Descomposición Espectral y SVD

Alan Reyes-Figueroa

Métodos Numéricos II (Aula 02) 12.julio.2021

Descomposición Espectral

Definición

*Una **descomposición espectral** de una matriz cuadrada $A \in \mathbb{R}^{n\times n}$ es una factoración de la forma*

$$A = X\Lambda X^{-1},$$

donde $X \in \mathbb{R}^{n\times n}$ es no singular, y $\Lambda \in \mathbb{R}^{n\times n}$ es una matriz diagonal, cuyas entradas $\lambda_1, \ldots, \lambda_n$ corresponden a los autovalores de A.

La definición anterior puede escribirse como $AX = X\Lambda$, esto es

$$\begin{pmatrix} & & \\ & A & \\ & & \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} = \begin{pmatrix} \lambda_1\mathbf{x}_1 & \lambda_2\mathbf{x}_2 & \ldots & \lambda_n\mathbf{x}_n \end{pmatrix}.$$

La descomposición espectral expresa un cambio de base en las coordenadas dadas por los autovectores de A.

UNIVERSIDAD DEL VALLE DE GUATEMALA

Descomposición Espectral

El conjunto de autovectores correspondientes a un solo autovalor λ de A, junto con el vector cero, forma un subespacio de E_λ conocido como el **autoespacio asociado a** λ.

Obs! $E_\lambda \subseteq \mathbb{C}^n$. Es un subespacio invariante: $AE_\lambda \subseteq E_\lambda$.

Definición

La ***multiplicidad geométrica*** *de* λ *es* $\dim_{\mathbb{R}} E_\lambda = \dim_{\mathbb{R}} \operatorname{Ker}(\lambda I - A)$.

Recordemos que el polinomio característico de $A \in \mathbb{R}^{n\times n}$ es el polinomio de grado n dado por $p_A(z) = \det(zI - A)$.

Teorema

λ *es un autovalor de* A *si, y sólo si,* $p_A(\lambda) = 0$.

Descomposición Espectral

Prueba:

$$\begin{aligned} \lambda \text{ es autovalor de } A \quad &\Leftrightarrow \quad \text{existe } \mathbf{x} \neq \mathbf{0}: \ A\mathbf{x} = \lambda\mathbf{x} \ \Leftrightarrow \ \lambda\mathbf{x} - A\mathbf{x} = \mathbf{0} \\ &\Leftrightarrow \quad \lambda I - A \text{ es singular} \ \Leftrightarrow \ \det(\lambda I - A) = 0. \ \square \end{aligned}$$

Por el Teorema Fundamental del Álgebra, una consecuencia del teorema anterior es que

$$p_A(z) = (z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_n),$$

donde los λ_j son todos autovalores de A.

Definición

*La **multiplicidad algebraica** del autovalor λ es su multiplicidad como raíz del polinomio característico $p_A(z)$.*

Descomposición Espectral

Observaciones:

- Toda matriz cuadrada $A \in \mathbb{R}^{n\times n}$ posee n autovalores (contados con multiplicidad).
- Si las raíces de $p_A(z)$ son todas simples, A posee n autovalores distintos.
- Cuando X es no singular, las matrices A y XAX^{-1} poseen igual polinomio característico, autovalores, y multiplicidades algebraicas y geométricas, ya que

$$p_{XAX^{-1}}(z) = \det(zI - XAX^{-1}) = \det\left(X(zI-A)X^{-1}\right) = \det X \det(zI-A)(\det X)^{-1} = \det(zI-A) = p_A(z).$$

y E_λ es autoespacio para $A \;\Leftrightarrow\; X^{-1}E_\lambda$ es autoespacio para XAX^{-1}.

Teorema

multiplicidad algebraica $\lambda \geq$ *multiplicidad geométrica* λ, $\forall\lambda$ *autovalor.*

Descomposición Espectral

Prueba: Sea k la multiplicidad geométrica de λ para la matriz $A \in \mathbb{R}^{n\times n}$. Formamos una matriz $V \in \mathbb{R}^{n\times k}$ cuyas k columnas constituyen una base ortonormal del autoespacio $E_\lambda = \{\mathbf{x} : A\mathbf{x} = \lambda\mathbf{x}\}$. Extendemos V a una matriz unitaria cuadrada $V \in \mathbb{R}^{n\times n}$, y obtenemos V^*AV en la forma

$$B = V^*AV = \begin{pmatrix} \lambda I_k & C \\ \mathbf{0} & D \end{pmatrix},$$

donde $C \in \mathbb{R}^{k\times(n-k)}$, $D \in \mathbb{R}^{(n-k)\times(n-k)}$.

Luego, $\det(zI - B) = \det(zI - \lambda I_k)\,\det(zI - D) = (z-\lambda)^k \det(\lambda I - D)$.

De ahí que la multiplicidad algebraica de λ como autovalor de B es al menos k, y dado que las transformaciones de semejanza preservan multiplicidades, lo mismo es cierto para A. $\square$

Descomposición Espectral

Definición

*Un autovalor cuya multiplicidad algebraica excede su multiplicidad geométrica es un **autovalor defectuoso**. Una matriz que tiene uno o más autovalores defectuosos se llama una **matriz defectuosa**.*

- Las matrices diagonales no son defectuosas.
- Veremos luego que las matrices simétricas no son defectuosas.

Definición

*Una matriz cuadrada $A \in \mathbb{R}^{n\times n}$ es **diagonalizable** si admite una descomposición espectral.*

Teorema

Una matriz $A \in \mathbb{R}^{n\times n}$ es diagonalizable si, y sólo si, no es defectuosa.

Descomposición Espectral

Prueba: ($\Rightarrow$) Dada una descomposición espectral $A = X\Lambda X^{-1}$ para A, $\Rightarrow \ \Lambda$ es similar a A, con los mismos autovalores y las mismas multiplicidades. Como Λ es diagonal, no es defectuosa y, por lo tanto, lo mismo vale para A.

($\Leftarrow$) Una matriz no defectuosa debe tener n autovectores l.i. (ya que autovectores de diferente autovalor son l.i., y cada cada autovalor contribuye exactamente con tantos autovectores l.i. como su multiplicidad geométrica). Si estos n autovectores l.i. se forman en las columnas de una matriz X, entonces X es no singular y tenemos $A = X\Lambda X^{-1}$. $\square$

Ejemplo: Considere las matrices

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \qquad \text{y} \qquad B = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

Tanto A como B tienen un polinomio característico $(z-2)^3$, por lo que hay un solo autovalor $\lambda = 2$ de multiplicidad algebraica 3.

En el caso de A, el autoespacio $E_\lambda = \langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle$, mientras que para B, $E_\lambda = \langle \mathbf{e}_1 \rangle$, de modo que λ es un autovalor defectuoso para B.
Portanto, A es diagonalizable, pero B no.

(Ver libro *Linear Algebra* de Hoffman, Kunze).

Descomposición Espectral

En ocasiones, no sólo ocurre que una matriz $A \in \mathbb{R}^{n\times n}$ admite n autovectores l.i., sino que éstos son ortogonales.

Definición

Decimos que A es ***unitariamente diagonalizable*** *si existe una matriz ortogonal U tal que $A = U\Lambda U^T$.*

Teorema (Teorema espectral / Descomposición espectral)

Sea $A \in \mathbb{R}^{d\times d}$ una matriz simétrica (operador auto-adjunto). Entonces, A admite una descomposición de la forma

$$A = U\Lambda U^T,$$

donde $\Lambda = diag(\lambda_1, \lambda_2, \dots, \lambda_d)$ es la matriz diagonal formada por los autovalores $\lambda_1 \geq \lambda_2 \geq \dots \lambda_d$ de A, y

$$U = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_d \end{pmatrix} \in \mathbb{R}^{d \times d}$$

es una matriz ortogonal cuyas columnas son los autovectores de A, con $\mathbf{u}_i$ el autovector correspondiente a λ_i, $i = 1, 2, \dots, d$.

En otras palabras, A es una suma de matrices de rango 1: $A = \sum_{i=1}^{d} \lambda_i \, \mathbf{u}_i \mathbf{u}_i^T$.

Comentarios:

- El teorema espectral dice que toda matriz cuadrada A, real, simétrica, posee una base ortonormal, formada por autovectores de A.
- Para $1 \leq k \leq d$, la suma $A = \sum_{i=1}^{k} \lambda_i \, \mathbf{u}_i \mathbf{u}_i^T$, es una matriz de rango k.

Descomposición Espectral

Observaciones:

- A simétrica y semi-definida positiva, $\Rightarrow$ existe $A^{1/2}$ tal que $A^{1/2}A^{1/2} = A$.
- Si todos los autovalores de A son no-negativos, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d \geq 0$, entonces $\Lambda^{1/2}$ existe y
$$\Lambda^{1/2} = diag(\lambda_1, \lambda_2, \ldots, \lambda_d)^{1/2} = diag(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_d^{1/2}).$$
- A partir de la descomposición espectral podemos calcular $A^{1/2}$. De hecho, si $A = U\Lambda U^T$, definimos $A^{1/2} = U\Lambda^{1/2}U^T$, y
$$\begin{aligned} A^{1/2}A^{1/2} &= (U\Lambda^{1/2}U^T)(U\Lambda^{1/2}U^T) = U\Lambda^{1/2}(U^TU)\Lambda^{1/2}U^T \\ &= U\Lambda^{1/2}\Lambda^{1/2}U^T = U\Lambda U^T = A. \end{aligned}$$
- En general $A \succeq 0 \ \Rightarrow A^p = U\Lambda^p U^T$, para todo $p \in \mathbb{R}$.

<u>Prueba</u>: (Teorema espectral).
La prueba es por inducción sobre $n = \dim A$. Para $n = 1$, el resultado es inmediato pues

$$A = [a] = [1][a][1] = [1][a][1]^T.$$

Suponga que el resultado es válido para $n - 1$. Mostramos que es posible extenderlo a n. Para $n > 1$, sea $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ una base ortonormal cualquiera para $\mathbb{R}^n$, y representamos un vector $\mathbf{x} \in \mathbb{R}^n$ arbitrario por

$$\mathbf{x} = \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \ldots + \alpha_n\mathbf{u}_n.$$

Podemos pensar $\mathbf{w} : \mathbb{R}^n \to \mathbb{R}^n$ como una función de las coordenadas, esto es $\mathbf{w} = \mathbf{w}(\alpha_1, \ldots, \alpha_n)$.
Consideramos el problema de optimización:

$$\text{maximizar } \langle \mathbf{w}, A\mathbf{w} \rangle = \mathbf{w}^T A\mathbf{w}, \quad \text{sujeto a} ||\mathbf{w}||_2 = \mathbf{w}^T\mathbf{w} = 1. \tag{1}$$

Teorema Espectral

El lagrangiano de este problema es $\mathcal{L}(\mathbf{w}) = \mathbf{w}^T A\mathbf{w} - \lambda(\mathbf{w}^T\mathbf{w} - 1)$. Del criterio de optimalidad, resulta

$$\nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w},\lambda) = 2A\mathbf{w} - 2\lambda\mathbf{w} = 0, \quad \nabla_{\lambda}\mathcal{L}(\mathbf{w},\lambda) = \mathbf{w}^T\mathbf{w} - 1 = 0. \tag{2}$$

La primera condición en (2) implica que $A\mathbf{w} = \lambda\mathbf{w}$, de modo que el $\mathbf{w}^*$ óptimo debe ser un autovector de A, y λ su autovalor. La segunda condición implica que $||\mathbf{w}^*||_2 = 1$.
De hecho $\max \mathbf{w}^T A\mathbf{w} = \max \mathbf{w}^T(\lambda\mathbf{w}) = \max \lambda||\mathbf{w}||_2^2 = \max \lambda$, implica que $\mathbf{w}^*$ corresponde al autovector de λ_1, el mayor autovalor de A.

Hallado el máximo $\mathbf{w}^*$, hacemos $\mathbf{u}_1 = \frac{\mathbf{w}}{||\mathbf{w}||_2}$ y $\lambda_1 = \lambda$. Consideramos ahora el subespacio $W_2 = (\langle\mathbf{w}\rangle)^\perp \equiv \mathbb{R}^{n-1}$. Como $A : \langle\mathbf{w}\rangle \to \langle\mathbf{w}\rangle$, entonces también $A : \langle\mathbf{w}\rangle^\perp \to \langle\mathbf{w}\rangle^\perp$

Teorema Espectral

Esto se debe a que si $\mathbf{x} \in \langle \mathbf{w} \rangle^\perp$, entonces
$\langle \mathbf{x}, \mathbf{w} \rangle = 0 \;\Rightarrow\; \langle A\mathbf{x}, \mathbf{w} \rangle = \langle \mathbf{x}, A\mathbf{w} \rangle = \langle \mathbf{x}, \lambda\mathbf{w} \rangle = \lambda \langle \mathbf{x}, \mathbf{w} \rangle = 0$, así $A\mathbf{x} \in \langle \mathbf{w} \rangle^\perp$.

Por inducción, el espacio $(n-1)$-dimensional W_0 tiene una base ortonormal $\{\mathbf{u}_2, \ldots, \mathbf{u}_n\}$ formada por autovectores de A, cada uno con un autovalor real, y por construcción $\mathbf{u}_1$ es un vector unitario ortogonal a cada uno de $\{\mathbf{u}_2, \ldots, \mathbf{u}_n\}$.

Entonces $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ es una base ortonormal para $\mathbb{R}^n$ formada por autovectores de A, lo que prueba el teorema. $\square$

Descomposición SVD

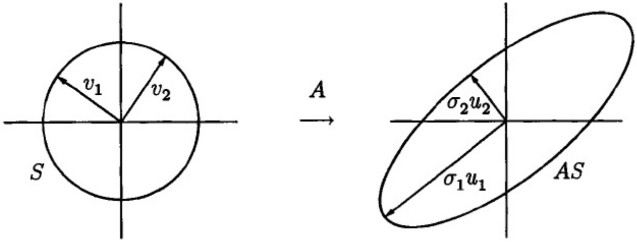
La descomposición en valores singulares, SVD, está motivada por el siguiente hecho geométrico:

La imagen de la esfera unitaria bajo cualquier matriz en $\mathbb{R}^{m\times n}$ es una hiperelipse en $\mathbb{R}^m$, obtenida al "estirar" la esfera unitaria en $\mathbb{R}^m$ por algunos factores $\sigma_1, \ldots, \sigma_m$, (posiblemente cero) en algunas direcciones ortogonales $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbb{R}^m$.

Por conveniencia, consideramos que los $\mathbf{u}_i$ son vectores unitarios, es decir, $||\mathbf{u}_i||_2 = 1$. Los vectores $\{\sigma_i \mathbf{u}_i\}$ son los semi-ejes principales de la hiperelipse, con longitudes $\sigma_1, \ldots, \sigma_m$. Si A tiene rango r, exactamente r de estas longitudes resultarán ser distintas de cero, y en particular, si $m > n$, como máximo n de ellas serán distinto de cero.

UVG UNIVERSIDAD DEL VALLE DE GUATEMALA

Descomposición SVD



SVD de una matriz en $\mathbb{R}^{2\times 2}$.

Sea $S = S^{n-1}$ la esfera unitaria en $\mathbb{R}^n$, y tome cualquier matriz $A \in^{m\times n}$ con $m > n$. Por simplicidad, supongamos que A tiene rango completo n. La imagen AS es una hiperelipse en $\mathbb{R}^m$.

Definimos algunas propiedades de A en términos de la forma de AS.

Descomposición SVD

Primero, definimos los n **valores singulares** de A. Estos son las longitudes de los n semi-ejes principales de AS, escritos $\sigma_l \geq \sigma_2 \geq \ldots \geq \sigma_n > 0$.

Los n **vectores singulares izquierdos** de A. Estos son los vectores unitarios $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n n\}$ orientados en las direcciones de los semi-ejes principales de AS, numerados para corresponder con los valores singulares. Así, $\sigma_i \mathbf{u}_i$ es el i-ésimo mayor semi-eje principal de AS.

Finalmente, los n **vectores singulares derechos** de A son los vectores unitarios $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subset S$ que son las preimágenes de los semi-ejes principales de AS, numerados de modo que $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$, $i = 1, 2, \ldots, n$.

De lo anterior tenemos una relación entre los σ_i, los $\mathbf{u}_i$ y los $\mathbf{v}_i$:

Descomposición SVD

$$A\mathbf{v}_j = \sigma_j \mathbf{u}_j, \ \forall\, 1 \leq j \leq n. \tag{3}$$

Esta colección de ecuaciones vectoriales se puede expresar como una ecuación matricial, como

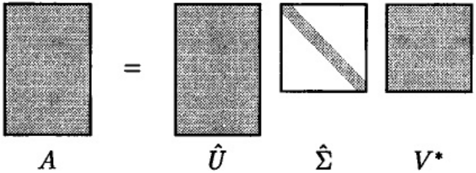
$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \left[\begin{array}{c|c|c|c} v_1 & v_2 & \cdots & v_n \end{array}\right] = \left[\begin{array}{c|c|c|c} u_1 & u_2 & \cdots & u_n \end{array}\right] \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}$$

o de forma más compacta como $AV = \widehat{U}\widehat{\Sigma}$, con $\widehat{U} \in \mathbb{R}^{m\times n}$, $V, \widehat{\Sigma} \in \mathbb{R}^{n\times n}$. $V \in O(n)$ es ortogonal, mientras que $\widehat{U}$ posee columnas ortogonormales.

Como V es ortonormal, podemos escribir $A = \widehat{U}\widehat{\Sigma}V^T$.

Definición

La factoración $A = \widehat{U}\widehat{\Sigma}V^T$ se llama la ***descomposición en valores singulares reducida*** *de A.*



Descomposición SVD reducida, (para $m \geq n$).

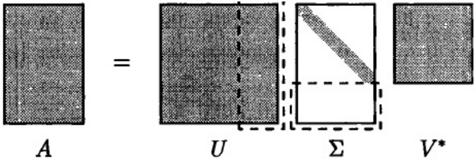
Introducimos una segunda factoriación SVD. La idea es la siguiente. Las columnas de $\widehat{U}$ son n vectores ortonormales en el espacio $\mathbb{R}^m$. A menos que $m = n$, no forman una base de $\mathbb{R}^m$, ni $\widehat{U}$ es una matriz ortogonal.

Adjuntamos $m - n$ columnas ortonormales adicionales, de modo que $\widehat{U}$ se extiende a una matriz ortogonal U. Si $\widehat{U}$ se reemplaza por U en (3), $\widehat{\Sigma}$ también cambia. Para que el producto permanezca inalterado, las últimas $m - n$ columnas de U deben multiplicarse por cero. En consecuencia, sea Σ la matriz $m \times n$ que consta de $\widehat{\Sigma}$ en el bloque $n \times n$ superior junto con $m - n$ filas de ceros a continuación.

Ahora tenemos una nueva factoración de la forma $AV = U\Sigma$ ó $A = U\Sigma V^T$, donde $\widehat{U} \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^n$ y $\Sigma \in \mathbb{R}^{m \times n}$. Además, $V \in O(n)$ y $U \in O(m)$ son matrices ortogonales.

Definición

*La factoración $A = U\Sigma V^T$ se llama la **descomposición en valores singulares completa** de A.*



Descomposición SVD completa, (para $m \geq n$).

Habiendo descrito la SVD completa, ahora podemos descartar el supuesto inicial que A tiene rango completo. Si A es deficiente en rango, i.e. $\text{rank}(A) = r < \min\{m, n\}$, la factoración SVD completa aún es válida.

Descomposición SVD

Todo lo que cambia es que ahora no n sino solo r de los vectores singulares izquierdos de A están determinados por la geometría de la hiperelipse. Al construir la matriz ortogonal U, añadimos $m - r$ en lugar de $m - n$ columnas ortonormales adicionales. La matriz V también necesitará $n - r$ columnas ortonormales adicionales para extender las r columnas determinadas por la geometría.

La matriz Σ ahora tendrá r entradas diagonales positivas, con las $n - r$ restantes igual a cero. De la misma manera, la SVD reducida también tiene sentido para matrices A de rango deficiente. Se puede tomar $\widehat{U}$ como $m \times n$, con $\widehat{\Sigma}$ de dimensiones $n \times n$ con algunos ceros en la diagonal, o comprimir aún más la representación de modo que $\widehat{U}$ sea $m \times r$ y $\widehat{\Sigma}$ sea $r \times r$ y estrictamente positiva en la diagonal.

Teorema (Existencia y unicidad SVD)

Toda matriz $A \in \mathbb{C}^{m\times n}$ posee una descomposición en valores singulares $A = U^*$. Además, los valores singulares $\{\sigma_j\}$ se determinan unívocamente y, si A es cuadrada y los σ_j son distintos, los vectores singulares izquierdos y derechos $\{\mathbf{v}_j\}$ y $\{\mathbf{u}_j\}$ se determinan de forma única hasta signos complejos (es decir, factores escalares complejos de módulo 1).

Prueba. Para probar la existencia de la SVD, procedemos por inducción sobre la dimensión de A.

Hagamos $\sigma_1 = ||A||_2$. Como $S = \{\mathbf{x} \in \mathbb{R}^n : ||\mathbf{x}||_2 = 1\}$ es compacto, existe un vector $\mathbf{v}_1 \in \mathbb{R}^n$ con $||\mathbf{v}_1||_2 = 1$, y $||\mathbf{u}_1||_2 = \sigma_1$, tales que $\mathbf{u}_1 = A\mathbf{v}_1$.

Extendemos $\mathbf{v}_1$ a una base ortonormal $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ de $\mathbb{R}^n$, y $\mathbf{u}_1$ a una base

Descomposición SVD

$\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ ortonormal de $\mathbb{R}^m$. Sean U_1 y V_1 las matrices ortogonales con columnas $\mathbf{u}_i$ y $\mathbf{v}_i$, respectivamente.

Tenemos
$$U^T \Sigma V = S = \begin{pmatrix} \sigma_1 & \mathbf{w}^T \\ \mathbf{o} & B \end{pmatrix},$$
donde $\mathbf{w}^T$ es un vector fila, $\mathbf{o}$ es un vector columna y $B \in \mathbb{R}^{(m-1)\times(n-1)}$. Además
$$\left|\left| \begin{pmatrix} \sigma_1 & \mathbf{w}^T \\ \mathbf{o} & B \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \mathbf{w} \end{pmatrix} \right|\right|_2 \geq \sigma_1^2 + \mathbf{w}^T\mathbf{w} = (\sigma_1^2 + \mathbf{w}^T\mathbf{w})^{1/2} \left|\left| \begin{pmatrix} \sigma_1 \\ \mathbf{w} \end{pmatrix} \right|\right|_2,$$
de modo que $||S||_2 \geq (\sigma_1^2 + \mathbf{w}^T\mathbf{w})^{1/2}$.
Como U_1 y V_1 son ortogonales, sabemos que $||S||_2 = ||A||_2 = \sigma_1$, y esto implica que $\mathbf{w} = \mathbf{o}$.

Descomposición SVD

Si $n = 1$ ó $m = 1$, acabó la prueba. De lo contrario, la submatriz B describe la acción de A en el subespacio ortogonal a $\langle \mathbf{v}_1 \rangle$, un espacio de dimensión $n - 1$.

Por la hipótesis de inducción, B admite una descomposición SVD de la forma $B = U_2 \Sigma_2 V_2^T$, y se verifica que

$$A = U_1 \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{pmatrix} \begin{pmatrix} \sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & V_2 \end{pmatrix}^T V_1^T.$$

es una descomposición SVD para A, como queríamos. Esto muestra la existencia.

Para la unicidad, leer el final del capítulo 4, libro de Trefethen. $\square$

Cambio de Bases

La SVD nos permite decir que cada matriz es diagonal, desde que se utilicen las bases adecuadas para los espacios de dominio y rango. DAdo cualquier $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{b}$ puede expandirse en la base de los vectores singulares izquierdos de A (columnas de U), y cualquier $\mathbf{x} \in \mathbb{R}^n$ puede representarse en la base de los vectores singulares derechos de A (columnas de V).

Los vectores de coordenadas para estas expansiones son $\mathbf{b}' = U^T\mathbf{b}$ y $\mathbf{x}' = V^T\mathbf{x}$. De $A = U\Sigma V^T$, la relación $\mathbf{b} = A\mathbf{x}$ se puede expresar en términos de $\mathbf{b}'$ y $\mathbf{x}'$ como

$$\mathbf{b} = A\mathbf{x} \;\Leftrightarrow\; U^T\mathbf{b} = U^TA = U^TU\Sigma V^T\mathbf{x} \;\Leftrightarrow\; \mathbf{b}' = \Sigma\mathbf{x}'.$$

Así, A se reduce a la matriz diagonal Σ cuando el rango se expresa en la base de columnas de U y el dominio en la base de columnas de V.

SVD y Descomposición Espectral

Una matriz cuadrada A de rango completo puede expresarse como una matriz diagonal de autovalores Λ, si el rango y el dominio están representados en una base de autovectores.

Relación entre la SVD y la Descomposición espectral:

Sea $A = U\Sigma V^T$ la SVD de A. Entonces

$$A^TA = (U\Sigma V^T)^T(U\Sigma V) = V\Sigma(U^TU)\Sigma V^T = V\Sigma^2V^T = V\Lambda V^T$$

$\Rightarrow$ las columnas de V son los autovectores de A^TA,
$\Rightarrow$ los valores singulares σ_i son las raíces de los autovalores λ_i de A^TA.

$$AA^T = (U\Sigma V^T)(U\Sigma V)^T = U\Sigma(V^TV)\Sigma U^T = U\Sigma^2U^T = U\Lambda U^T$$

$\Rightarrow$ las columnas de U son los autovectores de AA^T,
$\Rightarrow$ los valores singulares σ_i son las raíces de los autovalores λ_i de AA^T.

Propiedades a partir de la SVD

Sea $A = U\Sigma V^T$ su descomposición en valores singulares.
Las pruebas de los resultados siguientes están en el Capítulo 5 de Trefethen.

Teorema

El rango $\text{rank}(A)$ *es* r*, el número de valores singulares no-nulos.*

Teorema

$\text{Im}(A) = \langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle$ *y* $\text{Ker}(A) = \langle \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n \rangle$.

Teorema

$||A||_2 = \sigma_1$ *y* $||A||_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_r^2}$.

UVG UNIVERSIDAD DEL VALLE DE GUATEMALA

Propiedades a partir de la SVD

Teorema

Los valores singulares no-nulos de A son las raíces de los autovalores de A^TA o de AA^T. Esto es $\sigma_i = \sqrt{\lambda_i}$, donde $\Lambda = (\lambda_1, \ldots, \lambda_r, \ldots)$ corresponde a la diagonal en la descomposición espectral de A^TA o de AA^T.

Teorema

Si $A = AT$, los valores singulares de A son los valores absolutos $\sigma_i = |\lambda_i|$ de los autovalores de A.

Teorema

Para $A \in \mathbb{R}^{n\times n}$, vale $\det(A) = \prod_{i=1}^{n} \sigma_i$.

UVG UNIVERSIDAD DEL VALLE DE GUATEMALA

Aproximaciones de bajo rango

Teorema (Eckart-Young)

Sea $A \in \mathbb{R}^{n\times d}$, $n \geq d$, una matriz cuya descomposición SVD está dada por

$$A = USV^T = \sum_{i=1}^{d} \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

Entonces, la matriz $\widehat{A}_r$ de rango r, $1 \leq r \leq d$, que mejor aproxima A en el sentido de minimizar

$$\min_{rank\ \widehat{A}_r \leq r} ||A - \widehat{A}_r||_F^2$$

se obtiene de truncar la descomposición en valores dingulares de A:

$$\widehat{A}_r = U_r S_r V_r^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T,$$

Teorema (Eckart-Young)

donde

$$U_r = [\mathbf{u}_1\ \mathbf{u}_2\ \ \dots\ \ \mathbf{u}_r],\ \ S_r = diag(\sigma_1, \sigma_2 \dots, \sigma_r),\ \ V_r = [\mathbf{v}_1\ \mathbf{v}_2\ \ \dots\ \ \mathbf{v}_r].$$

En ese caso, el error de aproximación está dado por

$$||A - \widehat{A}_r||_F^2 = \sum_{i=r+1}^{d} \lambda_i,$$

o

$$||A - \widehat{A}_r||_2^2 = \lambda_{r+1}.$$

Descomposición de Schur

Otra factorización matricial, en realidad la que es más útil en análisis, es la llamada factoración de Schur.

Definición

Una ***factoración de Schur*** *de una matriz $A \in \mathbb{R}^{n \times n}$ es una factorización de la forma*

$$A = QTQ^T,$$

donde Q es ortogonal (unitaria) y T es triangular superior.

Observe que A y T son semejantes. De manera similar, los autovalores de A necesariamente aparecen en la diagonal de T.

Teorema

Toda matriz cuadrada A posee una factoración de Schur.

Descomposición de Schur

Prueba: Por inducción sobre la dimensión n de A.

El caso $n = 1$ es directo, ya que $A = [a] = [1][a][1]^T$, así que suponga que $n \geq 2$. Sea $\mathbf{x} \in \mathbb{R}^n$ cualquier autovector de A, con autovalor correspondiente λ.
Tomamos $\mathbf{u}_1 = \frac{\mathbf{x}}{||\mathbf{x}||_2}$ normalizado, y lo hacemos la primer columna de la matriz U. Entonces, al igual que en la descomposición espectral, vale

$$U^TAU = \begin{pmatrix} \lambda & B \\ \mathbf{0} & C \end{pmatrix},$$

con $B \in \mathbb{R}^{n-1}$, $C \in \mathbb{R}^{(n-1)\times(n-1)}$.
De la hipótesis inductiva, existe una factoración de Schur VTV^T para C.
Escribimos entonces

Descomposición de Schur

$$Q = U \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & V \end{pmatrix}.$$

Entonces, Q es una matriz ortogonal (unitaria), pues $QQ^T = U \left(\begin{smallmatrix} 1 & 0 \\ 0 & VV^T \end{smallmatrix}\right) U^T = UU^T = I$, y tenemos

$$Q^TAQ = \begin{pmatrix} 1 & 0 \\ 0 & V^T \end{pmatrix} U^TAU \begin{pmatrix} 1 & 0 \\ 0 & V \end{pmatrix} = \begin{pmatrix} \lambda & BV \\ \mathbf{0} & T \end{pmatrix} \Rightarrow A = Q \begin{pmatrix} \lambda & BV \\ \mathbf{0} & T \end{pmatrix} Q^T. \square$$

En resumen:

- Una diagonalización $A = X\lambda X^{-1}$ existe, si y sólo si, A es no defectuosa.
- Una diagonalización unitaria existe, si y sólo si, A es **normal** (esto es, $AA^* = A^*A$).
- Una descomposición SVD $A = U\Sigma V^T$ siempre existe.
- Una descomposición de Schur $A = QTQ^T$ siempre existe.