ \Rightarrow c_i = 0$$

Teorema

Sea $U \in \mathrm{Mat}(\mathbb{C}, n)$ **hermitiana**. Si λ es un valor propio de U, entonces $\lambda \in \mathbb{R}$.

Prueba: Sea $v \in \mathbb{C}^n$ un vector propio de U ($v \neq 0$, por definición) asociado a λ. Note que:

$$\begin{aligned}\overline{\lambda}\ \langle v, v\rangle &= \langle v, \lambda v\rangle = \langle v, Uv\rangle = (Uv)^*v = v^*U^*v \\ &= v^*Uv = \langle Uv, v\rangle = \langle \lambda v, v\rangle = \lambda\ \langle v, v\rangle\end{aligned}$$

Así, $(\overline{\lambda} - \lambda) \cdot \langle v, v\rangle\ = 0$, de donde $\overline{\lambda} = \lambda$ y de ahí se tiene el resultado.

Teorema

Sea $A \in \mathrm{Mat}(\mathbb{C}, m, n)$, entonces AA^* y A^*A son hermitianas.

Prueba: Basta ver que $(AA^*)^* = (A^*)^*A^* = AA^*$.

Teorema

Sea $A \in \mathrm{Mat}(\mathbb{C}, m, n)$. Los valores propios de A^*A y de AA^* son no negativos.

Prueba: Sea λ valor propio de A^*A (el otro caso es similar) y v un vector propio asociado a λ. Como A^*A es hermitiana, sabemos que $\lambda \in \mathbb{R}$. Luego,

$$\begin{aligned} \lambda \cdot \langle v, v\rangle &= \langle v, \lambda v\rangle = \langle v, A^*Av\rangle = (A^*Av)^*v = v^*A^*Av \\ &= (Av)^*Av = \langle Av, Av\rangle \geq 0 \end{aligned}$$

Así, $\lambda \cdot \langle v, v\rangle \geq 0$, de donde $\lambda \geq 0$.

Teorema

Sea $U \in \mathrm{Mat}(\mathbb{C}, n)$ **hermitiana**. Si λ_1 y λ_2 son autovalores de U, tales que $\lambda_1 \neq \lambda_2$, y v_1 y v_2 autovectores asociados a λ_1 y λ_2, respectivamente, entonces $v_1 \perp v_2$.

Prueba: Basta ver que:

$$\begin{aligned}
\lambda_1 \ \langle v_1, v_2 \rangle &= \langle \lambda_1 v_1, v_2 \rangle = \langle U v_1, v_2 \rangle = v_2^* U v_1 = v_2^* U^* v_1 \\
&= (U v_2)^* v_1 = \langle v_1, U v_2 \rangle = \langle v_1, \lambda_2 v_2 \rangle = \overline{\lambda_2} \ \langle v_1, v_2 \rangle \\
&= \lambda_2 \ \langle v_1, v_2 \rangle
\end{aligned}$$

Así, $(\lambda_1 - \lambda_2) \cdot \langle v_1, v_2 \rangle = 0$, de donde $v_1 \perp v_2$.

Definición de matriz unitaria

Sea $n \in \mathbb{N}$ y $U \in \mathrm{Mat}(\mathbb{C}, n)$. Se dice que U es unitaria si

$$U^* \cdot U = I_n = U \cdot U^*$$

Esta condición implica que una matriz U es unitaria si U^{-1} existe y $U^{-1} = U^*$.

Teorema

Sea $U \in \mathrm{Mat}(\mathbb{C}, n)$. U es unitaria si y solo si sus vectores columnas forman un conjunto ortonormal con el producto escalar complejo usual.

Prueba: Sea $U = (C_1\ C_2\ \ldots\ C_n)$. Note que:

$$\begin{aligned} U^*U &= \begin{pmatrix} C_1^* \\ C_2^* \\ \vdots \\ C_n^* \end{pmatrix} (C_1\ C_2\ \ldots\ C_n) = \begin{pmatrix} C_1^*C_1 & C_1^*C_2 & \ldots & C_1^*C_n \\ C_2^*C_1 & C_2^*C_2 & \ldots & C_2^*C_n \\ \vdots & & & \vdots \\ C_n^*C_1 & C_n^*C_2 & \ldots & C_n^*C_n \end{pmatrix} \\ &= \begin{pmatrix} \langle C_1, C_1\rangle & \langle C_2, C_1\rangle & \ldots & \langle C_n, C_1\rangle \\ \langle C_1, C_2\rangle & \langle C_2, C_2\rangle & \ldots & \langle C_n, C_2\rangle \\ \vdots & & & \vdots \\ \langle C_1, C_n\rangle & \langle C_2, C_n\rangle & \ldots & \langle C_n, C_n\rangle \end{pmatrix} \end{aligned}$$

De lo anterior se concluye que:

$$U \text{ unitaria} \quad \Longleftrightarrow \quad U^*U = I_n$$

$$\Longleftrightarrow \quad \begin{cases} \langle C_i, C_j \rangle \; = 0 & \text{si} \quad i \neq j \\ \langle C_i, C_i \rangle \; = \|C_i\|^2 = 1 & \text{si} \quad i \in \{1, \ldots, n\} \end{cases}$$

Esto finaliza la prueba.

Por lo tanto, las columnas de una matriz unitaria $U_{n\times n}$ forman una base ortonormal de C^n con respecto al producto escalar usual, y U **es de rango completo**.

SVD

Definición de SVD

Sean $m, n \in \mathbb{N}$ y $A \in \mathrm{Mat}(\mathbb{C}, m, n)$. Una descomposición en valores singulares de A es una factorización

$$A_{mn} = U_{mm} \cdot \Sigma_{mn} \cdot V^*_{nn},$$

donde U y V son unitarias y Σ es una matriz diagonal generalizada de entradas reales tal que $\sigma_{ij} = 0$, para $i \neq j$. Si A solo tiene entradas reales entonces la SVD se escribe:

$$A = U\Sigma V^{\mathbf{T}}$$

Como U y V^* son de rango completo (por ser unitarias y por ende invertibles), entonces $\mathrm{r}(A) = \mathrm{r}(U\Sigma V^*) = \mathrm{r}(\Sigma)$. Por ende, el rango de A (denótelo r, esto es $\mathrm{r}(A) = r$) será el número de entradas no nulas de la matriz Σ. La matriz A no tiene porqué ser de rango completo.

¿Qué forma tienen dichas matrices?

Suponiendo que dicha factorización existe para toda matriz $A \in \mathrm{Mat}(\mathbb{C}, m, n)$, se analizará qué forma deben tener las matrices U, Σ y V.

$$\begin{aligned}
\Sigma\Sigma^* &= \begin{pmatrix} \Sigma_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{m\times n} \cdot \begin{pmatrix} \Sigma_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{n\times m} \\
&= \begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \ddots & 0 & \dots & 0 \\ 0 & 0 & \sigma_r^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}_{m\times m} = \begin{pmatrix} \Sigma_r^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{m\times m}
\end{aligned}$$

¿Qué forma tienen dichas matrices?

Por otra parte,

$$A = U\Sigma V^* \quad \Rightarrow \quad A^* = V\Sigma^* U^* \Rightarrow AA^* = U\Sigma V^* V\Sigma^* U^*$$

$$\Rightarrow \quad AA^* = U\Sigma\Sigma^* U^* \Rightarrow AA^* U = U\Sigma\Sigma^*$$

Note que la igualdad anterior nos da una diagonalización de AA^*. En efecto, si llamamos u_j a las columnas de U, entonces de la última igualdad tendríamos que

$$AA^* u_j = \sigma_j^2 u_j, \;\; \forall j = 1, \ldots, m$$

Por lo tanto, las columnas de U corresponden a vectores propios de AA^* (asociados a los valores propios σ_j^2 de AA^*). A los σ_j se les denomina **valores singulares** de A.

¿Qué forma tienen dichas matrices?

Algunos hechos:

- Como AA^* es hermitiana, entonces los σ_j^2 son reales y no negativos.
- Los valores singulares no nulos de A se organizan de manera decreciente: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ y se ordenan de la misma forma en la diagonal de Σ.
- El vector propio u_i (normalizado) asociado al valor singular σ_i, se coloca como columna i de la matriz U.

¿Qué forma tienen dichas matrices?

Algunos hechos:

- Las últimas $m-r$ columnas de U se toman del núcleo de A^*, esto es, se determinan $m-r$ vectores ortonormales a partir del sistema $A^*u = \mathbf{0}$. De esta manera $u_1, \ldots, u_r, u_{r+1}, \ldots, u_m$ forman una base ortonormal para $\mathbb{C}^m$. A estos vectores se les suele llamar **vectores singulares por la izquierda**.

- Con respecto al punto anterior, en realidad debería resolverse $AA^*u = \mathbf{0}$ (es decir, analizar el núcleo de AA^*), sin embargo, note que $\mathrm{Ker}(A^*) \subset \mathrm{Ker}(AA^*)$. Por ello es suficiente con construir una base ortonormal para $\mathrm{Ker}(A^*)$.

¿Qué forma tienen dichas matrices?

Nótese que la construcción pudo haberse hecho así:

$$\begin{aligned} A = U\Sigma V^* \quad &\Rightarrow \quad A^* = V\Sigma^* U^* \Rightarrow A^*A = V\Sigma^* U^* U\Sigma V^* \\ &\Rightarrow \quad A^*A = V\Sigma^*\Sigma V^* \Rightarrow A^*AV = V\Sigma^*\Sigma \end{aligned}$$

Por lo tanto, las columnas de V corresponden a vectores propios de A^*A. Las columnas de V, pensadas como vectores de $\mathbb{C}^n$ generan una base ortonormal para este espacio y se conocen como **vectores singulares por la derecha** de A.

- AA^* y A^*A tienen los mismos valores propios no nulos, por ende no hay ambigüedad en cómo se empiece.
- Conviene elegir a A^*A o AA^*, según cual sea de menor tamaño, así el polinomio característico es de menor grado.

¿Qué forma tienen dichas matrices?

- Para construir la matriz V si **se tiene a** U note que $A = U\Sigma V^* \Rightarrow A^* = V\Sigma^* U^* \Rightarrow A^*U = V\Sigma^*$. Así, $A^*u_j = \sigma_j v_j$. Por lo tanto, para los $\sigma_j \neq 0$, $1 \leq j \leq \min\{m, n\}$, se tiene:

  $$v_j = \frac{1}{\sigma_j} A^* u_j \quad \hookrightarrow \text{ son ortogonales, } \textbf{pero hay que normalizar}$$

  Si ya se tienen todos los v_j, listo. Si no, se busca una base ortonormal de $\mathrm{Ker}(A)$ y se agregan para formar a V.

- Para construir la matriz U si **se tiene a** V note que $A = U\Sigma V^* \Rightarrow AV = U\Sigma$. Así, $Av_j = \sigma_j u_j$, de donde

  $$u_j = \frac{1}{\sigma_j} A v_j \quad \hookrightarrow \text{ son ortogonales, } \textbf{pero hay que normalizar}$$

  Si ya se tienen todos los u_j, listo. Si no, se busca una base ortonormal de $\mathrm{Ker}(A^*)$ y se agregan para formar a U.

Ejemplo #1

$$A = \begin{pmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \Rightarrow A^*A = \begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix}$$

Se calcula $P(\lambda) = |A^*A - \lambda I_2| = \lambda(\lambda - 4)$. Por ende, el único valor singular no nulo de A es $\sigma_1 = 2$ y así $r = 1$ (cantidad de valores singulares no nulos). Así,

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

Ahora resolvemos $A^*Av = 4v \iff \left\{ \begin{array}{l} 4x = 0 \\ 4y = 4y \end{array} \right.$. Por lo tanto, puede tomarse $v_1 = (0 \;\; 1)^{\mathbf{T}}$.

Ejemplo

Ahora, para las restantes $n-r$ columnas de V se se resuelve $Av = 0 \iff A\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \iff y = 0 \wedge x \in \mathbb{R}$. Por ende puede tomarse $v_2 = (1 \;\; 0)^{\mathbf{T}}$. Ahora se contruyen las $m = 3$ columnas de U. Para determinar la columna asociada al único valor singular no nulo:

$$u_1 = \frac{1}{\sigma_1} A v_1 \Rightarrow \frac{1}{2}\begin{pmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Para las dos restantes se toma una base ortonormal del núcleo de A^* (esto es suficiente porque $\mathrm{Ker}(A^*) \subset \mathrm{Ker}(AA^*)$):

$$A^* u = \mathbf{0} \iff \begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{0} \iff x = 0,\; y \in \mathbb{R}, \;\; z \in \mathbb{R}$$

Así, puede tomarse $u_2 = (0 \;\; 1 \;\; 0)^{\mathbf{T}}$ y $u_2 = (0 \;\; 0 \;\; 1)^{\mathbf{T}}$.

Finalmente,

$$A = U\Sigma V^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Ejemplo #2

$$A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} \Rightarrow AA^* = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{pmatrix} = \begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix}$$

Se calcula $P(\lambda) = |AA^* - \lambda I_2| = (25 - \lambda)(9 - \lambda)$. Por ende, los valores singulares no nulos de A son $\sigma_1 = 5$ y $\sigma_2 = 3$ y así $r = 2$. Luego,

$$\Sigma = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}$$

Ejemplo #2

Ahora resolvemos $AA^*u = 25u \iff \left\{ \begin{array}{l} y = x \\ y = x \end{array} \right.$. Por ende puede tomarse $u_1 = \frac{1}{\sqrt{2}}(1 \quad 1)^{\mathbf{T}}$. Similarmente, $AA^*u = 9u \iff \left\{ \begin{array}{l} y = -x \\ y = -x \end{array} \right.$. De donde, $u_2 = \frac{1}{\sqrt{2}}(1 \quad -1)^{\mathbf{T}}$. Por lo tanto,

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Ahora construimos las dos primeras columnas de V usando $v_j = \frac{1}{\sigma_j} A^* u_j$:

$$v_1 = \frac{1}{3\sqrt{2}} \begin{pmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{5}{3\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \frac{5}{6} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$$

Similarmente:

$$v_2 = \frac{1}{5\sqrt{2}} \begin{pmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{5\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 4 \end{pmatrix} = \frac{3}{5} \begin{pmatrix} \frac{1}{3\sqrt{2}} \\ \frac{-1}{3\sqrt{2}} \\ \frac{2\sqrt{2}}{3} \end{pmatrix}$$

Para obtener v_3 resolvemos

$$Av = \mathbf{0} \iff \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{0} \iff y = -x,\ x \in \mathbb{R}, \quad z \in \mathbb{R}$$

Por lo tanto, podemos tomar $v_3 = \frac{1}{\sqrt{3}}(1 \quad -1 \;\; 1)^{\mathbf{T}}$.

En resumen,

$$A = U\Sigma V^* = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{3\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{-1}{\sqrt{3}} \\ 0 & \frac{2\sqrt{2}}{3} & \frac{1}{\sqrt{3}} \end{pmatrix}^{\mathbf{T}}$$

Ejemplo #3

$$A = \begin{pmatrix} -2 & 1 & 1 & 2 \\ 2 & 1 & 1 & -2 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \end{pmatrix} \Rightarrow AA^* = \begin{pmatrix} 10 & -6 & 0 & 0 \\ -6 & 10 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

Se tiene que $P(\lambda) = (4-\lambda)^3(16-\lambda)$.

Ejemplo #3

Por lo tanto,

$$\Sigma = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

Con $\sigma_1^2 = 16$ queda $y = -x, \quad z = 0, \quad w = 0$ y $x \in \mathbb{R}$. Podemos tomar $u_1 = \frac{1}{\sqrt{2}}(1 \quad -1 \quad 0 \quad 0)^T$. Para $\sigma_2^2 = 4$ queda $y = x$ y $x, z, w \in \mathbb{R}$. Podemos tomar 3 vectores ortonormales del espacio propio respectivo.

$$u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad u_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad u_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Ejemplo #3

Usamos nuevamente la fórmula $v_j = \dfrac{1}{\sigma_j} A^* u_j$:

1. $v_1 = \dfrac{1}{\sigma_1} A^* u_1 = \dfrac{1}{4} A^* u_1 = (\frac{-1}{\sqrt{2}} \;\; 0 \;\; 0 \;\; \frac{1}{\sqrt{2}})^T$
2. $v_2 = \dfrac{1}{\sigma_2} A^* u_2 = \dfrac{1}{2} A^* u_2 = (0 \;\; \frac{1}{\sqrt{2}} \;\; \frac{1}{\sqrt{2}} \;\; 0)^T$
3. $v_3 = \dfrac{1}{\sigma_3} A^* u_3 = \dfrac{1}{2} A^* u_3 = (\frac{1}{2} \;\; \frac{-1}{2} \;\; \frac{1}{2} \;\; \frac{1}{2})^T$
4. $v_4 = \dfrac{1}{\sigma_4} A^* u_3 = \dfrac{1}{2} A^* u_4 = (\frac{1}{2} \;\; \frac{1}{2} \;\; \frac{-1}{2} \;\; \frac{1}{2})^T$

Ejemplo #3

En resumen,

$$
\begin{aligned}
A &= U\Sigma V^* \\
&= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} \frac{-1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & \frac{-1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{-1}{2} & \frac{1}{2} \end{pmatrix}
\end{aligned}
$$

Reducción de la SVD

Considere la descomposición en valores singulares de A dada por:

$$A_{mn} = U_{mm} \cdot \Sigma_{mn} \cdot V_{nn}^*,$$

y, como antes, $r = \mathrm{r}(A) = \mathrm{r}(\Sigma)$, que para Σ corresponde al número de entradas no nulas (o el número de valores singulares de A no nulos, contando multiplicidades). Denotemos a U por

$$U = [u_1, u_2, \ldots, u_r, u_{r+1}, \ldots, u_m]$$

donde u_i es un vector que denota a la $i-$ésima columna de U, y por ende $u_i \in \mathbb{C}^m$. En función de lo anterior, adoptemos la notación por bloques:

$$U = [U_r \mid U_{m-r}],$$

donde U_r es una matriz de tamaño $m \times r$ y U_{m-r} es de tamaño $m \times (m-r)$.

Reducción de la SVD

Análogamente, escribimos $V = [V_r \mid V_{n-r}]$, donde V_r es una matriz de tamaño $n \times r$ y V_{n-r} es de tamaño $n \times (n-r)$.

Posteriormente, usamos la siguiente representación por bloques de la matriz Σ:

$$\Sigma = \left(\begin{array}{c|c} \Sigma_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right)_{m\times n} = \left(\begin{array}{c|c} (\Sigma_r)_{r\times r} & \mathbf{0}_{r\times(n-r)} \\ \hline \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{array} \right),$$

donde Σ_r corresponde a una matriz diagonal cuadrada, cuya diagonal contiene los r valores singulares no nulos de A.

Ahora procedemos a realizar el producto $U\Sigma V^*$, usando los bloques correspondientes.

Reducción de la SVD

Primero note que:

$$U\Sigma = \left[(U_r)_{m\times r} \mid (U_{m-r})_{m\times(m-r)}\right]\left(\begin{array}{c|c} (\Sigma_r)_{r\times r} & \mathbf{0}_{r\times(n-r)} \\ \hline \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{array}\right)$$

$$= \left[(U_r\Sigma_r)_{m\times r} \mid \mathbf{0}_{m\times(n-r)}\right]$$

Luego,

$$A = U\Sigma V^* = \left[(U_r\Sigma_r)_{m\times r} \mid \mathbf{0}_{m\times(n-r)}\right]\left(\begin{array}{c} (V_r^*)_{r\times n} \\ \hline \left(V_{n-r}^*\right)_{(n-r)\times n} \end{array}\right)$$

$$= U_r\Sigma_rV_r^*$$

Reducción de la SVD

SVD reducida

Sea $A \in \mathrm{Mat}(\mathbb{C}, m, n)$. La descomposición **reducida** en valores singulares de A está dado por:

$$A = U_r \Sigma_r V_r^*,$$

donde $r = \mathrm{r}(A)$, que corresponde al número de valores singulares no nulos de A, que a su vez son los valores propios no nulos de las matrices AA^* y A^*A. Las matrices U_r, Σ_r y V_r^* son como se definieron previamente.

Esta descomposición es la que normalmente se aplica en la práctica. Entendiendo que al hacer la reducción se pierden propiedades sobre las matrices. Por ejemplo, las matrices U y V en la descomposición completa son unitarias, sin embargo, U_r y V_r no necesariamente lo son (ni siquiera tienen porqué ser cuadradas).

Reducción de la SVD

Si visualizamos a U_r como una matriz de 1 bloque-fila y 1 bloque columna (es decir, un único bloque, la matriz total) y a Σ_r como de 1 bloque-fila y r bloques-columna, entonces:

$$
\begin{aligned}
U_r\Sigma_r &= \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1r} \\ u_{21} & u_{22} & \dots & u_{2r} \\ \vdots & & \dots & \vdots \\ u_{m1} & u_{m2} & \dots & u_{mr} \end{pmatrix}_{1\times 1} \left(\begin{array}{c|c|c|c} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & \sigma_r \end{array}\right)_{1\times r} \\
&= \left(\begin{array}{c|c|c|c} \sigma_1 u_{11} & \sigma_2 u_{12} & \dots & \sigma_r u_{1r} \\ \sigma_1 u_{21} & \sigma_2 u_{22} & \dots & \sigma_r u_{2r} \\ \vdots & & & \vdots \\ \sigma_1 u_{m1} & \sigma_2 u_{m2} & \dots & \sigma_r u_{mr} \end{array}\right)_{1\times r} \\
&= [\sigma_1 u_1 \mid \sigma_2 u_2 \mid \dots \mid \sigma_r u_r]_{1\times r}
\end{aligned}
$$

Reducción de la SVD

Ahora, note que V_r^* es de dimensión $r \times n$ pues V tiene tamaño $n \times r$. Si denotamos a V por su columnas como $V = [v_1, v_2, \ldots, v_r]$, entonces V_r^* puede ser visualizada como una matriz por bloques de r bloques-fila y 1 bloque-columna. Es decir,

$$V_r^* = \begin{pmatrix} v_1^* \\ \hline v_2^* \\ \hline \vdots \\ \hline v_r^* \end{pmatrix}_{r \times 1}$$

Por ende,

$$\begin{aligned} U_r \Sigma_r V_r^* &= [\sigma_1 u_1 \mid \sigma_2 u_2 \mid \ \ldots \ \mid \sigma_r u_r]_{1 \times r} \cdot \begin{pmatrix} v_1^* \\ \hline v_2^* \\ \hline \vdots \\ \hline v_r^* \end{pmatrix}_{r \times 1} \\ &= \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \ldots + \sigma_r u_r v_r^* \end{aligned}$$

Ahora una aplicación...