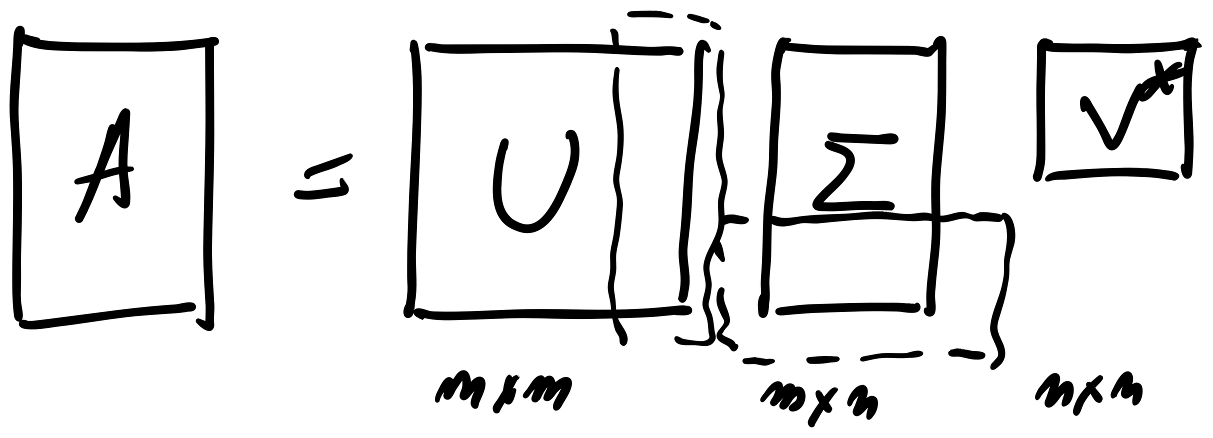
ALN
Clase 3
30/8/2022

DESCOMPOSICIÓN EN VALORES SINGULARES II

Hemos probado que toda matriz $A \in \mathbb{C}^{m\times n}$ se puede escribir como $A = U \Sigma V^*$ con $U \in \mathcal{U}(m)$, $V \in \mathcal{U}(n)$ y $\Sigma \in \mathbb{C}^{m\times n}$ "diagonal".

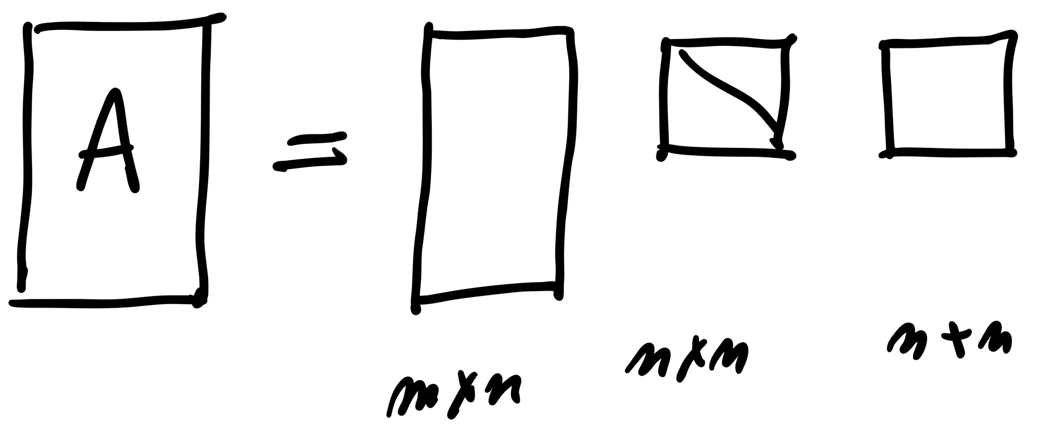
Vimos dos versiones



Full SVD

$A v_i = \sigma_i u_i$

↳ (cambio de notación respecto a clase anterior, ... mal yo!)



Una forma de interpretar la SVD es que se pueden elegir bases ortonormales en el dominio y codominio de manera tal que el operador es diagonal.

$Ax = b$, $\underset{\in\,\mathbb{C}^n}{x} \overset{A}{\longmapsto} \underset{\in\,\mathbb{C}^m}{b}$ podemos expresarlo en las bases adecuadas y ver que efectivamente es diagonal.

Escribimos $x \in \mathbb{C}^n$ en la base generada por columnas de V, i.e. $x' = V^* x$ y hacemos lo mismo con $b \in \mathbb{C}^m$, expresado en la base de generada por columnas de U, i.e. $b' = U^* b$.

Luego tenemos $b = Ax \iff b' = U^* A x = U^* A (V V^*) x = \Sigma x'$

i.e. $b' = \Sigma x'$

SVD vs EVD (eigenvalue decomposition)

Vale mencionar primero que esta "pelea" solo tiene sentido si $m=n$, dado que EVD solo tiene sentido cuando el operador A "opera" en el mismo espacio: $A : \mathbb{C}^n \to \mathbb{C}^n$.

Asumiendo $m=n$, y que A es diagonalizable podemos escribir

$$A = X \Lambda X^{-1}$$ donde columnas de X son vectores propios y Λ es la diagonal de valores propios. Más precisamente,

$$X = (v_1 | \cdots | v_n) \quad , \quad A v_i = \lambda_i v_i \quad (i=1,\ldots,n)$$

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) .$$

Si tomamos $\{v_1, \ldots, v_n\}$ como base de $\mathbb{C}^n$ y escribimos los vectores en esta base (como hicimos en sección anterior) tenemos

$$x' = X^{-1} x \ , \quad b' = X^{-1} b \quad \text{luego} \quad b = Ax \iff b' = \Lambda x'.$$

O sea A opera también como un operador diagonal, PERO: en la misma base!!

En particular, esto implica que el comportamiento de iterar A varias veces, A^k, es fácilmente analizado via el espectro.

$$A^k = (X \Lambda X^{-1}) \cdots (X \Lambda X^{-1}) = X \Lambda^k X^{-1} \quad (A^k v_i = \lambda_i^k v_i)$$

Sin embargo la SVD no nos ayuda a entender la dinámica A^k, ~~pero~~ como veremos la SVD sí ~~da~~ información muy valiosa sobre la "geometría" de A, o ~~más~~ menos rayo, la SVD encapsula la información relevante del operador A.

Para ser más precisos, veamos algunas propiedades.

Propiedades de Matrices vía SVD

$$A = (u_1|\cdots|u_m)\begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_k & \\ & & & 0 \end{pmatrix}(v_1|\cdots|v_n)^*$$

- Rango: El rango de A es el nº de valores singulares no nulos.

  Es fácil ver que como el rango de U y V son maximales resulta que el rango de A es el mismo que el rango de Σ.

- Imagen y núcleo: $\operatorname{Im}(A) = \langle u_1, \dots, u_k \rangle$, $\operatorname{Ker} A = \langle v_1, \dots, v_k \rangle$ siendo $k = rk(A)$.

- Normas $\|A\|_2$ y $\|A\|_F$: Resulta $\|A\|_2 = \|\Sigma\|_2 = \sigma_1$

  $$\|A\|_F = \|\Sigma\|_F = (\sigma_1^2 + \dots + \sigma_k^2)^{1/2}$$

- Caso Hermítico:

  $A^* = A$ entonces $|\lambda_i|$, $\lambda_i \in Sp(A)$ son los valores singulares.

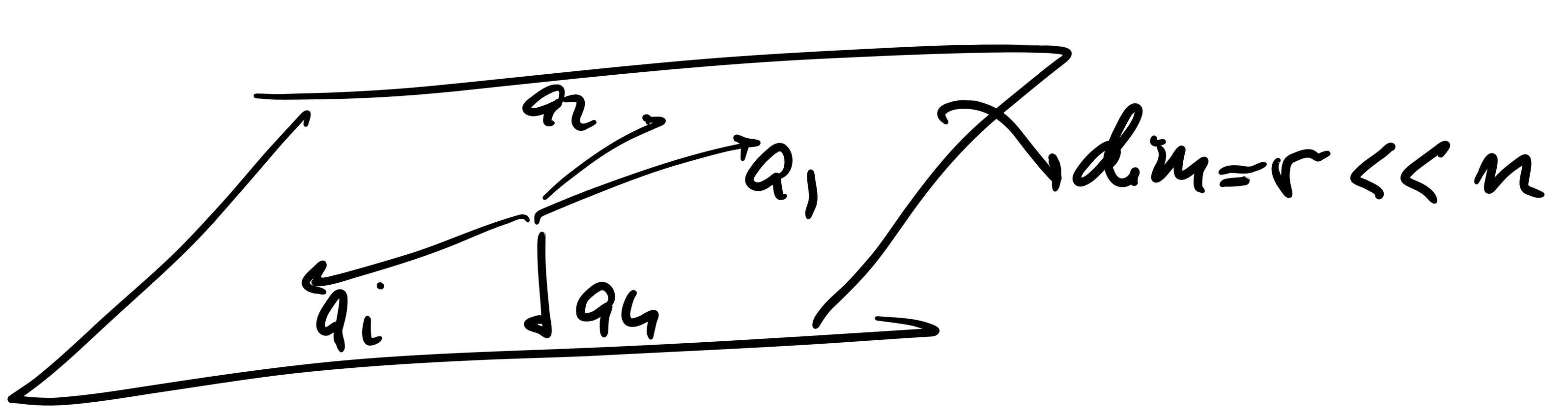
- Determinante: $A \in \mathbb{C}^{n\times n} \Rightarrow |\det A| = \prod_{i=1}^{n} \sigma_i$

  Esto resulta de que si $U \in \mathcal{U}(n) \Rightarrow U^* U = Id$ y por lo tanto $1 = \det(U^* U) = \det U^* \cdot \det U = \overline{\det U}\, \det U = |\det U|^2$

LOW-RANK Approximation

He aquí una de las ~~propiedades~~ aplicaciones más importantes de SVD.

Supongamos que tenemos una matriz $A \in \mathbb{C}^{m\times n}$, $A = (a_1|\cdots|a_m)$ de rango $\underline{r}$, i.e $\operatorname{Im}(A) = \langle a_1, \dots, a_n \rangle$ es un espacio de dimensión $\underline{r}$.



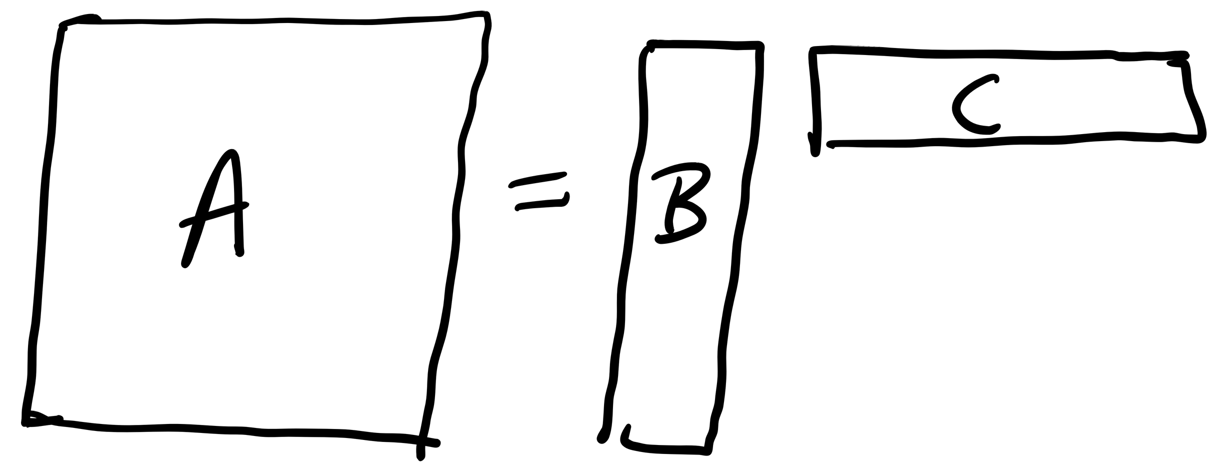
Luego tomando una base de ese espacio, por ejemplo $\{b_1, \dots, b_r\}$ tenemos que podemos escribir cada columna de A como combinación lineal de los vectores $\{b_i\}$, a saber,

$$a_1 = c_{11} b_1 + \dots + c_{r1} b_r$$

$$a_n = c_{1n} b_1 + \dots + c_{rn} b_r$$

$$A = B \cdot C$$

$$\underbrace{\left[a_1 | \dots | a_n \right]}_{m \times n} = \underbrace{\left[b_1 | \dots | b_r \right]}_{m \times r} \underbrace{\boxed{\quad C \quad}}_{r \times n}$$



$r << \{m, n\}$

Lo que acabamos de probar es que toda matriz ~~...~~ $m \times n$ de rango r puede ser ~~escrita~~ representada por un producto de matrices de tamaños $m \times r$ y $r \times n$.

Observar que en el caso $r << \min(m,n)$ resulta que A puede ser descrita por $r(m+n)$ coeficientes (en vez de $m \cdot n$ que es mucho mayor).

~~Aproximar~~ Obs: Lo que acabamos de hacer es una generalización ~~...~~ de lo hecho para matrices de rango 1 (producto tensorial). Básicamente estamos diciendo que matrices de rango k las podemos escribir tipo $(b_1 | \dots | b_r) \begin{pmatrix} z_1^* \\ \vdots \\ z_r^* \end{pmatrix}$.

Aproximar matrices por matrices de rango menor es muy utilizado en diversas aplicaciones (compresión, "de-noising", completación de matrices, procesamiento de imágenes, etc...)

Observar que del SVD podemos escribir $A = \sigma_1 u_1 v_1^* + \cdots + \sigma_r u_r v_r^*$
que es una suma de r matrices de rango 1, y las cuales son ortonormales entre sí con $\langle \cdot, \cdot \rangle_F$.

Definimos $A_k := \sigma_1 u_1 v_1^* + \cdots + \sigma_r u_r v_r^*$ $\quad (k = 1, \dots, r)$ $\quad A_r = A$

$$\boxed{\begin{array}{l} \underline{\text{Teorema (Low-Rank Aprox)}} \quad (k \leq r) \\[1em] \displaystyle \min_{\substack{B \in \mathbb{C}^{m \times n} \\ rk(B) \leq k}} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}. \end{array}}$$

~~Para el caso~~ Un caso particular importante (a veces llamado Eckart-Young) es cuando $m = n$ y A invertible.

Luego $\displaystyle \min_{rk B \leq n-1} \|A - B\| = \|A - A_{n-1}\| = \sigma_n$.

Comenzamos haciendo la prueba de este caso:

Si $B \in \mathbb{C}^{n \times n}$ / $\exists w \in \mathbb{C}^n$ ($\|w\| = 1$), $Bw = 0 \Longrightarrow \|A - B\| \geq \|(A - B)w\| = \|Aw\|$

$\underline{\text{Afirmación}}$: $\|Aw\| \geq \sigma_n$ $\quad (\|w\| = 1)$

Esto resulta del hecho ~~que~~ ~~$\|A^{-1}\| = 1/\sigma_n$. Y las ortonormales~~

$$\|Aw\|^2 = \|U \Sigma V^* w\|^2 = \|\Sigma V^* w\|^2 = \|\Sigma y\|^2 = \left\| \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\|^2 = |\sigma_1 y_1|^2 + \cdots + |\sigma_n y_n|^2$$

$y = V^* w$

$$= \sigma_1^2 |y_1|^2 + \cdots + \sigma_n^2 |y_n|^2 \geq \sigma_n^2 \cdot (|y_1|^2 + \cdots + |y_n|^2) = \sigma_n^2 \|y\|^2 = \sigma_n^2$$

$\sigma_1 \geq \cdots \geq \sigma_n$ $\quad\quad \|y\|^2 = \|V^* w\|^2 = \|w\|^2 = 1$

(Alternativo: $1 = \|w\| = \|A^{-1} A w\| \leq \|A^{-1}\| \cdot \|Aw\| = \dfrac{\|Aw\|}{\sigma_n}$

$\Longrightarrow \|Aw\| \geq \sigma_n$.

Luego tenemos que $\|A-B\| \geq \sigma_n \quad \forall B$ de rango $\leq n-1$.

Basta ver que esta desigualdad se alcanza en $B = A_{n-1} = \sigma_1 u_1 v_1^* + \cdots + \sigma_{n-1} u_{n-1} v_{n-1}^*$

$$\|A - A_{n-1}\| = \|\sigma_n u_n v_n^*\| = |\sigma_n| = \sigma_n. \qquad \square$$

Observar que la misma prueba vale para el caso de norma de Frobenius.

<u>Dem Teo:</u> Sea $B \in \mathbb{C}^{m\times n}$ de rango k ($k \leq \text{rango}(A) = r$)

Luego $\dim \ker B \geq n-k$. Por otro lado sea $W = \langle v_1, \ldots, v_{k+1} \rangle$ es decir el espacio generado por los primeros $k+1$ vectores singulares.

Por estar W y $\ker B$ en $\mathbb{C}^n$ resulta $\exists\, w \in \underset{\dim \geq n-k}{\ker B} \cap \underset{\dim = k+1}{W}$ no trivial.

Es decir $\exists\, w = \alpha_1 v_1 + \cdots + \alpha_{k+1} v_{k+1}$, con $|\alpha_1|^2 + \cdots + |\alpha_{k+1}|^2 = 1$ y que satisface $Bw = 0$.

$$\text{Luego } \|A-B\|_2 \geq \|(A-B)w\| \overset{Bw=0}{=} \|Aw\| = \|\alpha_1 \sigma_1 u_1 + \cdots + \alpha_{k+1} \sigma_{k+1} u_{k+1}\|$$

$$\underset{u_i \perp u_j}{\overset{\text{pitagoras}}{=}} \left(|\alpha_1|^2 \sigma_1^2 + \cdots + |\alpha_{k+1}|^2 \sigma_{k+1}^2\right)^{1/2} \geq \left(|\alpha_1|^2 + \cdots + |\alpha_{k+1}|^2\right)^{1/2} \sigma_{k+1} = \sigma_{k+1}.$$

Veamos que el mínimo se alcanza en $B = A_k \in \{rk = k\}$

$$\|A - A_k\| = \|\underbrace{\sigma_{k+1} u_{k+1} v_{k+1}^* + \cdots + \sigma_n u_n v_n^*}_{=\,(u_1 \cdots u_{k+1} \cdots u_n) \begin{bmatrix} 0 & & & & \\ & \ddots & & & \\ & & \sigma_{k+1} & & \\ & & & \ddots & \\ & & & & \sigma_n \\ & & 0 & & \end{bmatrix} V^*}\| = \sigma_{k+1}$$

$$\tilde{\Sigma}$$

multiplicando por U^* a izq. y V^* a derecha queda una matriz tipo $\begin{pmatrix} 0 & & & & \\ & \ddots & & & \\ & & \sigma_{k+1} & & \\ & & & \ddots & \\ & & & & \sigma_n \\ & & 0 & & \end{pmatrix}$,

y para ese caso es fácil ver $\|\tilde{\Sigma}\| = \sigma_{k+1}$.

Un resultado similar ocurre si cambiamos $\|\cdot\|_2$ por $\|\cdot\|_F$.

Teo: A como antes.

$$\Rightarrow \min_{\substack{B\in\mathbb{C}^{m\times n} \\ rk(B)\leq k}} \|A-B\|_F = \|A-A_k\|_F = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2}.$$

Omitimos la prueba. De cualquier manera la segunda desigualdad es muy fácil.

$$\|A-A_k\|_F^2 = \|\sigma_{k+1}u_{k+1}v_{k+1}^* + \cdots + \sigma_r u_r v_r^*\|_F^2$$

$$= \sigma_{k+1}^2 + \cdots + \sigma_r^2 \quad \text{dado que } \langle u_i v_i^*, u_j v_j^* \rangle = \delta_{ij}.$$

$\square$