Tema 14

Pseudo-inversa y descomposición en valores singulares

14.1. Pseudo-inversa

El concepto de inversa de una matriz cuadrada cuyo rango coincide con el número de filas (o columnas) es bien conocido. A^{-1} es la inversa de A si y sólo si $A\,A^{-1} = A^{-1}\,A = I$. Podríamos decir que A^{-1} es la inversa "bilateral" de A. También sabemos que cuando una matriz es invertible, el sistema de ecuaciones $A\,x = b$ tiene solución única dada por $x = A^{-1}\,b$.

En este tema cerramos el curso generalizando el concepto de inversa a otras matrices, lo cual nos permitirá, por ejemplo, proporcionar una alternativa a resolver el sistema $A\,x = b$, cuando A no tiene inversa, tal y como hicimos cuando abordamos el problema de mínimos cuadrados. Para ello, se introduce la matriz conocida como **pseudo-inversa de Moore-Penrose** (o simplemente como *pseudo-inversa*) de la siguiente manera:

Si $A \in \mathbb{R}^{m\times n}$, entonces existe una *<u>única</u>* matriz $A^+ \in \mathbb{R}^{n\times m}$ que satisface las siguientes cuatro condiciones:

1. $A\,A^+\,A = A$.
2. $A^+\,A\,A^+ = A^+$.
3. $A^+\,A = (A^+A)^t$.
4. $A\,A^+ = (A\,A^+)^t$.

Estas condiciones se denominan **condiciones de Penrose**.

Obsérvese que para cualquier matriz A la transformación lineal $T_{A\,A^+}$ asociada a $A\,A^+$ es siempre una proyección, pues

$$(T_{A\,A^+} \circ T_{A\,A^+})(b) = T_{A\,A^+}(A\,A^+\,b) = A\,A^+\,A\,A^+\,b \underset{\substack{\uparrow \\ A\,A^+\,A=A}}{=} A\,A^+\,b = T_{A\,A^+}(b)\,.$$

Si $m = n$ y A es no singular, es evidente que A^{-1} satisface de manera trivial las cuatro condiciones de Penrose, es decir, la pseudo-inversa de A coincide con la inversa de A.

En caso de que $m > n$ y que las $\boxed{\text{columnas de } A \text{ sean linealmente independientes}}$ (lo que llamaremos de rango columna completo), su pseudo-inversa se obtiene con las siguiente expresión:

$$\boxed{A^+ = (A^t\,A)^{-1}\,A^t.}$$

Dicha expresión ha de resultarnos familiar del Tema 13, en el que abordamos el problema de mínimos cuadrados. De hecho, al final de este capítulo estudiaremos la relación entre la pseudo-inversa y la resolución de un sistema por mínimos cuadrados.

Ejemplo

Consideremos la matriz

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}.$$

Es evidente que las columnas de A son linealmente independientes. Por tanto, calculamos su pseudo-inversa mediante

$$\begin{aligned}
A^{+} = (A^{t}A)^{-1}A^{t} &= \left[\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}\right]^{-1}\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \\
&= \begin{pmatrix} 35 & 44 \\ 44 & 56 \end{pmatrix}^{-1}\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \\
&= \frac{1}{24}\begin{pmatrix} 56 & -44 \\ -44 & 35 \end{pmatrix}\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \\
&= \frac{1}{12}\begin{pmatrix} -16 & -4 & 8 \\ 13 & 4 & -5 \end{pmatrix}.
\end{aligned}$$

Podemos verificar que se cumplen las condiciones de Penrose:

1. $A\,A^{+}\,A = A$:

$$\begin{aligned}
A\,A^{+}\,A &= \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}\frac{1}{12}\begin{pmatrix} 16 & -4 & 8 \\ 13 & 4 & -5 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = A\,.
\end{aligned}$$

2. $A^+ A A^+ = A^+$:

$$\begin{aligned} A^+AA^+ &= \frac{1}{12}\begin{pmatrix} -16 & -4 & 8 \\ 13 & 4 & -5 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}\frac{1}{12}\begin{pmatrix} -16 & -4 & 8 \\ 13 & 4 & -5 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\frac{1}{12}\begin{pmatrix} -16 & -4 & 8 \\ 13 & 4 & -5 \end{pmatrix} = A^+ . \end{aligned}$$

3. $A^+ A = (A^+ A)^t$:

$$A^+ A = \frac{1}{12}\begin{pmatrix} -16 & -4 & 8 \\ 13 & 4 & -5 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = (A^+ A)^t .$$

4. $A A^+ = (A A^+)^t$:

$$\begin{aligned} A A^+ &= \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}\frac{1}{12}\begin{pmatrix} -16 & -4 & 8 \\ 13 & 4 & -5 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix} \\ &= (A A^+)^t . \end{aligned}$$

De manera análoga, si $m < n$ y las $\boxed{\text{filas de } A \text{ son linealmente independientes}}$ (lo que llamaremos de rango fila completo), su pseudo-inversa se obtiene con las siguiente expresión:

$$\boxed{A^+ = A^t\,(A\,A^t)^{-1}.}$$

No obstante, en el caso de que A no sea de rango completo no se pueden usar estas fórmulas.

Ejemplo

Sea la matriz:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}.$$

Obviamente $\det(A) = 0$ y por tanto no tiene inversa: las columnas son linealmente dependientes (o no tiene rango completo). Si intentamos usar la expresión $A^+ = (A^t A)^{-1} A^t$ obtenemos que la matriz

$$A^t A = \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix} = \begin{pmatrix} 10 & 20 \\ 20 & 40 \end{pmatrix}$$

también es singular y no tiene inversa.

Para abordar este caso necesitamos introducir el concepto de valores singulares de una matriz.

14.2. Valores singulares

Consideremos una matriz cuadrada $A \in \mathbb{R}^{n \times n}$. Sabemos que, desde un punto de vista geométrico, los vectores propios de A indican las direcciones de “estiramiento puro” de la transformación lineal asociada, mientras que los valores propios corresponden a la magnitud de tal estiramiento. Como vimos anteriormente, una clase particularmente importante de matrices es la de las matrices simétricas reales, con cuyos vectores propios se puede formar una base ortogonal de $\mathbb{R}^n$.

En el caso opuesto, si A es una matriz rectangular, ésta no tiene valores propios. Sin embargo, sí podemos calcular los valores propios de la matriz cuadrada (simétrica) asociada $K = A^t A$; sus correspondientes raíces cuadradas, de gran utilidad en diversos campos como la compresión de imágenes o la recuperación de información, se denominan *valores singulares* de A. Veamos algunas propiedades de la matriz K.

Como ya sabemos, K es simétrica (ver Tema 1) y, en consecuencia, la matriz es diagonalizable y tendrá valores propios reales. Además:

Los valores propios de $A^t A$ son no negativos.

Demostración. En efecto, si $\lambda \in \sigma(A^t A)$ tiene vector propio asociado v_1 (que podemos considerar unitario respecto al producto escalar usual, es decir, $\|v_1\| = 1$), se obtiene que

$$\lambda = \lambda\|v_1\|^2 = \lambda \langle v_1, v_1\rangle = \langle v_1, \lambda v_1\rangle = \langle v_1, A^t A v_1\rangle = \langle A v_1, A v_1\rangle = \|A v_1\|^2 \geqslant 0 \,.$$

□

Así, podemos definir los **valores singulares** de una matriz arbitraria $A \in \mathbb{R}^{m\times n}$ como las n raíces cuadradas (positivas) de los n valores propios no negativos de $A^t A$. Estos valores singulares se denotan de manera habitual por $\sigma_i \left(= \sqrt{\lambda_i}\right)$ $(1 \leqslant i \leqslant n)$ y se ordenan de manera decreciente, teniendo en cuenta sus multiplicidades algebraicas:

$$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n \,.$$

Ejemplo

Consideremos la matriz

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} .$$

Para determinar sus valores singulares, calculamos

$$A^t A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

y sus correspondientes valores propios:

$$\sigma(A^t A) = \{1, 3\}.$$

De este modo, los valores singulares de A son $\sigma_1 = \sqrt{3}$ y $\sigma_2 = 1$.

14.2.1. Descomposición en valores singulares

El cálculo de los valores singulares permite además encontrar una factorización especialmente útil para *cualquier* matriz $A \in \mathbb{R}^{m\times n}$. Esta descomposición, a la que nos referiremos como SVD, es de la forma:

$$\boxed{A = U\,S\,V^t}$$

siendo U ortogonal, S diagonal y V ortogonal.

Descomposición en valores singulares

Sea la matriz $A \in \mathbb{R}^{m\times n}$. Entonces existen matrices ortogonales $U \in \mathbb{R}^{m\times m}$ y $V \in \mathbb{R}^{n\times n}$, de manera que la matriz puede descomponerse como $A = U\,S\,V^t$, siendo S una matriz $m \times n$ de la forma:

$$S = \left(\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 \\ \hline 0 & 0 & \cdots & 0 & \underbrace{0}_{n-r \text{ columnas}} \end{array}\right) \Big\} \; m-r \text{ filas}$$

con $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ y siendo $r \leqslant \min(m, n)$ el número de valores singulares no nulos.

Las $m-r$ filas y $n-r$ columnas de ceros de la matriz S existirán o no en función de los valores de m y n y del número de valores singulares no nulos.
Los valores $\sigma_1, \sigma_2, \dots, \sigma_r$ de la diagonal principal de S son los valores singulares de A.
Las columnas de la matriz U se denominan **vectores singulares por la izquierda** y las columnas de V se denominan **vectores singulares por la derecha** de A.

Obviamente A es equivalente a la matriz S y tienen, por tanto, el mismo rango. Por esta razón

$$r = rg(A)\,.$$

Las matrices U y V no son únicas, pero una forma habitual de obtener tales matrices es la siguiente:

- Matriz V:

$$V = (v_1, \cdots, v_n)\ ,$$

  con $v_1, \dots, v_n$ los vectores propios de $K = A^t\,A$ normalizados y en el orden dado por los valores singulares de A.

- Matriz U:

$$U = (u_1, \cdots, u_m)\ ,$$

  con las columnas $u_1, \cdots, u_r$ calculadas mediante

$$\boxed{u_i = \frac{1}{\sigma_i}\,A\,v_i}$$

  y con $u_{r+1}, \cdots, u_m$, si existen, elegidos de manera que U sea ortogonal (usando Gram-Schmidt, si es necesario).

Ejemplo

Vamos a calcular una descomposición en valores singulares de la matriz 2×3:

$$A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}.$$

Calculamos la matriz $A^t A$:

$$A^t A = \begin{pmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{pmatrix}$$

y sus valores propios:

$$\sigma(A^t A) = \{0, 9, 25\},$$

por lo que los valores singulares de A son $\sigma_1 = 5$, $\sigma_2 = 3$ y $\sigma_3 = 0$. Los vectores propios de $A^t A$ son:

$$v_1' = (1,1,0)^t \ , \quad v_2' = (1,-1,4)^t \ , \quad v_3' = (2,-2,-1)^t \ ,$$

que normalizados resultan:

$$v_1 = \frac{1}{\sqrt{2}}(1,1,0)^t, \quad v_2 = \frac{1}{3\sqrt{2}}(1,-1,4)^t, \quad v_3 = \frac{1}{3}(2,-2,-1)^t.$$

Por tanto, tenemos que

$$\begin{aligned} S &= \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} \\ V &= \frac{1}{3\sqrt{2}} \begin{pmatrix} 3 & 1 & 2\sqrt{2} \\ 3 & -1 & -2\sqrt{2} \\ 0 & 4 & -\sqrt{2} \end{pmatrix}. \end{aligned}$$

Vamos a calcular la matriz U, de dimensión 2×2:

$$
\begin{aligned}
u_1 &= \frac{1}{5} A\, v_1 = \frac{1}{\sqrt{2}} (1,1)^t \,, \\
u_2 &= \frac{1}{3} A\, v_2 = \frac{1}{\sqrt{2}} (1,-1)^t \,.
\end{aligned}
$$

En resumen:

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

Se puede comprobar que efectivamente $A = U\,S\,V^t$.

Las matrices obtenidas en la SVD verifican la siguiente propiedad:

Teorema

Sea $A = U\,S\,V^t$ una descomposición en valores singulares de $A \in \mathbb{R}^{m \times n}$ con $rg(A) = r$. Entonces:

i) $(u_1, \dots, u_r)$ es una base (ortonormal) de $\mathcal{C}(A)$.

ii) $(u_{r+1}, \dots, u_m)$ es una base (ortonormal) de $N(A^t)$.

iii) $(v_1, \dots, v_r)$ es una base (ortonormal) de $\mathcal{C}(A^t)$.

iv) $(v_{r+1}, \dots, v_n)$ es una base (ortonormal) de $N(A)$.

A efectos prácticos, en ocasiones es interesante tener en cuenta el siguiente resultado:

Sea la matriz $A \in \mathbb{R}^{m \times n}$; entonces:

i) El número de valores singulares no nulos de A es $r = rg(A) \leqslant \text{mín}(m, n)$.

ii) Las matrices $A^t\,A$ y $A\,A^t$ tienen los MISMOS valores propios NO NULOS.

En consecuencia, si la matriz A, de dimensión $m \times n$, con $m < n$, tiene descomposición

SVD dada por $A = U\,S\,V^t$, resulta que $A^t = V\,S^t\,U^t$ es la descomposición SVD de A^t. Como $A^t\,A$ es de dimensión $n \times n$ y $A\,A^t$ es de dimensión $m \times m$, los cálculos son más sencillos si trabajamos con esta segunda matriz. Todo lo anterior puede esquematizarse como se indica en la figura 14.1.

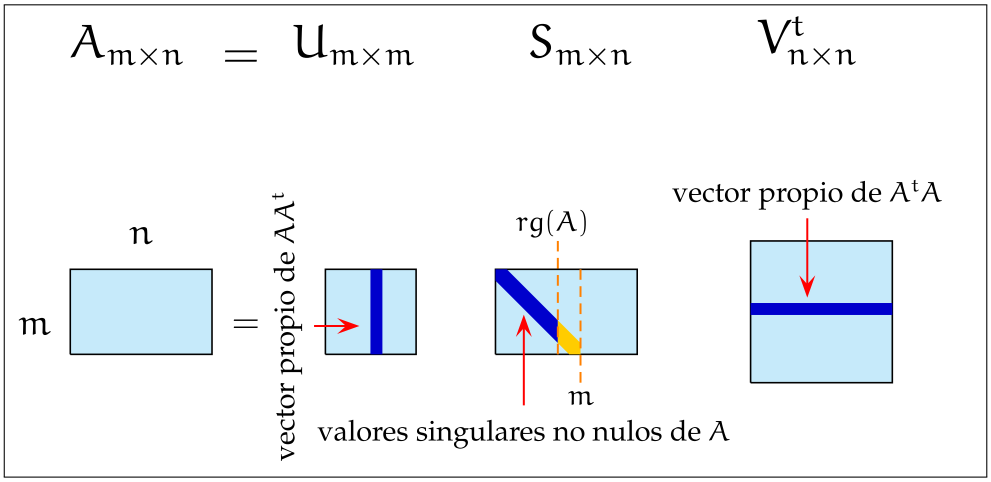

Figura 14.1: Descomposición SVD de una matriz $m \times n$.

Ejemplo

Consideremos de nuevo la matriz

$$A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}.$$

Vamos a obtener la descomposición SVD de A^t. Calculamos

$$A\,A^t = \begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix},$$

cuya ecuación característica es de grado 2; sus valores propios son $\sigma(A\,A^t) = \{9, 25\}$. Por tanto, los valores singulares de A^t son $\sigma_1 = 5, \sigma_2 = 3$ y los vectores propios de $A\,A^t$, normalizados, son

$$v_1 = \frac{1}{\sqrt{2}}(1,1)^t\,, \qquad v_2 = \frac{1}{\sqrt{2}}(1,-1)^t\,.$$

Ahora calculamos los vectores

$$\begin{aligned} u_1 &= \frac{1}{\sigma_1} A^t v_1 = \frac{1}{\sqrt{2}} (1,1,0)^t \ , \\ u_2 &= \frac{1}{\sigma_2} A^t v_2 = \frac{1}{3\sqrt{2}} (1,-1,4)^t \ . \end{aligned}$$

Un vector ortonormal a u_1 y u_2 sería

$$u_3 = \frac{1}{3} (2,-2,-1)^t \ ,$$

que podemos obtener buscando un vector linealmente independiente de u_1 y u_2 y utilizando Gram-Schmidt.

Obsérvese que la descomposición SVD de la matriz A^t obtenida con estos cálculos coincide con la traspuesta de la SVD de A que obtuvimos anteriormente.

14.3. Pseudo-inversa de matrices arbitrarias

Finalmente, damos el siguiente resultado que permite obtener la pseudo-inversa de una matriz arbitraria, incluidas las que no son de rango completo.

Teorema

Sea la descomposición SVD de la matriz $A \in \mathbb{R}^{m\times n}$ de la forma $A = U\,S\,V^t$ con $U \in \mathbb{R}^{m\times m}$ y $V \in \mathbb{R}^{n\times n}$ ortogonales y $\boxed{S \in \mathbb{R}^{m\times n}}$ dada por:

$$S = \left(\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 \\ \hline 0 & 0 & \cdots & 0 & 0 \end{array}\right) \begin{array}{l} \\ \\ \\ \\ \} \ m-r \text{ filas} \end{array}$$

$\underbrace{\qquad}_{n-r \text{ columnas}}$

siendo $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ y siendo $r \leqslant \text{mín}(m, n)$ el número de valores singulares no nulos (e igual al rango de A). La pseudo-inversa de A se calcula mediante la expresión:

$$A^+ = V\,S^+\,U^t$$

donde la matriz $\boxed{S^+ \in \mathbb{R}^{n\times m}}$ se calcula con

$$S^+ = \left(\begin{array}{cccc|c} 1/\sigma_1 & 0 & \cdots & 0 & 0 \\ 0 & 1/\sigma_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_r & 0 \\ \hline 0 & 0 & \cdots & 0 & 0 \end{array}\right) \begin{array}{l} \\ \\ \\ \\ \}\ n-r \text{ filas} \end{array}.$$

$$\underbrace{\qquad}_{m-r \text{ columnas}}$$

Ejemplo

Dada la matriz

$$A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix},$$

cuya descomposición SVD es

$$A = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} \frac{1}{3\sqrt{2}} \begin{pmatrix} 3 & 1 & 2\sqrt{2} \\ 3 & -1 & -2\sqrt{2} \\ 0 & 4 & -\sqrt{2} \end{pmatrix}^t,$$

la pseudo-inversa de A se escribe como

$$A^{+} = V\,S^{+}\,U^{t} \;=\; \frac{1}{3\sqrt{2}}\begin{pmatrix} 3 & 1 & 2\sqrt{2} \\ 3 & -1 & -2\sqrt{2} \\ 0 & 4 & -\sqrt{2} \end{pmatrix}\begin{pmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{3} \\ 0 & 0 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$= \frac{1}{45}\begin{pmatrix} 7 & 2 \\ 2 & 7 \\ 10 & -10 \end{pmatrix}.$$

Se puede comprobar que A^{+} verifica las cuatro condiciones de Penrose.

14.4. Pseudo-inversa y mínimos cuadrados

Recordemos que el sistema de ecuaciones lineales $A\,x = b$ tiene al menos una solución cuando b pertenece al espacio columna de A, pero si $b \notin \mathcal{C}(A)$, entonces el sistema no tiene solución. Sin embargo, si fuera deseable encontrar algún x_0 que sea lo más cercano posible a una solución, calculamos la solución de mínimos cuadrados dada por

$$A^{t}\,A\,x_0 = A^{t}\,b\,.$$

En el caso de que la matriz $A^{t}\,A$ fuese invertible (o en otras palabras, si las columnas de A fueran linealmente independientes), el problema de mínimos cuadrados tiene solución única y ésta viene expresada por

$$x_0 = (A^{t}\,A)^{-1}\,A^{t}\,b\,,$$

que en este caso podemos escribir mediante

$$\boxed{x_0 = A^{+}\,b.}$$

Por otro lado, si la matriz $A^t A$ no fuese invertible, el problema de mínimos cuadrados tiene infinitas soluciones. En esta situación es posible demostrar que, si tenemos la SVD $A = U S V^t$, entonces

$$\boxed{x_0 = A^+ b = V S^+ U^t b}$$

proporciona *una* solución de mínimos cuadrados de todas las posibles, con la propiedad de tener norma euclídea (es decir, asociada al producto escalar usual) mínima.

Ejemplo

Consideremos el sistema $A x = b$, con

$$A = \begin{pmatrix} 6 & 2 & -4 \\ 3 & 1 & -2 \end{pmatrix}, \quad b = \begin{pmatrix} 35 \\ 70 \end{pmatrix}.$$

La matriz

$$A^t A = \begin{pmatrix} 45 & 15 & -30 \\ 15 & 5 & -10 \\ -30 & -10 & 20 \end{pmatrix}$$

no es invertible, por lo que existe un número infinito de soluciones s de mínimos cuadrados, de la forma

$$S = \left\{ s = (s_1, s_2, s_3)^t : s_1 = \frac{28}{3} + \frac{2}{3} s_3 - \frac{1}{3} s_2 \right\}.$$

Todas ellas verifican que la $\boxed{\text{diferencia}}$ $r = A s - b = (21, -42)^t$ tiene la mínima norma posible,

$$\|r\| = \|A s - b\| = \sqrt{(21)^2 + (-42)^2} = \sqrt{2205}.$$

La $\boxed{\text{solución}}$ proporcionada por la pseudo-inversa es

$$x_0 = (6, 2, -4)^t, \quad \text{con } \|x_0\| = \sqrt{56}.$$

Si consideramos cualquier otra solución de mínimos cuadrados, por ejemplo

$$x_1 = (10, 0, 1)^t ,$$

ésta tendrá una norma mayor, en este caso $\|x_1\| = \sqrt{101}$.