Repaso de SVD

Dada $A \in \mathbb{K}^{m\times n}$, existen matrices:

- $U \in \mathbb{K}^{m\times m}$ unitaria $(U^* = U^{-1})$
- $\Sigma \in \mathbb{K}^{m\times n}$ diagonal
- $V \in \mathbb{K}^{n\times n}$ unitaria $(V^* = V^{-1})$

tales que $A = U \Sigma V^*$ y

$$\Sigma = \left(\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & 0 \\ & & & \sigma_m & \end{array}\right) \quad \text{si } m \leq n$$

ó

$$\Sigma = \left(\begin{array}{ccc} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \hline & 0 & \end{array}\right) \quad \text{si } m > n$$

$\sigma_1 \geq \sigma_2 \geq \dots \geq 0$ son los valores singulares.

$$A = U\Sigma V^* \iff \overbrace{A\cdot V}^{n\times m} = \overbrace{U\cdot\Sigma}^{n\times m}$$

si $V = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_m \\ | & | & & | \end{pmatrix}$, $U = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{pmatrix}$

$$\Rightarrow A\cdot V = \begin{pmatrix} | & | & & | \\ Av_1 & Av_2 & & Av_m \\ | & | & & | \end{pmatrix}$$

si $m \leq n$

$$U\cdot\Sigma = \left(\begin{array}{cccc} | & | & & | \\ \sigma_1 u_1 & \sigma_2 u_2 & & \sigma_m u_m \\ | & | & & | \end{array}\ \overbrace{\begin{array}{ccc} 0 & \cdots & 0 \\ & & \\ 0 & \cdots & 0 \end{array}}^{n-m \text{ ceros}}\right)$$

y

$m > n$

$$U\cdot\Sigma = \begin{pmatrix} | & & | \\ \sigma_1 u_1 & & \sigma_n u_n \\ | & & | \end{pmatrix}$$

En ambos casos $AV = U\cdot\Sigma$ dice que

$Av_i = \sigma_i u_i \quad i = 1\ldots,\min\{n,m\}$

si $m \leq n$ $\quad Av_i = \sigma_i\cdot u_i \quad i = 1\ldots m$

$\qquad\qquad Av_i = 0 \quad i = m+1,\ldots n$

Si $m > n$ $\quad Av_i = \sigma_i u_i \quad i = 1 \dots n$

Av_i libre $\quad i = n+1, \dots, m.$

OBS: En ambos casos para los $\sigma_i = 0$ se tendrá $Av_i = 0$.
y en este caso vemos que $\sigma_i \cdot u_i = 0$ también $\Rightarrow$ esos u_i son libres.

¿Como construimos U, Σ, V?

1) Planteamos $A^* A \in \mathbb{K}^{n \times n}$ que es hermitiana
$\Rightarrow$ existe una BON de autovectores
$v_1 \dots v_n$ y autovalores no negativos
$\lambda_1 \geq \dots \geq \lambda_n \geq 0$ (los ordenamos)

$$\Rightarrow V = \begin{pmatrix} \vdots & & \vdots \\ v_1 & \cdots & v_n \\ \vdots & & \vdots \end{pmatrix}, \quad \sigma_i = \sqrt{\lambda_i}$$

$$\Sigma = \left(\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_n & \end{array}\right) \text{ ó } \Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \hline & 0 & \end{pmatrix}$$

$$\sigma_i = \sqrt{\lambda_i}$$

2) Para $\sigma_i \neq 0$ $\quad u_i = \frac{1}{\sigma_i} A v_i$
$i = 1, \dots, s$

Para $\sigma_i = 0$ tenemos $A v_i = 0$

Para los $u_{s+1}, \dots, u_m$ que faltan completamos a una BON

Valores singulares y norma 2

Vimos que:

Teo: $A \in \mathbb{K}^{m \times n} \Rightarrow \|A\|_2 = \sigma_1$

Además, si $A \in \mathbb{K}^{n \times n}$ inversible

$\Rightarrow \|A^{-1}\|_2 = \frac{1}{\sigma_n}$ ya que

val sing de A^{-1} son $1/\sigma_i$, σ_i val sing de A.

Ej: Probar $A \in \mathbb{K}^{n\times n}$ es inversible si y solo si $\sigma_n > 0$.

Luego $\text{Cond}_2(A) = \sigma_1/\sigma_n$.

Recordar que para matrices simétricas $\text{Cond}_2(A) = \dfrac{|\lambda_{max}|}{|\lambda_{min}|}$

Si bien vimos que

$$\text{Cond}_2(A) \geq \frac{\|A\|}{\|A-B\|}$$

para B singular y que entonces podría pensarse mal condicionamiento de A con "cuan cercana está A de una matriz singular" vemos que

$\varepsilon \cdot Id$ está cerca de ser singular pero está super bien condic. $\quad Cond_2(\varepsilon Id) = 1$

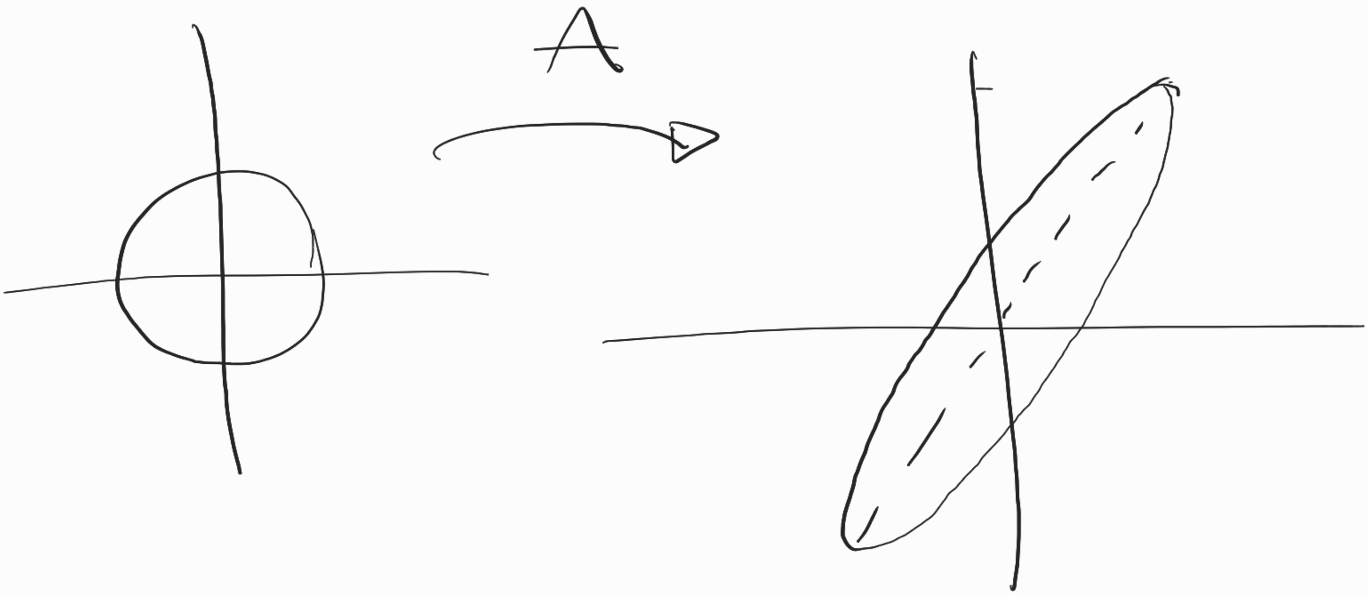
Para A invertible, podemos pensar que la condición sí tiene que ser con si hay mucha distancia entre los valores singulares de la matriz, es decir si

$$\sigma_1 >> \sigma_n$$

ó $\dfrac{\sigma_1}{\sigma_n}$ será grande.

Geométricamente, estamos mirando si los círculos tienden a estirar-

se mucho en un eje y compri-
mirse mucho en otro.



Para matrices no cuadradas,
qué podemos decir?

Vimos la forma reducida:

Si $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = 0$

$\Rightarrow A = U_r \Sigma_r V_r^* \in \mathbb{K}^{m \times n}$

donde $V_r = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_r \\ | & & | \end{pmatrix} \in \mathbb{K}^{r \times n}$

$$\Sigma_r = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \in \mathbb{K}^{r \times r}$$

$$U_r = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{pmatrix} \in \mathbb{K}^{m \times r}$$

Por otro lado, si $k \leq r$

$$V_k = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_k \\ | & & | \end{pmatrix} \quad \Sigma_k = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix}$$

$$U_k = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_k \\ | & & | \end{pmatrix}$$

$$A_k = V_k \Sigma_k U_k^*$$

$$= V_k \left(D_1 + \cdots + D_k \right) U_k^*$$

(con $D_1 = \begin{pmatrix} \sigma_1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}$, $D_2 = \begin{pmatrix} 0 & & & & \\ & \sigma_2 & & & \\ & & 0 & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix}$

$D_i = \begin{pmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & \sigma_i & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \to$ lugar i

$$= V_k D_1 U_k^* + \cdots + V_k D_k U_k^*$$

$$V_k D_i U_k^* =$$

$$\begin{pmatrix} \vdots & & \vdots \\ v_1 & \cdots & v_k \\ \vdots & & \vdots \end{pmatrix} \begin{pmatrix} 0 & & & & \\ & \ddots 0 & & & \\ & & \sigma_i & & \\ & & & 0 \ddots & \\ & & & & 0 \end{pmatrix} \begin{pmatrix} \cdots & \bar{u}_1 & \cdots \\ & \vdots & \\ \cdots & \bar{u}_k & \cdots \end{pmatrix}$$

$$= \begin{pmatrix} \vdots & & \vdots \\ v_1 & & v_k \\ \vdots & & \vdots \end{pmatrix} \begin{pmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ \cdots & \sigma_i \bar{u}_i & \cdots \\ 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \vdots & & \vdots \\ \sigma_i v_i \bar{u}_i^1 & \cdots & \sigma_i v_i \bar{u}_i^k \\ \vdots & & \vdots \end{pmatrix}$$

$$= \sigma_i \boxed{v_i} \cdot \boxed{u_i^*}$$

$\Rightarrow A_k = \sigma_1 v_1 u_1^* + \dots + \sigma_k v_k u_k^*$

que tiene rango k.

<u>Teorema:</u>

Vale que si $A \in \mathbb{K}^{m\times n}$ y $rg(A) = r \Rightarrow \forall\, k \leq r$ se tiene que A_k satisface

$$\|A - A_k\|_2 = \min_{B \in \mathbb{K}^{m\times n} :\, rg(B) = k} \|A - B\|_2 = \sigma_{k+1}$$

Es decir A_k es la "mejor" aproximación en norma 2 de matrices de la matriz A a una matriz de rango k.

Aplicaciones

$$\underbrace{A}_{m\times n} = \underbrace{U}_{m\times m}\ \underbrace{\begin{pmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \hline & 0 & \end{pmatrix}}_{m\times n}\ \underbrace{V^*}_{n\times n}$$

Si $\operatorname{rg}(A) = r < n \Rightarrow \sigma_{r+1} = \cdots = \sigma_n = 0.$

$$\Rightarrow A = A_r = \underbrace{\begin{pmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{pmatrix}}_{m\times r} \underbrace{\begin{pmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{pmatrix}}_{r\times r} \underbrace{\begin{pmatrix} - & v_1 & - \\ & \vdots & \\ - & v_r & - \end{pmatrix}}_{r\times n}$$

OBS:

$$\underbrace{\begin{pmatrix} - & v_1 & - \\ - & v_r & - \end{pmatrix}}_{r\times n} \underbrace{\begin{pmatrix} | & & | \\ v_1 & \cdots & v_r \\ | & & | \end{pmatrix}}_{n\times m} = Id_r$$

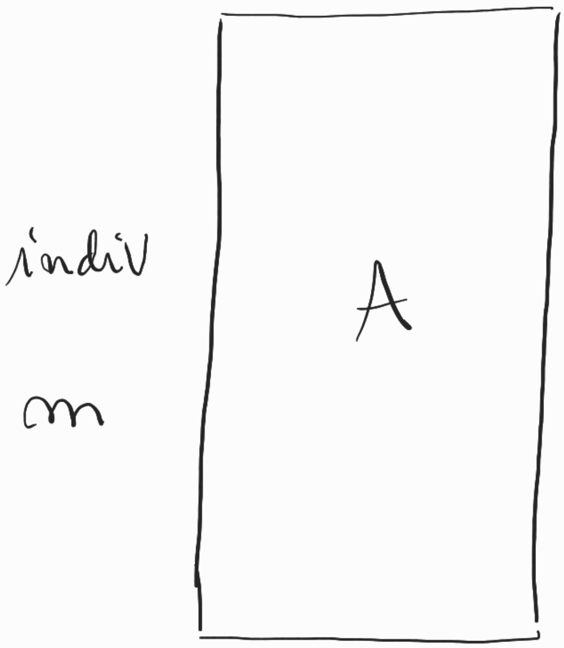
$$\Rightarrow A \cdot \begin{pmatrix} | & & | \\ v_1 & \cdots & v_r \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ u_1 & & u_r \\ | & & | \end{pmatrix} \begin{pmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{pmatrix}$$

El resto $v_{r+1} \cdots v_n$ están en $Nu(A)$.

Aplicaciones:

- Reducción de dimensionalidad.

n características (altura, peso, etc)



Puede haber características redundantes

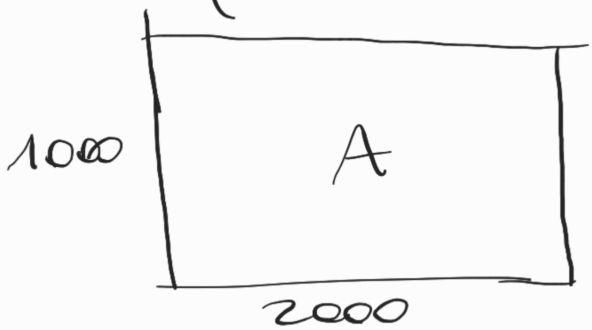
$rg(A) = r$

si $k < r$

σ_{k+1} nos dice cuánta información perdemos si usamos A_k en lugar de $A = A_r$.

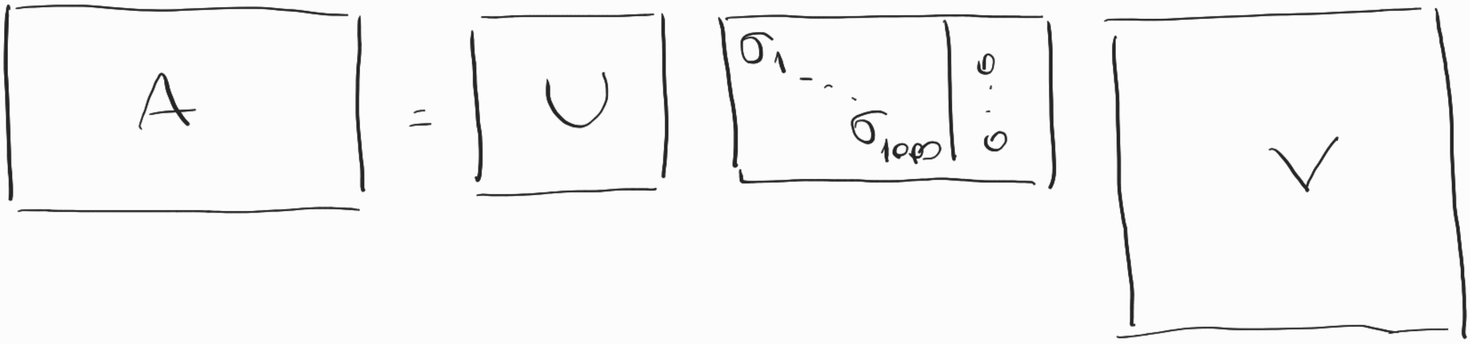
- Imágenes (en color, grises)

Grises

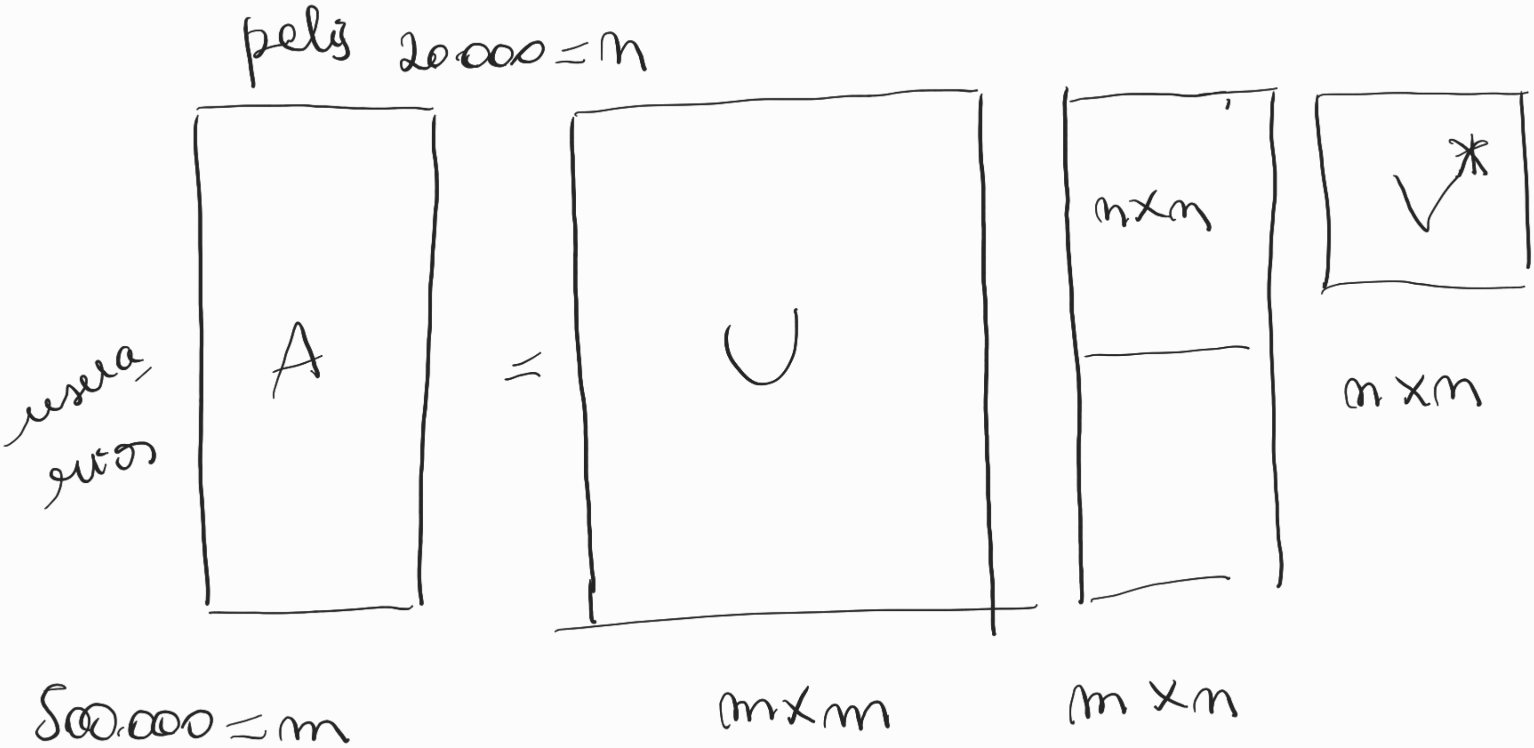


pixeles
1000×2000

¿Cuánto puedo comprimir?



- Algoritmos de recomendación



Uso reducida: $r = rg(A)$

$$A = A_r = \underbrace{\begin{pmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{pmatrix}}_{m \times r} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \underbrace{\begin{pmatrix} - & v_1 & - \\ & \vdots & \\ - & v_r & - \end{pmatrix}}_{r \times n}$$

$$
A = \underbrace{\overbrace{\begin{bmatrix} \\ U_r \\ \\ \end{bmatrix}}^{\text{"características"}}}_{\text{caract "más relevante"}} \;\; \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}}_{\text{importancia de c/ característica}} \;\; \overbrace{\begin{bmatrix} V_r^* \end{bmatrix}}^{\text{pelis}} \Big\} \text{"características"}
$$

usuarios

Las "características" son desconocidas.

Cada

fila i de $\sqrt{2}\, U_r = Q_r$

col j de $\sqrt{2}\, V_r^* = P_r$

$$
a_{ij} = \underbrace{q_i^t}_{\substack{\text{concentra} \\ \text{características} \\ \text{de usuario } i}} \cdot \underbrace{p_j}_{\substack{\text{concentra relación} \\ \text{entre caract y} \\ \text{peli } j.}}
$$

Se busca

$$\min_{p,q} \sum_{(i,j)\in\Omega} (r_{ij} - \overbrace{q_i^t p_j}^{a_{ij} \text{ de la q' queremos}})^2 + \lambda \|Q_r\|_F^2 + \lambda \|P_r\|_F^2$$

$\Omega = \{(i,j) : r_{ij}$ es conocido$\}$

Conjunto de entrenamiento.

De Hastie (Tao 2009, Mazumder 2010)

$\Omega = \{(i,j) : a_{ij}$ es conocido$\}$

$P_\Omega(M) =$ dejar los m_{ij} con $(i,j)\in\Omega$ y cero en el resto

$P_{\Omega^\perp}(M) =$ dejar los m_{ij} con $(i,j)\notin\Omega$ y cero en el resto.

A es la que tiene entradas originales

Sea M : minimiza

$$\frac{1}{2} \| P_\Omega (A - M) \|_F + \lambda \underbrace{\| M \|_*}_{\substack{\sum \text{ valores} \\ \text{sing} \\ \text{de } M}}$$

$(\sum \text{val sing}^2)^{1/2}$

Lo resuelven iterativamente:

1. M_0 cualquiera

$$A_1 = P_\Omega (A) + P_{\Omega^\perp}(M_0)$$

2. Si $A_1 = U \Sigma V^*$

$$\Rightarrow \quad M_1 = U \, S_\lambda(\Sigma) \, V^*$$

$S_\lambda(\Sigma) = \operatorname{diag}(d_{ii} - \lambda)_+$.

$\Rightarrow$ M_1 matriz de bajo rango si λ es grande.

$$\| A_1 - M_1 \|_F = \| U (\Sigma - S_\lambda(\Sigma)) V^* \|_F$$

$$= \| \Sigma - S_\lambda(\Sigma) \|_F$$

Si λ es chico $\quad \Sigma \sim S_\lambda(\Sigma)$

Si λ es grande $\quad \Sigma - S_\lambda(\Sigma) \sim \Sigma$