# La descomposición de valor singular

María Eugenia Torio
Dto de Física-ECEN-FCEIA

del libro: "Álgebra Lineal: una introducción moderna"
David Poole

# La descomposición de valor singular

toda matriz simétrica $A$ puede factorizarse como $A = PDP^T$,

$P$ es una matriz ortogonal y $D$ es una matriz diagonal

$A$ no es simétrica, diagonalizable, $A = PDP^{-1}$,

$P$ ahora matriz invertible $D$ es una matriz diagonal

no toda matriz es diagonalizable !!!!!!

Lo nuevo:

*toda* matriz (simétrica o no, cuadrada o no) tiene una factorización

$A = PDQ^T$, ¡donde $P$ y $Q$ son ortogonales y $D$ es una matriz diagonal!

*descomposición de valor singular* (DVS)

## Los valores singulares de una matriz

Para cualquier matriz $A$ de $m \times n$, la matriz $A^TA$ de $n \times n$ es simétrica y en consecuencia puede ser diagonalizable ortogonalmente , por el teorema espectral. No sólo todos los eigenvalores de $A^TA$ son reales (Teorema 5.18), todos son *no negativos*. Para demostrar esto, sea $\lambda$ un eigenvalor de $A^TA$ con su correspondiente eigenvector unitario $\mathbf{v}$. Entonces

$$
\begin{aligned}
0 \le \|A\mathbf{v}\|^2 &= (A\mathbf{v}) \cdot (A\mathbf{v}) = (A\mathbf{v})^T A\mathbf{v} = \mathbf{v}^T A^T A\mathbf{v} \\
&= \mathbf{v}^T \lambda \mathbf{v} = \lambda(\mathbf{v} \cdot \mathbf{v}) = \lambda \|\mathbf{v}\|^2 = \lambda
\end{aligned}
$$

Por tanto, tiene sentido sacar raíces cuadradas (positivas) de dichos eigenvalores.

**Definición** Si $A$ es una matriz de $m \times n$, los ***valores singulares*** de $A$ son las raíces cuadradas de los eigenvalores de $A^TA$ y se denotan mediante $\sigma_1, \ldots, \sigma_n$. Es convencional ordenar los valores singulares de modo que $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n$.

## Ejemplo

Encuentre los valores singulares de

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

**Solución** La matriz

$$A^T A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

tiene eigenvalores $\lambda_1 = 3$ y $\lambda_2 = 1$. En consecuencia, los valores singulares de $A$ son $\sqrt{\lambda_1} = \sqrt{3}$ y $\sigma_2 = \sqrt{\lambda_2} = 1$.

Para comprender el significado de los valores singulares de una matriz $A$ de $m \times n$, considere los eigenvectores de $A^TA$. Puesto que $A^TA$ es simétrica, se sabe que existe una base *ortonormal* para $\mathbb{R}^n$ que consiste de los eigenvectores de $A^TA$. Sea $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ una base tal que corresponde a los eigenvalores de $A^TA$, ordenados de modo que $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. A partir de los cálculos justo antes de la definición,

$$\lambda_i = \|A\mathbf{v}_i\|^2$$

Por tanto,

$$\sigma_i = \sqrt{\lambda_i} = \|A\mathbf{v}_i\|$$

En otras palabras, los valores singulares de $A$ son las longitudes de los vectores $A\mathbf{v}_1, \ldots, A\mathbf{v}_n$.

## La descomposición de valor singular

Se quiere demostrar que una matriz $A$ de $m \times n$ puede factorizarse como

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

donde $U$ es una matriz ortogonal de $m \times m$, $V$ es una matriz ortogonal de $n \times n$ y $\Sigma$ es una matriz "diagonal" de $m \times n$. Si los valores singulares *distintos de cero* de $A$ son

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$$

y $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$, entonces $\Sigma$ tendrá la forma de bloque

$$\Sigma = \begin{array}{cc} \begin{array}{cc} \overbrace{\phantom{D}}^{r} & \overbrace{\phantom{O}}^{n-r} \end{array} & \\ \left[\begin{array}{c|c} D & O \\ \hline O & O \end{array}\right] & \begin{array}{l} \}\, r \\ \}\, m-r \end{array} \end{array}, \quad \text{donde} \quad D = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix} \tag{1}$$

y cada matriz $O$ es una matriz cero del tamaño adecuado. (Si $r = m$ o $r = n$, alguna de éstas no aparecerá.) Algunos ejemplos de tal matriz $\Sigma$ con $r = 2$ son

$$\Sigma = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(¿Cuál es $D$ en cada caso?)

## La descomposición de valor singular

Se quiere demostrar que una matriz $A$ de $m \times n$ puede factorizarse como

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

donde $U$ es una matriz ortogonal de $m \times m$, $V$ es una matriz ortogonal de $n \times n$ y $\Sigma$ es una matriz "diagonal" de $m \times n$. Si los valores singulares *distintos de cero* de $A$ son

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$$

y $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$, entonces $\Sigma$ tendrá la forma de bloque

$$\Sigma = \begin{array}{cl} \begin{array}{cc} \overbrace{\quad}^{r} & \overbrace{\quad}^{n-r} \end{array} & \\ \left[\begin{array}{c:c} D & O \\ \hdashline O & O \end{array}\right] & \begin{array}{l} \}\, r \\ \}\, m-r \end{array} \end{array}, \quad \text{donde} \quad D = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix} \tag{1}$$

y cada matriz $O$ es una matriz cero del tamaño adecuado. (Si $r = m$ o $r = n$, alguna de éstas no aparecerá.) Algunos ejemplos de tal matriz $\Sigma$ con $r = 2$ son

$$\Sigma = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(¿Cuál es $D$ en cada caso?)

Para construir la matriz ortogonal $V$, primero encuentre una base ortonormal $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ para $\mathbb{R}^n$ que consiste de eigenvectores de la matriz simétrica $A^TA$ de $n \times n$. Entonces

$$V = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n]$$

es una matriz ortogonal de $n \times n$.

Para la matriz ortogonal $U$, note primero que $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_n\}$ es un conjunto ortogonal de vectores en $\mathbb{R}^m$. Para ver esto, suponga que $\mathbf{v}_i$ es el eigenvector de $A^TA$ correspondiente al eigenvalor $\lambda_i$. Entonces, para $i \neq j$, se tiene

$$\begin{aligned}
(A\mathbf{v}_i) \cdot (A\mathbf{v}_j) &= (A\mathbf{v}_i)^T A\mathbf{v}_j \\
&= \mathbf{v}_i^T A^T A\mathbf{v}_j \\
&= \mathbf{v}_i^T \lambda_j \mathbf{v}_j \\
&= \lambda_j(\mathbf{v}_i \cdot \mathbf{v}_j) = 0
\end{aligned}$$

dado que los eigenvectores $\mathbf{v}_i$ son ortogonales. Ahora recuerde que los valores singulares satisfacen $\sigma_i = \|A\mathbf{v}_i\|$ y que los primeros $r$ de éstos son distintos de cero. Por tanto, puede normalizar $A\mathbf{v}_1, \ldots, A\mathbf{v}_r$ al establecer

$$\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i \quad \text{para } i = 1, \ldots, r$$

Esto garantiza que $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ es un conjunto ortonormal en $\mathbb{R}^m$, pero si $r < m$ no será una base para $\mathbb{R}^m$. En este caso, se extiende el conjunto $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ a una base ortonormal $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ para $\mathbb{R}^m$. (Esta es la única parte complicada de la construcción; en los ejemplos siguientes y en los ejercicios se describirán técnicas para realizarla.) Entonces se establece

$$U = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_m]$$

¿ Todo lo que hicimos, funciona????? ?

es necesario verificar que con $U$, $V$ y $\Sigma$ como se describieron $A = U\Sigma V^T$.

$V^T = V^{-1}$ $\longrightarrow$ equivale a demostrar que $AV = U\Sigma$

Se sabe que $$A\mathbf{v}_i = \sigma_i \mathbf{u}_i \quad \text{para } i = 1, \ldots, r$$

y $\|A\mathbf{v}_i\| = \sigma_i = 0$ para $i = r + 1, \ldots, n$. En consecuencia,

$$A\mathbf{v}_i = \mathbf{0} \quad \text{para } i = r + 1, \ldots, n$$

Por tanto, $$AV = A[\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n] = [A\mathbf{v}_1 \quad \cdots \quad A\mathbf{v}_n]$$

$$= [\sigma_1\mathbf{u}_1 \quad \cdots \quad \sigma_r\mathbf{u}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0}]$$

$$= [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_m]\left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O \\ 0 & \cdots & \sigma_r & \\ \hline & O & & O \end{array}\right] = U\Sigma$$

Por tanto, $\boxed{AV} = A[\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n] = [A\mathbf{v}_1 \quad \cdots \quad A\mathbf{v}_n]$

$$= [\sigma_1 \mathbf{u}_1 \quad \cdots \quad \sigma_r \mathbf{u}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0}]$$

$$= [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_m] \left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O \\ 0 & \cdots & \sigma_r & \\ \hline & O & & O \end{array}\right] \boxed{= U\Sigma}$$

como se requería.

Acaba de demostrarse el siguiente teorema extremadamente importante.

Por tanto, $$\boxed{AV} = A[\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n] = [A\mathbf{v}_1 \quad \cdots \quad A\mathbf{v}_n]$$

$$= [\sigma_1\mathbf{u}_1 \quad \cdots \quad \sigma_r\mathbf{u}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0}]$$

$$= [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_m]\left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O \\ 0 & \cdots & \sigma_r & \\ \hline & O & & O \end{array}\right] \boxed{= U\Sigma}$$

como se requería.

Acaba de demostrarse el siguiente teorema extremadamente importante.

## La descomposición de valor singular

Sea $A$ una matriz de $m \times n$ con valores singulares $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ y $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$. Entonces existe una matriz ortogonal $U$ de $m \times m$, una matriz ortogonal $V$ de $n \times n$ y una matriz $\Sigma$ de $m \times n$ de la forma que se muestra en la ecuación (1) tal que

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

Por tanto,
$$\boxed{AV} = A[\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n] = [A\mathbf{v}_1 \quad \cdots \quad A\mathbf{v}_n]$$
$$= [\sigma_1\mathbf{u}_1 \quad \cdots \quad \sigma_r\mathbf{u}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0}]$$
$$= [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_m]\left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O \\ 0 & \cdots & \sigma_r & \\ \hline & O & & O \end{array}\right] \boxed{= U\Sigma}$$
como se requería.

Acaba de demostrarse el siguiente teorema extremadamente importante.

**La descomposición de valor singular**

Sea $A$ una matriz de $m \times n$ con valores singulares $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ y $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$. Entonces existe una matriz ortogonal $U$ de $m \times m$, una matriz ortogonal $V$ de $n \times n$ y una matriz $\Sigma$ de $m \times n$ de la forma que se muestra en la ecuación (1) tal que
$$A = U\Sigma V^T$$

Una factorización de $A$ como en el Teorema se llama ***descomposición de valor singular***

## La descomposición de valor singular

Sea $A$ una matriz de $m \times n$ con valores singulares $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ y $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$. Entonces existe una matriz ortogonal $U$ de $m \times m$, una matriz ortogonal $V$ de $n \times n$ y una matriz $\Sigma$ de $m \times n$ de la forma que se muestra en la ecuación (1) tal que

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

Una factorización de $A$ como en el Teorema

se llama ***descomposición de valor singular (DVS)*** de $A$.

Las columnas de $U$ se llaman ***vectores singulares izquierdos*** de $A$

las columnas de $V$ se llaman ***vectores singulares derechos*** de $A$.

# Compresión de imágenes digitales

En el caso de las imágenes digitales, suponga que tiene una imagen en escala de grises con un tamaño de $340 \times 280$ píxeles. Cada píxel es una de los 256 tonos de gris, que pueden representarse mediante un número entre 0 y 255. Esta información se puede almacenar en una matriz $A$ $340 \times 280$, pero transmitir y manipular estos 95,200 números es muy costoso. La idea detrás de la compresión de imágenes es que algunas partes de la imagen son menos interesantes que otras. Por ejemplo, en una fotografía de alguien de pie en exteriores, puede haber mucho cielo en el fondo, mientras que el rostro de la persona contiene muchos detalles. Probablemente podría transmitir cada segundo o tercer píxel en el fondo, pero le gustaría mantener todos los píxeles en la región del rostro.

Es evidente que los pequeños valores singulares en la DVS de la matriz $A$ provienen de las partes "aburridas" de la imagen, y puede ignorar muchas de ellas. Suponga, entonces, que tiene la DVS de $A$ en forma de producto externo

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

Sea $k \leq r$ y defina

$$A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$$

Entonces $A_k$ es una aproximación de $A$ que corresponde a mantener sólo los primeros $k$ valores singulares y los correspondientes vectores singulares. Para el ejemplo de $340 \times 280$, puede descubrir que es suficiente con transmitir sólo los datos correspondientes a los primeros 20 valores singulares. Entonces, en lugar de transmitir 95,200 números, sólo es necesario enviar 20 valores singulares más los 20 vectores $\mathbf{u}_1, \ldots, \mathbf{u}_{20}$ en $\mathbb{R}^{340}$ y los 20 vectores $\mathbf{v}_1, \ldots, \mathbf{v}_{20}$ en $\mathbb{R}^{280}$, para un total de

$$20 + 20 \cdot 340 + 20 \cdot 280 = 12{,}420$$

números. ¡Esto representa un ahorro sustancial!

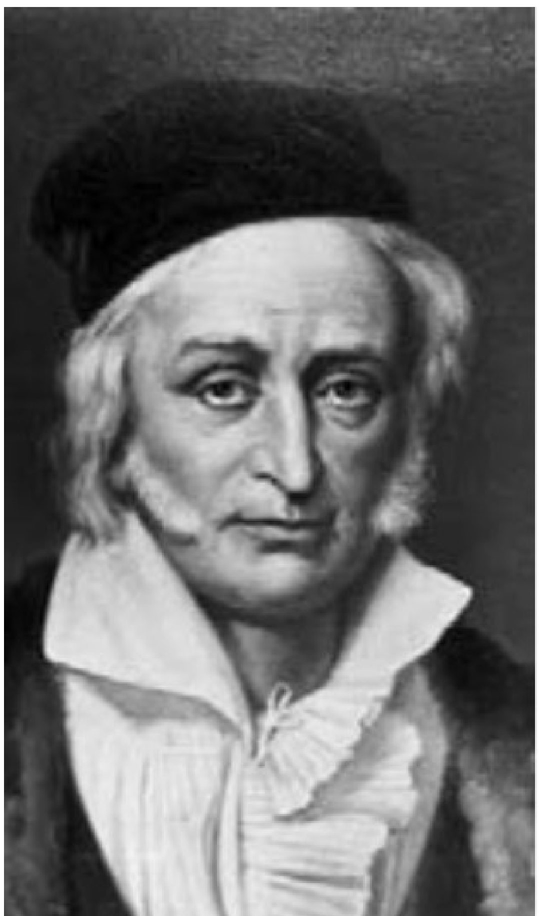


imagen del matemático Gauss

es una imagen de $340 \times 280$ píxeles.

256 tonos de gris

$A$ tiene $340 \times 280$, con entradas entre 0 y 255.

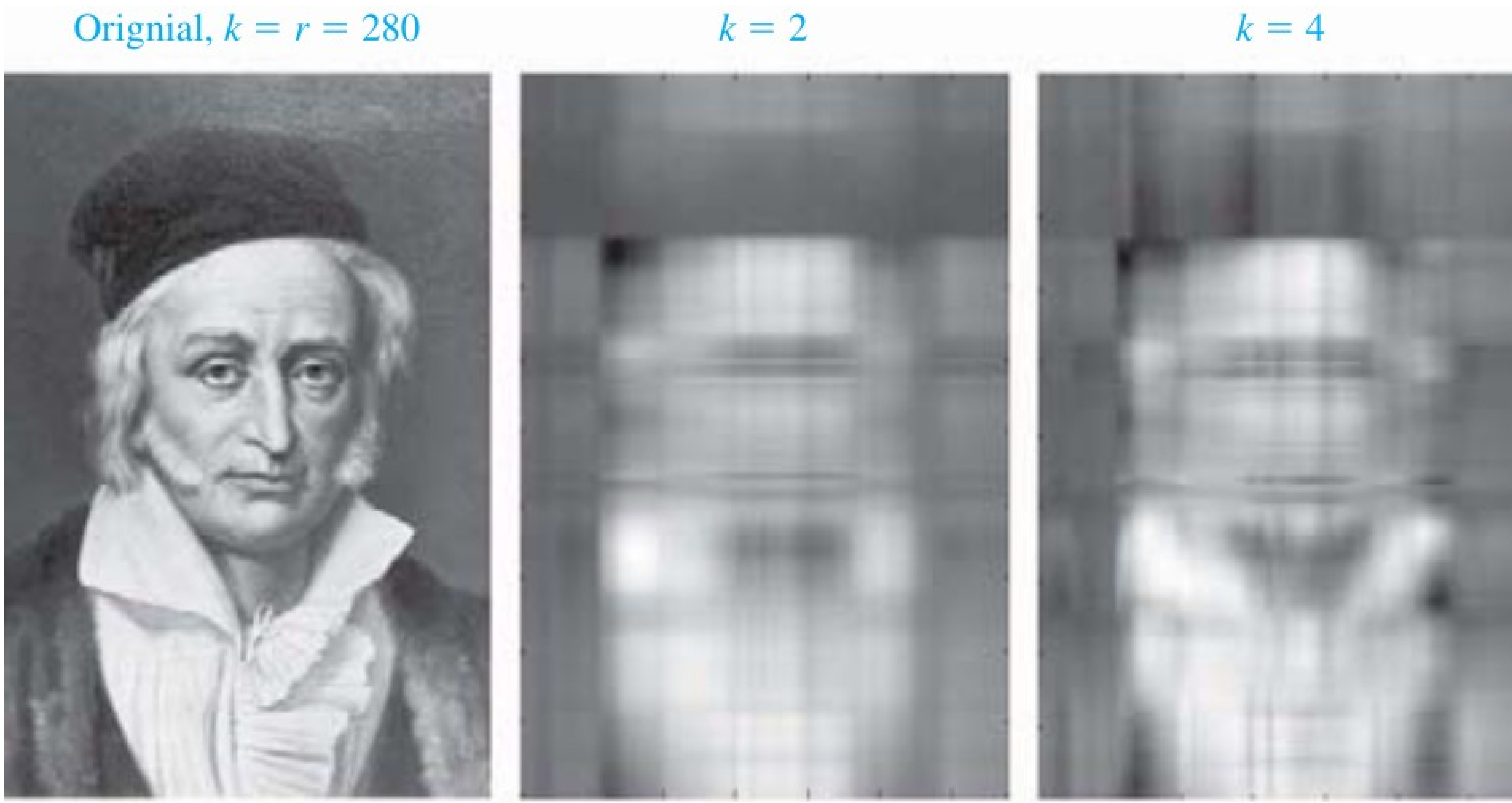
$A$ tiene rank 280.

Si aproxima $A$ mediante $A_k$: $\quad A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$

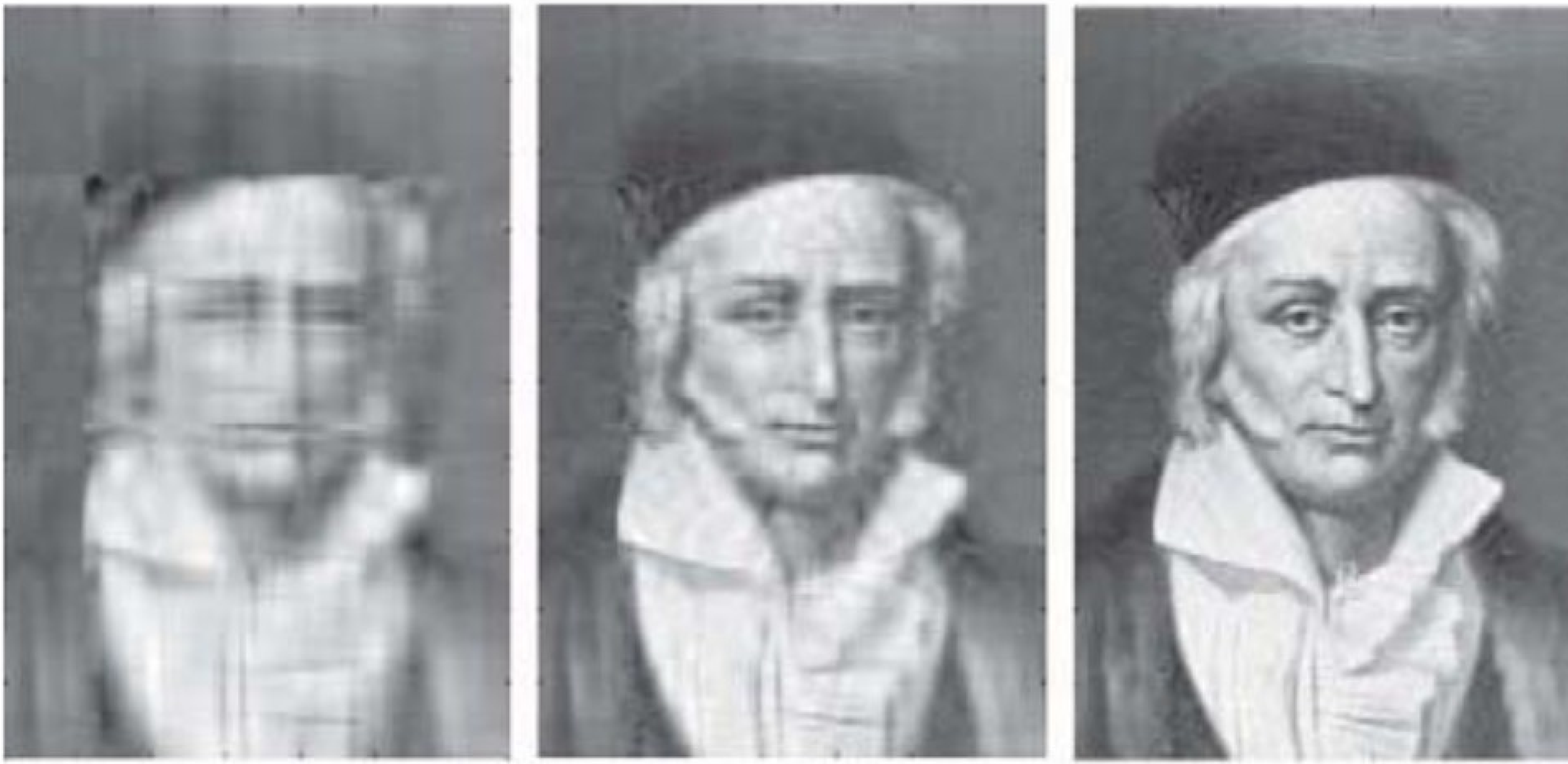
se obtiene una imagen que corresponde a los primeros $k$ valores singulares

Algunos de los valores singulares de $A$ son $\sigma_1 = 49{,}096$, $\sigma_{16} = 22{,}589$, $\sigma_{32} = 10{,}187$, $\sigma_{64} = 484$, $\sigma_{128} = 182$, $\sigma_{256} = 5$ y $\sigma_{280} = 0.5$. Los valores singulares más pequeños contribuyen muy poco a la imagen, razón por la cual las aproximaciones rápidamente parecen tan cercanas al original.

Algunos de los valores singulares de $A$ son $\sigma_1 = 49{,}096$, $\sigma_{16} = 22{,}589$, $\sigma_{32} = 10{,}187$, $\sigma_{64} = 484$, $\sigma_{128} = 182$, $\sigma_{256} = 5$ y $\sigma_{280} = 0.5$. Los valores singulares más pequeños contribuyen muy poco a la imagen, razón por la cual las aproximaciones rápidamente parecen tan cercanas al original.



$k = 8$ $k = 16$ $k = 32$



$k = 64$ $k = 128$ $k = 256$