Regresión Lineal
Parte II

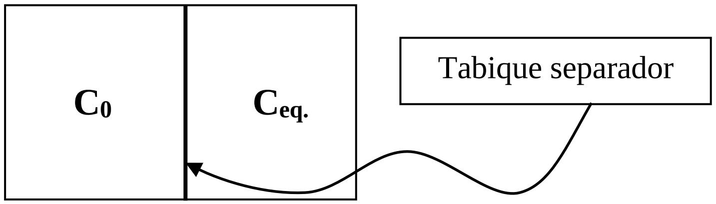
Profesor: Dr. Alejandro S. M. Santa Cruz
JTP: Dr. Juan Ignacio Manassaldi
Auxiliar: Srta. Amalia Rueda

Regresión Lineal

- En general, en análisis de datos, lo que se pretende es ajustar una curva a una serie de datos experimentales.
- Una forma de realizar el ajuste es efectuar una regresión (lineal o no lineal).
- La regresión lineal no significa que debemos necesariamente ajustar los datos a una recta, sino que desarrollamos un algoritmo que reduce el problema a tratar de resolver un SEAL.

Ejemplo: Determinación del Coeficiente de Transferencia de Materia

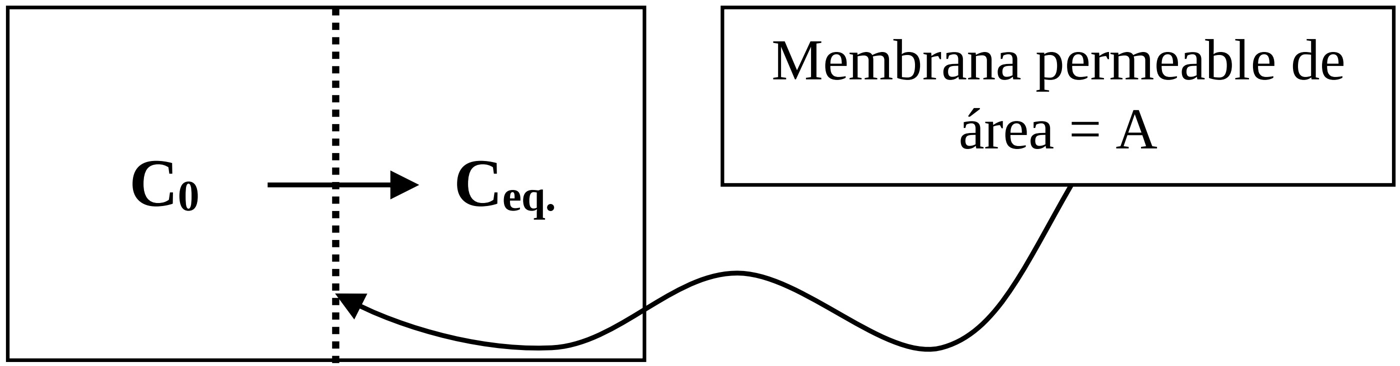
- Supongamos que hemos diseñado un experimento de transferencia de materia que consiste de dos reservorios que contienen soluciones diluidas y que están separados por un tabique divisorio que impide el intercambio de materia:



- En el reservorio de la izquierda se tiene una solución diluida con una concentración C_0 de soluto. Se supone que el volumen del reservorio de la derecha es mucho mayor que el de la izquierda y la concentración de soluto se mantiene constante en el valor $C_{eq}<C_0$.

Ejemplo: Determinación del Coeficiente de Transferencia de Materia (cont.)

- Si el volumen del reservorio de la izquierda es V y en un instante determinado ($t = 0$) se retira el tabique de separación dejando expuesta una membrana permeable al soluto, entonces comenzará el proceso de transferencia de soluto desde el reservorio de la izquierda hacia el de la derecha.



Ejemplo: Determinación del Coeficiente de Transferencia de Materia (cont.)

- La modelización del proceso de transferencia se describe a través de la siguiente ecuación diferencial:

$$V\frac{dC}{dt} = -Ah\left(C - C_{eq}\right)$$

sujeta a la condición inicial: $C(t=0)=C_0$

- Resolviendo esta ecuación diferencial se obtiene:

$$C - C_{eq} = \left(C_0 - C_{eq}\right)e^{-\left(\frac{hA}{V}t\right)}$$

- Deseamos determinar h midiendo C como función del tiempo.

- Para que podamos utilizar regresión lineal el modelo debe ser lineal en los parámetros de modelización; en este caso, es posible hacerlo transformando los datos y redefiniendo los parámetros de modelización, así

$$ln\left(C - C_{eq}\right) = ln\left(C_0 - C_{eq}\right) - \frac{hA}{V}t$$

Ejemplo: Determinación del Coeficiente de Transferencia de Materia (cont.)

- Si graficamos:

$$ln\left(C-C_{eq}\right)\ vs.\ \frac{At}{V}=t^{*}$$

obtenemos una recta de pendiente h y ordenada al origen $ln(C_0 - C_{eq})$

- ¿Como se obtiene la recta que mejor ajusta los datos experimentales?
- Veamos la desviación de los datos de la recta e intentemos minimizar esta distancia de alguna manera. Hagamos un cambio de variable,

$$y\equiv ln\left(C-C_{eq}\right)$$

y redefinamos los parámetros del modelo, así:

$$a\equiv-\frac{Ah}{V}\quad,\quad b\equiv ln\left(C_0-C_{eq}\right)$$

- Queremos ajustar los parámetros del modelo $y = at + b$ a N datos experimentales: $(t_i\ , y_i)$ con $i=1, 2, 3, \ldots, N$

Método de Mínimos Cuadrados

- Formemos la siguiente expresión:

$$E=\sum_{i=1}^{N}\left[y_i-\left(at_i+b\right)\right]^2$$

que representa la suma de los cuadrados de las distancias en ordenada entre los valores medidos de concentración y los que suministra el modelo.

- Planteamos las condiciones necesarias de existencia de extremo (mínimo) para una función de dos variables, esto es:

$$\frac{\partial E}{\partial a}=0$$

$$\frac{\partial E}{\partial b}=0$$

Método de Mínimos Cuadrados (cont.)

- Explicitamos esas derivadas:

$$\frac{\partial E}{\partial a} = \sum_{i=1}^{N} -2\left[y_i - \left(at_i + b\right)\right]t_i = \sum_{i=1}^{N} -2y_i t_i + 2a\sum_{i=1}^{N} t_i^2 + 2b\sum_{i=1}^{N} t_i = 0$$

$$\frac{\partial E}{\partial b} = \sum_{i=1}^{N} -2\left[y_i - \left(at_i + b\right)\right] = \sum_{i=1}^{N} -2y_i + 2a\sum_{i=1}^{N} t_i + 2Nb = 0$$

- Definimos los siguientes valores medios:

$$\bar{t} \equiv \frac{1}{N}\sum_{i=1}^{N} t_i \;\; ; \;\; \bar{y} \equiv \frac{1}{N}\sum_{i=1}^{N} y_i \;\; ; \;\; \overline{ty} \equiv \frac{1}{N}\sum_{i=1}^{N} t_i y_i \;\; ; \;\; \overline{t^2} \equiv \frac{1}{N}\sum_{i=1}^{N} t_i^2$$

Método de Mínimos Cuadrados (cont.)

- Se obtiene un SEAL de 2 ecuaciones con 2 incógnitas, a y b:

$$a\overline{t^2} + b\bar{t} = \overline{yt}$$

$$a\bar{t} + b = \bar{y}$$

- De aquí podemos despejar a y b, y a partir de ellos los parámetros originales del modelo, así:

$$a = \frac{\overline{yt} - \bar{y}\,\bar{t}}{\overline{t^2} - \bar{t}^2} = -\frac{Ah}{V} \quad \Rightarrow h = -\frac{Va}{A}$$

$$b = \bar{y} - a\bar{t} = ln\left(C_0 - C_{eq}\right) \Rightarrow C_0 = e^b + C_{eq}$$

Ecuaciones Normales

- Podemos hacer lo mismo que antes, pero desde otra perspectiva.
- Hagamos que los datos satisfagan el modelo, así:

$$\begin{array}{lll} y_1 & = at_1 & + b \\ y_2 & = at_2 & + b \\ \vdots & \vdots & \vdots \\ y_N & = at_N & + b \end{array}$$

- En forma matricial:

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}}_{\substack{\text{Vector} \\ \text{de} \\ \text{datos}}} \approx \underbrace{\begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_N & 1 \end{pmatrix}}_{\substack{\text{Matriz} \\ \text{de las} \\ \text{funciones} \\ \text{de modelización}}} \underbrace{\begin{pmatrix} a \\ b \end{pmatrix}}_{\substack{\text{Vector} \\ \text{de} \\ \text{parámetros}}}$$

Ecuaciones Normales (cont.)

- Reemplazaremos a los parámetros del modelo con la letra x como el vector que contiene los parámetros del modelo, así: $x_1 = a$, $x_2 = b$.
- Esto es, queremos que satisfagan el SEAL: $\mathbf{A} \,.\, \underline{x} = \underline{y}$ donde $\mathbf{A}$ representa a la matriz de las funciones de modelización e $\underline{y}$ es el vector de datos experimentales.
- En general, el número de datos experimentales, N, es mayor que el número de parámetros a ajustar, razón por la cual el sistema será un sistema sobredimensionado (mayor número de ecuaciones que de incógnitas) y por lo tanto conducirá a un sistema incompatible.
- En realidad, lo mejor que podemos hacer es minimizar el residuo $\underline{r} = \underline{y} - \mathbf{A} \,.\, \underline{x}$, o determinar $\underline{x}$ tal que la norma 2 del residuo

$$\| \underline{r} \|_2 \equiv \left\| \underline{y} - \underline{\underline{A}} . \underline{x} \right\|_2$$

sea tan pequeña como sea posible.

Ecuaciones Normales (cont.)

- En forma equivalente, queremos escoger $\underline{x}$ tal que:

$$\| \underline{r} \|_2^2 = \underline{r}^T . \underline{r} \equiv \left(\underline{y} - \underline{\underline{A}} . \underline{x} \right)^T \cdot \left(\underline{y} - \underline{\underline{A}} . \underline{x} \right) = \underline{y}^T . \underline{y} - \underline{y}^T . \underline{\underline{A}} . \underline{x} - \underline{x}^T . \underline{\underline{A}}^T . \underline{y} + \underline{x}^T . \underline{\underline{A}}^T . \underline{\underline{A}} . \underline{x}$$

  sea mínima.
- Esto es equivalente a lo que hicimos anteriormente con mínimos cuadrados.
- En efecto, planteamos la condición necesaria de existencia de extremo, pero ahora sobre la norma 2 al cuadrado del vector residuo $\underline{r}$, así:

$$\underline{\nabla}_{\underline{x}} \| \underline{r} \|_2^2 = \underline{0} \quad donde \quad \underline{\nabla}_{\underline{x}} \equiv \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \ldots, \frac{\partial}{\partial x_n} \right]$$

- Así, el problema se reduce a resolver el siguiente problema:

$$\left(\underline{\underline{A}}^T . \underline{\underline{A}} \right) \cdot \underline{x} = \underline{\underline{A}}^T \cdot \underline{y}$$

  que representa a un sistema de n ecuaciones lineales con n incógnitas (las componentes del vector $\underline{x}$ o vector de parámetros de modelización). Estas ecuaciones se conocen con el nombre de **Ecuaciones Normales**.

Ecuaciones Normales (cont.)

- Las ecuaciones normales son apropiadas, pero no son un buen camino para resolver el problema de regresión lineal para un número grande de parámetros a ajustar, ya que, en general, conducirá a elevados errores de redondeo.

- Con respecto al número de condición del sistema, se puede demostrar que,

$$\kappa\left[\left(\underline{\underline{A}}^T \cdot \underline{\underline{A}}\right),2\right]=\left[\kappa\left(\underline{\underline{A}},2\right)\right]^2$$

  igualdad que vale sólo en norma 2, dado que, en general, A es una matriz rectangular.

- Por lo tanto, el problema conducirá a un problema mal condicionado. Podemos mejorarlo cambiando de metodología.

Factorización QR de la Matriz A

Definiciones Preliminares

1. ***Transformación ortogonal:*** Una matriz ***Q*** se dice ortogonal si cumple que:

$$\underline{\underline{Q}}^T . \underline{\underline{Q}} = \underline{\underline{I}}$$

Dado que esto nos dice que:

$$\underline{\underline{Q}}^T = \underline{\underline{Q}}^{-1}$$

se deduce entonces que:

$$\underline{\underline{Q}}^T . \underline{\underline{Q}} = \underline{\underline{Q}} . \underline{\underline{Q}}^T = \underline{\underline{I}}$$

2. ***Propiedad de las transformaciones ortogonales:*** Las matrices ortogonales preservan la norma 2, esto es, $\left\| \underline{\underline{Q}} \cdot \underline{x} \right\|_2 = \left\| \underline{x} \right\|_2$

Demostración:

$$\left\| \underline{\underline{Q}} \cdot \underline{x} \right\|_2^2 = \left(\underline{\underline{Q}} \cdot \underline{x} \right)^T . \left(\underline{\underline{Q}} \cdot \underline{x} \right) = \underline{x}^T . \underline{\underline{Q}}^T \cdot \underline{\underline{Q}} \cdot \underline{x} = \underline{x}^T . \left(\underline{\underline{Q}}^T \cdot \underline{\underline{Q}} \right) \cdot \underline{x} = \underline{x}^T . \underline{\underline{I}} . \underline{x} = \underline{x}^T . \underline{x} = \left\| \underline{x} \right\|_2^2$$

Factorización QR de la Matriz A (cont.)

Definiciones Preliminares

3. Reflector elemental: Un reflector elemental ***P*** es una matriz de la forma:

$$\underline{\underline{P}} = \underline{\underline{I}} - 2\,\hat{\underline{v}} \otimes \hat{\underline{v}}^T \quad con \quad \left\| \hat{\underline{v}} \right\|_2 = 1 \; (versor)$$

donde $\otimes$ representa el producto tensorial o exterior de dos vectores.

- El producto exterior de dos vectores de $\mathbf{R}^n$ nos suministra un elemento de otro espacio vectorial, en este caso un elemento perteneciente al espacio vectorial de las matrices de $\mathbf{R}^{(nxn)}$.
- Estas matrices también se conocen con el nombre de *matrices hermíticas elementales* o *Transformaciones de Householder*. Reflejan el espacio vectorial en el hiperplano ortogonal o perpendicular a $\hat{\underline{v}}$.

Factorización QR de la Matriz A (cont.)

Definiciones Preliminares

4. ***Propiedad de los reflectores elementales:*** Dados dos vectores $\underline{x}$ e $\underline{y}$ de igual longitud se puede encontrar una matriz $\boldsymbol{P}$ tal que $\boldsymbol{P}\,.\,\underline{x} = \underline{y}$

Demostración: Adoptamos como versor de la transformación:

$$\hat{\underline{v}} = \frac{\underline{x}-\underline{y}}{\left\|\underline{x}-\underline{y}\right\|_2}$$

Luego, construimos $\underline{\underline{P}} = \left[\underline{\underline{I}} - \frac{2\left(\underline{x}-\underline{y}\right)\otimes\left(\underline{x}-\underline{y}\right)^T}{\left\|\underline{x}-\underline{y}\right\|_2^2}\right]$ y lo aplicamos al vector $\underline{x}$,

$$\left[\underline{\underline{I}} - \frac{2\left(\underline{x}-\underline{y}\right)\otimes\left(\underline{x}-\underline{y}\right)^T}{\left\|\underline{x}-\underline{y}\right\|_2^2}\right].\underline{x} = \underline{x} - \frac{2\left(\underline{x}^T.\underline{x}-\underline{y}^T.\underline{x}\right)}{\left(\underline{x}-\underline{y}\right)^T\left(\underline{x}-\underline{y}\right)}\left(\underline{x}-\underline{y}\right)$$

$$= \underline{x} - \frac{2\left(\underline{x}^T.\underline{x}-\underline{y}^T.\underline{x}\right)}{\left(\underline{x}^T.\underline{x}-\underline{y}^T.\underline{x}-\underline{x}^T.\underline{y}+\underline{y}^T.\underline{y}\right)}\left(\underline{x}-\underline{y}\right)$$

Factorización QR de la Matriz A (cont.)

Definiciones Preliminares

4. Propiedad de los reflectores elementales: Dados dos vectores $\underline{x}$ e $\underline{y}$ de igual longitud se puede encontrar una matriz $\boldsymbol{P}$ tal que $\boldsymbol{P}\,.\,\underline{x} = \underline{y}$

Demostración (cont.): Dado que:

1. $\underline{x}^T . \underline{x} = \underline{y}^T . \underline{y}$ por hipótesis.
2. $\underline{y}^T . \underline{x} = \underline{x}^T . \underline{y}$ ya que un escalar es invariante ante transposición.

Se deduce que el denominador de esta última expresión se reduce a: ; $2\left(\underline{x}^T . \underline{x} - \underline{y}^T . \underline{x}\right)$ por lo tanto resulta:

$$\left[\underline{\underline{I}} - \frac{2\left(\underline{x}-\underline{y}\right)\otimes\left(\underline{x}-\underline{y}\right)^T}{\left\|\underline{x}-\underline{y}\right\|_2^2}\right].\underline{x} = \underline{x} - \frac{2\left(\underline{x}^T . \underline{x} - \underline{y}^T . \underline{x}\right)}{2\left(\underline{x}^T . \underline{x} - \underline{y}^T . \underline{x}\right)}\left(\underline{x}-\underline{y}\right) = \underline{x} - \left(\underline{x}-\underline{y}\right) = \underline{y}$$

Factorización QR de la Matriz A (cont.)

TAREA: Probar que los reflectores elementales son matrices simétricas y ortogonales.

Factorización QR de la Matriz A (cont.)

- Nos formulamos la siguiente pregunta: ¿Para qué necesitamos esta propiedad de los reflectores elementales en el problema de regresión lineal?
- Dado que las matrices ortogonales preservan la norma vectorial 2, entonces, el problema de resolver:

$$\underset{\underline{x}}{mín}\left\| \underline{\underline{A}}.\underline{x} - \underline{b} \right\|_2$$

es equivalente a resolver:

$$\underset{\underline{x}}{mín}\left\| \underline{\underline{A}}.\underline{x} - \underline{b} \right\|_2 = \underset{\underline{x}}{mín}\left\| \underline{\underline{P}}.\left(\underline{\underline{A}}.\underline{x} - \underline{b}\right) \right\|_2 = \underset{\underline{x}}{mín}\left\| \left(\underline{\underline{P}}.\underline{\underline{A}}\right).\underline{x} - \left(\underline{\underline{P}}.\underline{b}\right) \right\|_2$$

donde ***P*** es una matriz ortogonal, pero no cualquier matriz ortogonal.

Factorización QR de la Matriz A (cont.)

- Queremos encontrar ***P*** tal que ***P.A*** sea de forma ***triangular superior***. ¿Por qué?
- Supongamos que existe esa matriz ortogonal ***P*** tal que:

$$\underline{\underline{P}}.\underline{\underline{A}}=\begin{pmatrix}\underline{\underline{R}}\\ \underline{\underline{0}}\end{pmatrix}=\begin{pmatrix}3&4&1\\0&2&2\\0&0&7\\0&0&0\\0&0&0\end{pmatrix}\quad y\quad \underline{\underline{P}}.\underline{b}=\underline{c}=\begin{pmatrix}14\\10\\21\\6\\2\end{pmatrix}$$

- Entonces el problema de regresión se reduciría a encontrar el vector $\underline{x}$ que minimiza $\left\|\begin{pmatrix}\underline{\underline{R}}\\ \underline{\underline{0}}\end{pmatrix}.\underline{x}-\underline{c}\right\|_2$ donde $\underline{x}$ es un vector de 3 componentes (igual al número de columnas de ***R***).

Factorización QR de la Matriz A (cont.)

- Por otra parte, resolver:

$$\min_{\underline{x}} \left\| \begin{pmatrix} \underline{\underline{R}} \\ \underline{\underline{0}} \end{pmatrix} . \underline{x} - \underline{c} \right\|_2$$

sería equivalente a resolver:

$$\min_{\underline{x}} \left\| \begin{pmatrix} \underline{\underline{R}} \\ \underline{\underline{0}} \end{pmatrix} . \underline{x} - \underline{c} \right\|_2^2$$

- Observemos que las tres primeras ecuaciones se pueden resolver exactamente y que las dos últimas filas no tienen efecto sobre $\underline{x}$, así

$$\left\| \begin{pmatrix} \underline{\underline{R}} \\ \underline{\underline{0}} \end{pmatrix} . \underline{x} - \underline{c} \right\|_2^2 = \left\| \begin{pmatrix} 3 & 4 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} - \begin{pmatrix} 14 \\ 10 \\ 21 \end{pmatrix} \right\|_2^2 + \left\| \begin{matrix} 6 \\ 2 \end{matrix} \right\|_2^2$$

Factorización QR de la Matriz A (cont.)

- En efecto, la norma al cuadrado es la suma de dos términos no negativos.
- Para que esta expresión sea mínima el primer término debería ser nulo; esta condición se cumple si $\underline{x}$ satisface el SEAL.

$$\begin{pmatrix} 3 & 4 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 7 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 14 \\ 10 \\ 21 \end{pmatrix}$$

- Podemos resolver por sustitución hacia atrás:

$$x_3 = \frac{21}{7} = 3 \quad ; \quad x_2 = \frac{10-6}{2} = 2 \quad ; \quad x_1 = \frac{14-3-8}{3} = 1$$

- El residuo es:

$$\left\| \begin{matrix} 6 \\ 2 \end{matrix} \right\|_2 = \sqrt{40}$$

Factorización QR de la Matriz A (cont.)

- Entonces, esta clase de problema es fácil de resolver si podemos encontrar una matriz ***P*** tal que reduzca a la matriz ***A*** de las funciones de modelización a una forma triangular superior.

- Con ayuda de este resultado podemos construir una *secuencia de transformaciones de Householder* P_1, P_2, P_3, …, P_n tal que al ser aplicadas a la matriz ***A***, una columna a la vez, la reduzcan a una matriz triangular superior:

$$\underline{\underline{P_n}} \cdots \underline{\underline{P_2}} \cdot \underline{\underline{P_1}} \cdot \underline{\underline{A}} = \begin{pmatrix} \underline{\underline{R}} \\ \underline{\underline{0}} \end{pmatrix}$$

- Dado que cada ***P*** es ortogonal, resulta:

$$\left(\underline{\underline{P_n}} \cdots \underline{\underline{P_3}} \cdot \underline{\underline{P_2}} \cdot \underline{\underline{P_1}}\right)^T \cdot \begin{pmatrix} \underline{\underline{R}} \\ \underline{\underline{0}} \end{pmatrix} = \underline{\underline{P_1}}^T \cdot \underline{\underline{P_2}}^T \cdots \underbrace{\underline{\underline{P_n}}^T \cdot \underline{\underline{P_n}}}_{\underline{\underline{I}}} \cdots \underline{\underline{P_2}} \cdot \underline{\underline{P_1}} \cdot \underline{\underline{A}} = \underline{\underline{A}}$$

Factorización QR de la Matriz A (cont.)

- Luego, definimos:

$$\underline{\underline{Q}} \equiv \left(\underline{\underline{P}}_n \cdots \underline{\underline{P}}_3 \cdot \underline{\underline{P}}_2 \cdot \underline{\underline{P}}_1 \right)^T$$

tal que: $\underline{\underline{A}} \equiv \underline{\underline{Q}} \cdot \begin{pmatrix} \underline{\underline{R}} \\ \underline{\underline{0}} \end{pmatrix}$ con $\underline{\underline{Q}}^T \cdot \underline{\underline{Q}} = \underline{\underline{I}}$

- Supongamos por el momento que hemos encontrado una matriz $\boldsymbol{Q}$ de esas características.

- Si consideramos el SEAL: $\boldsymbol{A} \cdot \underline{x} = \underline{b}$ entonces, resulta:

$$\underline{\underline{Q}}^T \cdot \underline{\underline{A}} = \begin{pmatrix} \underline{\underline{R}} \\ \underline{\underline{0}} \end{pmatrix} \quad y \quad \underline{\underline{Q}}^T \cdot \underline{b} = \begin{pmatrix} \underline{c} \\ \underline{d} \end{pmatrix}$$

Esto es, separamos al vector $\boldsymbol{Q}^T \cdot \underline{b}$ en un vector $\underline{c}$ cuyo número de componentes es igual al número de filas de $\boldsymbol{R}$ y un vector $\underline{d}$ cuyo número de componentes es igual al número de filas nulas.

Factorización QR de la Matriz A (cont.)

- Planteamos el vector residuo: $\underline{r} = A\,.\,\underline{x} - \underline{b}$ de donde obtenemos:

$$\underline{\underline{Q}}^T.\underline{r} = \underline{\underline{Q}}^T\left(\underline{\underline{A}}\cdot\underline{x}-\underline{b}\right) = \left(\underline{\underline{Q}}^T.\underline{\underline{A}}\right)\cdot\underline{x}-\underline{\underline{Q}}^T.\underline{b} = \begin{pmatrix}\underline{\underline{R}}\\ \underline{\underline{0}}\end{pmatrix}\cdot\underline{x}-\begin{pmatrix}\underline{c}\\ \underline{d}\end{pmatrix} = \begin{pmatrix}\underline{\underline{R}}\cdot\underline{x}-\underline{c}\\ -\underline{d}\end{pmatrix}$$

- Dado que $\boldsymbol{Q}$ es ortogonal,

$$\|\underline{r}\|_2^2 = \left\|\underline{\underline{Q}}^T.\underline{r}\right\|_2^2 = \left[\left(\underline{\underline{R}}\cdot\underline{x}-\underline{c}\right)^T - \quad \underline{d}^T\right].\begin{pmatrix}\underline{\underline{R}}\cdot\underline{x}-\underline{c}\\ -\underline{d}\end{pmatrix} = \left\|\underline{\underline{R}}\cdot\underline{x}-\underline{c}\right\|_2^2 + \left\|\underline{d}\right\|_2^2$$

- Planteamos ahora el problema de regresión:

$$\min_{\underline{x}\neq\underline{0}}\|\underline{r}\|_2^2$$

- Esta última condición se verifica si: $\boldsymbol{R}\,.\,\underline{x} = \underline{c}$, o bien, $\underline{x} = \boldsymbol{R}^{-1}\,.\,\underline{c}$ con el vector residuo dado por:

$$\underline{r} = \underline{\underline{Q}}\cdot\begin{pmatrix}\underline{0}\\ -\underline{d}\end{pmatrix}$$

Factorización QR de la Matriz A (cont.)

- El sistema de ecuaciones lineales $\underline{\underline{R}} \cdot \underline{x} = \underline{c}$ está mejor acondicionado que las ecuaciones normales $\left(\underline{\underline{A}}^T \cdot \underline{\underline{A}}\right) \cdot \underline{x} = \underline{\underline{A}}^T \cdot \underline{y}$
- En efecto, podemos escribir:

$$\underline{\underline{A}}^T \cdot \underline{\underline{A}} = \underline{\underline{A}}^T \cdot \underbrace{\underline{\underline{Q}} \cdot \underline{\underline{Q}}^T}_{\underline{\underline{I}}} \cdot \underline{\underline{A}} = \left(\underline{\underline{Q}}^T \cdot \underline{\underline{A}}\right)^T \cdot \left(\underline{\underline{Q}}^T \cdot \underline{\underline{A}}\right) = \underline{\underline{R}}^T \cdot \underline{\underline{R}}$$

- Luego se verifica que:

$$\kappa\left[\left(\underline{\underline{A}}^T \cdot \underline{\underline{A}}\right)\right] = \kappa\left[\left(\underline{\underline{R}}^T \cdot \underline{\underline{R}}\right)\right] = \kappa^2\left(\underline{\underline{R}}\right)$$

donde κ representa al número de condición en norma 2, definido como la relación entre el mayor y el menor valor singular de la matriz,

$$\kappa = \frac{|\sigma_1|}{|\sigma_n|}$$

- De aquí que el método se recomienda para resolver problemas de regresión lineal dado que:

$$\kappa\left(\underline{\underline{R}}\right) = \sqrt{\kappa\left(\underline{\underline{A}}^T \cdot \underline{\underline{A}}\right)}$$

Problemas Degenerados

¿Qué sucede si el problema está sub-determinado? En este caso, la matriz $\boldsymbol{R}$ no es de rango completo.

En general, podemos decir que $\boldsymbol{R}$ tendrá una forma trapezoidal en lugar de una triangular, así:

$$\begin{bmatrix} \underline{\underline{R}} \\ \underline{\underline{0}} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Algunas definiciones

1. *Rango de una matriz:* Dada una matriz cualquiera $\boldsymbol{A}$ de orden (mxn), se denomina rango de la matriz $\boldsymbol{A}$ y se denota $rg(\boldsymbol{A})$ al máximo número de vectores columna (fila) linealmente independientes. Además, se verifica que $rg_f(\boldsymbol{A}) = rg_c(\boldsymbol{A})$ y que $rg(\boldsymbol{A}) \leq mín(m,n)$

2. *Definición alternativa de rango de una matriz:* Orden del determinante no nulo de máximo orden que se puede construir con las filas y las columnas de $\boldsymbol{A}$.

Problemas Degenerados (cont.)

Clasificación de las matrices en función del rango

Sea $\boldsymbol{A} \in M^{(mxn)}$ entonces,

a) Si $m \neq n$ y $r_g(\boldsymbol{A}) = mín(m,n)$ diremos que la matriz es de rango completo; en caso contrario diremos que A no es de rango completo.

b) Si $m = n$ entonces:

- b.1) *Si* $rg(\boldsymbol{A}) = n$ diremos que $\boldsymbol{A}$ es una matriz no singular o regular.
- b.2) Si $rg(\boldsymbol{A}) < n$ diremos que $\boldsymbol{A}$ es una matriz singular.

Problemas Degenerados (cont.)

- Los problemas degenerados en regresión lineal surgen si las columnas de la matriz de las funciones de modelización son linealmente dependientes.
- Para estos problemas habrá un infinito número de modos de aproximar los datos con el mismo residuo (mínimo).

Problemas Degenerados (cont.)

- Sea el modelo:

$$b(t) = x_1 \cdot (1) + x_2 \cdot (t) + x_3 \cdot (2t+1)$$

- Supongamos que se dispone de N datos experimentales (t_i , b_i) con $i=1, 2, 3, \ldots , N$ ($N > 3$ = *nro. de parámetros*); por consiguiente éstos deberán satisfacer al modelo propuesto, esto es:

$$b_1 = b(t_1) = x_1 \cdot (1) + x_2 \cdot (t_1) + x_3 \cdot (2t_1 + 1)$$
$$b_2 = b(t_2) = x_1 \cdot (1) + x_2 \cdot (t_2) + x_3 \cdot (2t_2 + 1)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$b_N = b(t_N) = x_1 \cdot (1) + x_2 \cdot (t_N) + x_3 \cdot (2t_N + 1)$$

En forma matricial:

$$\underset{\underline{b}}{\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix}} = \underbrace{\begin{pmatrix} 1 & t_1 & 2t_1+1 \\ 1 & t_2 & 2t_2+1 \\ \vdots & \vdots & \vdots \\ 1 & t_N & 2t_N+1 \end{pmatrix}}_{\underline{\underline{A}}} \cdot \underset{\underline{x}}{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}$$

Problemas Degenerados (cont.)

- Entonces, si $\underline{x}^* = (x_1^*,\ x_2^*,\ x_3^*)$ es solución del problema de regresión, también lo es:

$$\underline{x} = \underline{x}^* + \alpha \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \quad \forall \alpha \in R$$

En efecto,

$$\underline{\underline{A}}.\underline{x} - \underline{b} = \underline{\underline{A}}.\left[\underline{x}^* + \alpha \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \right] - \underline{b} = \underline{\underline{A}}.\underline{x}^* + \underbrace{\underline{\underline{A}}.\begin{pmatrix} \alpha \\ 2\alpha \\ -\alpha \end{pmatrix}}_{\equiv \underline{0}} - \underline{b} = \underline{\underline{A}}.\underline{x}^* - \underline{b}$$

- Concluimos que $\underline{x}^*$ y $\underline{x}$ suministran el mismo residuo (mínimo).

Problemas Degenerados (cont.)

- El límite entre la degeneración y la no-degeneración puede estar desdibujado por los errores de redondeo.
- Por ejemplo, supongamos que tenemos el siguiente SEAL:

$$\begin{pmatrix} 1 & 1 \\ 0 & 10^{-k} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad con\ k>1$$

- Este SEAL tiene como solución:

$$x_2 = 10^k \quad ; \quad x_1 = \frac{1-10^k}{1} \approx -10^k$$

- Ahora supongamos que $k>>1$, siendo a_{22} distinto de cero debido a errores de redondeo (en realidad debería considerarse 0).
- En esta situación preferimos ajustar los datos con $x_2 = 0$ y $x_1 = 1$ puesto que en este caso $\|\underline{x}\|_2 = 1$ en lugar de $\sqrt{2}10^k$

Problemas Degenerados (cont.)

- Concluimos que en estos casos lo que se estila es adoptar para el modelo parámetros pequeños, esto es, para problemas degenerados deseamos que el vector de parámetros de modelización tenga la longitud más corta.

Problemas Degenerados (cont.)

- Algunas veces resulta difícil determinar por simple observación cuán cercana se halla una matriz de la degeneración. Consideremos la matriz:

$$\underline{\underline{A}} = \begin{pmatrix} 1 & -1 & -1 & -1 & \cdots & -1 \\ 0 & 1 & -1 & -1 & \cdots & -1 \\ 0 & 0 & 1 & -1 & \cdots & -1 \\ 0 & 0 & 0 & 1 & \cdots & -1 \\ 0 & 0 & 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- El determinante de esta matriz es igual a 1.
- Si $a_{n1} = -2^{-(n-2)}$ la matriz es singular.
- A pesar que es de forma triangular superior, aún podemos efectuarle una factorización ***QR***, pero en este caso aplicando pivoteo de columna.

Problemas Degenerados (cont.)

- Para ello, elegimos la columna de mayor norma, la movemos a la columna 1 y la reducimos (mediante transformaciones de Householder).
- Luego seleccionamos la próxima columna más grande de la submatriz restante [nx(n-1)] y la movemos a la columna 2, y así sucesivamente. Este procedimiento nos conduce a la siguiente matriz:

$$\begin{pmatrix} 1 & -1 & -1 & -1 & -1 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \Rightarrow \begin{pmatrix} 2.24 & 0.89 & 0.45 & 0 & -0.45 \\ 0 & 1.79 & 0.34 & 0 & -0.34 \\ 0 & 0 & -1.64 & 0 & 0.42 \\ 0 & 0 & 0 & -1.4 & 0.71 \\ 0 & 0 & 0 & 0 & 0.108 \end{pmatrix}$$

- Obsérvese que el último elemento de la diagonal es un orden de magnitud más pequeño que los otros elementos diagonales. Esto se debe a que la matriz $\boldsymbol{A}$ se halla próxima a la singularidad.

Problemas Degenerados (cont.)

- Lo que hicimos es equivalente a efectuar la factorización ***QR*** de la matriz ***A*** con pivoteo de columna, esto es,

$$\underline{\underline{A}} = \underline{\underline{Q}}.\underline{\underline{R}}.\underline{\underline{P}}^T$$

donde ***P*** representa a una matriz de permutación de columnas.

- Por otra parte, si ***A*** es singular o no es de rango completo, ***R*** es de la forma:

$$\underline{\underline{R}} = \begin{pmatrix} x & x & x & x & x & x \\ 0 & x & x & x & x & x \\ 0 & 0 & x & x & x & x \\ 0 & 0 & 0 & x & x & x \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

- El número de filas no nulas de ***R*** es el rango de ***A***.

Problemas Degenerados (cont.)

- Supongamos ahora que en la factorización ***QR*** con pivoteo de columna de la matriz ***A*** representamos la estructura de la matriz ***R*** de arriba de la siguiente manera:

$$\underline{\underline{R}} = \begin{pmatrix} x & x & x & x & x & x \\ 0 & x & x & x & x & x \\ 0 & 0 & x & x & x & x \\ 0 & 0 & 0 & x & x & x \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \underline{\underline{R_1}} & \underline{\underline{R_2}} \\ \underline{\underline{0}} & \underline{\underline{0}} \end{pmatrix}$$

donde $\boldsymbol{R_1}$ representa a una matriz triangular superior y $\boldsymbol{R_2}$ a una matriz rectangular no nula.

Problemas Degenerados (cont.)

- La factorización con pivoteo de columna se puede utilizar para resolver el problema de mínimos cuadrados, así:

$$\min_{\underline{x}\neq\underline{0}} \left\| \underline{\underline{A}}.\underline{x} - \underline{b} \right\|_2 = \min_{\underline{x}\neq\underline{0}} \left\| \underline{\underline{Q}}^T.\left(\underline{\underline{A}}.\underline{x} - \underline{b}\right) \right\|_2 =$$

$$= \min_{\underline{x}\neq\underline{0}} \left\| \underline{\underline{Q}}^T.\left(\underline{\underline{Q}}.\underline{\underline{R}}.\underline{\underline{P}}^T\right).\underline{x} - \underline{\underline{Q}}^T.\underline{b} \right\|_2$$

$$= \min_{\underline{x}\neq\underline{0}} \left\| \underline{\underline{R}}.\underline{\underline{P}}^T.\underline{x} - \underline{\underline{Q}}^T.\underline{b} \right\|_2$$

- Definimos un nuevo vector de parámetros:

$$\underline{y} \equiv \underline{\underline{P}}^T.\underline{x}$$

- Y un nuevo vector de datos,

$$\underline{c} \equiv \underline{\underline{Q}}^T.\underline{b}$$

- Particionamos a los vectores $\underline{y}$ y $\underline{c}$ en dos partes, de esta manera:

$$\left.\underline{y} = \begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \end{pmatrix} \;;\; \underline{c} = \begin{pmatrix} \underline{c}_1 \\ \underline{c}_2 \end{pmatrix}\right\} \textit{El nro. de componentes de } \underline{y}_1 \textit{ y } \underline{c}_1 \textit{ es igual al nro. de filas no nulas de } \underline{\underline{R}}$$

Problemas Degenerados (cont.)

- Por consiguiente, resolver el problema de regresión:

$$\min_{\underline{x}\neq\underline{0}} \left\| \underline{\underline{A}}.\underline{x} - \underline{b} \right\|_2^2$$

es equivalente a resolver:

$$\min_{\underline{y}\neq\underline{0}} \left\| \begin{pmatrix} \underline{\underline{R_1}} & \underline{\underline{R_2}} \\ \underline{\underline{0}} & \underline{\underline{0}} \end{pmatrix} . \begin{pmatrix} \underline{y_1} \\ \underline{y_2} \end{pmatrix} - \begin{pmatrix} \underline{c_1} \\ \underline{c_2} \end{pmatrix} \right\|_2^2 = \min_{\underline{y}\neq\underline{0}} \left[\left\| \underline{\underline{R_1}}.\underline{y_1} + \underline{\underline{R_2}}.\underline{y_2} - \underline{c_1} \right\|_2^2 + \left\| \underline{c_2} \right\|_2^2 \right]$$

- El segundo término no puede ser afectado por los parámetros $\underline{y}$. Sin embargo el primer término puede hacerse cero de infinitas maneras.
- En efecto, si fijamos $\underline{y_2}$ en forma arbitraria, entonces se obtiene resolviendo el siguiente SEAL:

$$\underline{\underline{R_1}}.\underline{y_1} + \underline{\underline{R_2}}.\underline{y_2} = \underline{c_1} \;\Rightarrow\; \underline{\underline{R_1}}.\underline{y_1} = \underline{c_1} - \underline{\underline{R_2}}.\underline{y_2}$$

Problemas Degenerados (cont.)

- En particular, si eligiésemos $\underline{y}_2 = \underline{0}$ el sistema anterior se reduce a:

$$\underline{\underline{R}}_1 \cdot \underline{y}_1 = \underline{c}_1$$

que resolvemos por sustitución hacia atrás.

- Pero todavía el problema no está resuelto. En efecto, tenemos que:

$$\underline{y} = \begin{pmatrix} \underline{y}_1 \\ \underline{0} \end{pmatrix}$$

- Luego, de la definición del vector $\underline{y}$ resulta:

$$\underline{x} = \underline{\underline{P}} \cdot \underline{y}$$

ya que la inversa de $\boldsymbol{P}^T$ es $\boldsymbol{P}$ por ser matrices de permutación (en este caso de columnas). Esto es, reordenamos variables para obtener finalmente los parámetros del modelo siendo $\left\| \underline{c}_2 \right\|$ el residuo.

Problemas Degenerados
Descomposición en valores singulares de A

- Una forma alternativa de resolver problemas de regresión lineal degenerados es plantear la *descomposición en valores singulares de la matriz* $A \in M^{(mxn)}$ con $m>n$.

$$\underset{m\times n}{\underline{\underline{A}}} = \underset{\substack{ortogonal \\ m\times m}}{\underline{\underline{U}}} \cdot \overset{\substack{diagonal \\ m\times n}}{\underline{\underline{\Sigma}}} \cdot \underset{\substack{ortogonal \\ n\times n}}{\underline{\underline{V}}^T}$$

- donde $\boldsymbol{U}$ y $\boldsymbol{V}$ representan matrices ortogonales, mientras que $\boldsymbol{\Sigma}$ es una matriz diagonal cuyos elementos diagonales representan a los valores singulares de la descomposición.
- Mediante la utilización de pivoteo de columna, los elementos de $\boldsymbol{\Sigma}$ se pueden ordenar de manera decreciente, así: $|\sigma_1| \geq |\sigma_2| \geq |\sigma_3| \geq \cdots \geq |\sigma_n|$
- Si $\sigma_n = 0$, luego $\boldsymbol{A}$ no es de rango completo!!
- Esto nos dice que estamos ante un problema degenerado.

Problemas Degenerados

Descomposición en valores singulares de A (cont.)

- Resolvamos el problema de regresión lineal:

$$\min_{\underline{x}\neq\underline{0}} \left\| \underline{\underline{A}}.\underline{x}-\underline{b} \right\|_2 = \min_{\underline{x}\neq\underline{0}} \left\| \underline{\underline{U}}.\underline{\underline{\Sigma}}.\underline{\underline{V}}^T.\underline{x}-\underline{b} \right\|_2 = \min_{\underline{x}\neq\underline{0}} \left\| \underline{\underline{U}^T}.\left(\underline{\underline{U}}.\underline{\underline{\Sigma}}.\underline{\underline{V}}^T.\underline{x}-\underline{b}\right) \right\|_2 = \min_{\underline{x}\neq\underline{0}} \left\| \underline{\underline{\Sigma}}.\underline{\underline{V}}^T.\underline{x}-\underline{\underline{U}}^T.\underline{b} \right\|_2$$

- Definamos dos nuevos vectores:

$$\underline{z} \equiv \underline{\underline{V}}^T.\underline{x} \quad y \quad \underline{d} \equiv \underline{\underline{U}}^T.\underline{b}$$

- Por lo tanto, resulta:

$$\min_{\underline{x}\neq\underline{0}} \left\| \underline{\underline{A}}.\underline{x}-\underline{b} \right\|_2 = \min_{\underline{z}\neq\underline{0}} \left\| \underline{\underline{\Sigma}}.\underline{z}-\underline{d} \right\|_2$$

- En forma equivalente,

$$\min_{\underline{x}\neq\underline{0}} \left\| \underline{\underline{A}}.\underline{x}-\underline{b} \right\|_2^2 = \min_{\underline{z}\neq\underline{0}} \left\| \underline{\underline{\Sigma}}.\underline{z}-\underline{d} \right\|_2^2$$

Problemas Degenerados

Descomposición en valores singulares de A (cont.)

- Separamos a los vectores $\underline{d}$ y $\underline{z}$ en dos vectores de manera que el número de componentes del primer vector sea igual al número de filas no nulas de $\mathbf{\Sigma}$ y el número de componentes del segundo igual al número de filas nulas de $\mathbf{\Sigma}$, luego resulta:

$$\min_{\underline{z}\neq 0}\left\|\begin{pmatrix}\sigma_1 & 0 & 0 & 0 & 0\\ 0 & \sigma_2 & 0 & 0 & 0\\ 0 & 0 & \cdots & 0 & 0\\ 0 & 0 & 0 & \sigma_{n-1} & 0\\ 0 & 0 & 0 & 0 & \sigma_n\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{pmatrix}.\begin{pmatrix}\underline{z}_1\\ \underline{z}_2\end{pmatrix}-\begin{pmatrix}\underline{d}_1\\ \underline{d}_2\end{pmatrix}\right\|_2^2=\min_{\underline{z}\neq 0}\left\{\left\|\begin{pmatrix}\sigma_1 & 0 & 0 & 0 & 0\\ 0 & \sigma_2 & 0 & 0 & 0\\ 0 & 0 & \cdots & \vdots & 0\\ 0 & 0 & \cdots & \sigma_{n-1} & 0\\ 0 & 0 & \cdots & 0 & \sigma_n\end{pmatrix}.\underline{z}_1-\underline{d}_1\right\|_2^2+\left\|\underline{d}_2\right\|_2^2\right\}$$

- Si $\sigma_n \neq 0$ *estamos ante un problema no degenerado* ya que resolver el problema anterior equivale a resolver el SEAL:

$$\begin{pmatrix}\sigma_1 & 0 & \cdots & \cdots & 0\\ 0 & \sigma_2 & 0 & \cdots & 0\\ 0 & \cdots & \cdots & \cdots & 0\\ 0 & 0 & \cdots & \cdots & 0\\ 0 & 0 & \cdots & 0 & \sigma_n\end{pmatrix}.\underline{z}_1=\underline{d}_1$$

Problemas Degenerados
Descomposición en valores singulares de A (cont.)

- Dado que $\underline{z}_2$ no intervine en el cálculo, podemos considerarlo cero.
- Si resolvemos el SEAL despejando los $z_{1,i}$ obtenemos:

$$z_{1,i} = \frac{d_i}{\sigma_i} \;;\; i=1,2,\ldots,n$$

- y finalmente el vector de parámetros $\underline{x}$ haciendo:

$$\underline{\underline{V}}.\underline{\underline{V}}^T.\underline{x} \equiv \underline{x} = \underline{\underline{V}}.\underline{z} \text{ siendo } \underline{z} = \begin{pmatrix} \underline{z}_1 \\ \underline{0} \end{pmatrix}$$

ya que $\boldsymbol{V}$ es una matriz ortogonal.

- El residuo en este caso es:

$$\left\|r\right\|_2^2 = \left\|\underline{d}_2\right\|_2^2 = \sum_{i=n+1}^{m} d_i^2$$

Problemas Degenerados
Descomposición en valores singulares de A (cont.)

- Si $\sigma_n = 0$ estamos en presencia de un problema degenerado.
- Si $\sigma_n \sim 0$ o del orden del ε épsilon de la computadora podríamos considerarlo cero dado que no interviene en el cálculo de $\underline{x}$. Esto es,

$$z_{1,i} = \begin{cases} \dfrac{d_i}{\sigma_i} & si\ \sigma_i > \varepsilon \\ 0 & si\ \sigma_i < \varepsilon \end{cases}$$

- Esta forma de determinar los parámetros tiene la propiedad que entre el conjunto de vectores $\underline{x}^*$ que minimizan el residuo, éste tendrá la mínima norma, esto es:

$$\left\| \underline{x}_{SVD} \right\|_2 \leq \left\| \underline{x} \right\|_2$$