Problemas de Mínimos Cuadrados

- **Ajuste de curvas.**
- **Sistemas rectangulares:** Solución en el sentido de mínimos cuadrados. Ecuaciones normales.
- **Ortogonalización:** Gram-Schmidt. Factorización **QR**.
- **Problemas de cuadrados mínimos no lineales:** Reducción a problemas lineales.

Considere un sistema rectangular de ecuaciones

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

donde $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, con $n < m$, es una matriz rectangular de m filas y n columnas y $\boldsymbol{b} \in \mathbb{R}^m$.

Este problema, en general, no tiene solución: **sistema sobredeterminado**.

Una alternativa es buscar una solución en el sentido generalizado siguiente:

Hallar $\boldsymbol{x} \in \mathbb{R}^n$ tal que $\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2$ sea **mínima**.

Definición. El vector $\boldsymbol{x}$ que minimiza $\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2$ es la **solución en el sentido de mínimos cuadrados** del sistema rectangular.

Ojo! En general:

$$\boldsymbol{A}\boldsymbol{x} \neq \boldsymbol{b}.$$

Ejemplo. Ajuste de polinomios.

Dado un conjunto de puntos

$$(x_1, y_1), \ldots, (x_m, y_m),$$

nos proponemos encontrar el polinomio

$$p(x) = c_0 + c_1 x + \ldots + c_{n-1} x^{n-1},$$

con $n < m$ que esté **más cerca** de estos puntos en el sentido que

$$\sum_{i=1}^{m} |p(x_i) - y_i|^2,$$

sea mínima.

Esta suma de cuadrados es el cuadrado de la norma del residuo del sistema rectangular:

$$\begin{pmatrix} 1 & x_1 & \dots & x_1^{n-1} \\ 1 & x_2 & \dots & x_2^{n-1} \\ 1 & x_3 & \dots & x_3^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \dots & x_m^{n-1} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{pmatrix}.$$

Teorema. *Sean* $\boldsymbol{A} \in \mathbb{R}^{m \times n}(m \geq n)$ *y* $\boldsymbol{b} \in \mathbb{R}^m$*. Un vector* $\boldsymbol{x} \in \mathbb{R}^n$ *minimiza la norma del residuo* $\|\boldsymbol{r}\|_2 = \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2$ *si y sólo si el residuo* $\boldsymbol{r}$ *es ortogonal a la imagen de* $\boldsymbol{A}$*; esto es si*

$$\boldsymbol{A}^{\mathrm{t}}\boldsymbol{r} = \boldsymbol{0},$$

donde $\boldsymbol{A}^{\mathrm{t}}$ *es la matriz transpuesta de* $\boldsymbol{A}$*.*

Consecuencia: $\boldsymbol{x}$ debe satisfacer

$$\boldsymbol{A}^{\mathrm{t}}\boldsymbol{r} = \boldsymbol{0} \quad \Longleftrightarrow \quad \boldsymbol{A}^{\mathrm{t}}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) = \boldsymbol{0} \quad \Longleftrightarrow \quad \boldsymbol{A}^{\mathrm{t}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{t}}\boldsymbol{b}.$$

Estas últimas ecuaciones reciben el nombre de **ecuaciones normales**.

Observación: En el caso en que $m = n$ y que la matriz $\boldsymbol{A}$ sea una matriz no singular, entonces las ecuaciones normales entregan como solución la solución del sistema lineal $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.

Las ecuaciones normales tienen solución única si y sólo si todas las columnas de $\boldsymbol{A}$ son l.i.; es decir, si rango$(\boldsymbol{A}) = n$.

En este caso, además, la matriz $\boldsymbol{A}^{\mathrm{t}}\boldsymbol{A}$ es **simétrica y definida positiva**, de donde, las ecuaciones normales tienen solución única y se pueden utilizar los métodos estudiados para estas matrices, en particular, el **método de Cholesky**.

Para resolver las ecuaciones normales se puede proceder del siguiente modo:

1. Calcular la matriz $\boldsymbol{A}^{\mathrm{t}}\boldsymbol{A}$ y el vector $\boldsymbol{A}^{\mathrm{t}}\boldsymbol{b}$.
2. Obtener la matriz $\boldsymbol{L}$ de la factorización de Cholesky: $\boldsymbol{A}^{\mathrm{t}}\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{t}}$.
3. Resolver el sistema triangular inferior $\boldsymbol{L}\boldsymbol{y} = \boldsymbol{A}^{\mathrm{t}}\boldsymbol{b}$.
4. Resolver el sistema triangular superior $\boldsymbol{L}^{\mathrm{t}}\boldsymbol{x} = \boldsymbol{y}$.

Inconveniente: El condicionamiento de la matriz $\boldsymbol{A}^{\mathrm{t}}\boldsymbol{A}$ es en general malo, lo que genera gran sensibilidad respecto a errores de redondeo.

Por ejemplo, si $\boldsymbol{A}$ es cuadrada, entonces $\mathrm{cond}_2(\boldsymbol{A}^{\mathrm{t}}\boldsymbol{A}) = \mathrm{cond}_2(\boldsymbol{A})^2$.

Solución: **factorización QR**. Ortogonalizar las columnas de $\boldsymbol{A}$ mediante, por ejemplo, **Gram-Schmidt**.

Para esto, escribamos:

$$\boldsymbol{A} = \left(\begin{array}{c|c|c|c} \\ \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \\ \\ \\ \end{array}\right) \in \mathbb{R}^{m\times n}$$

donde $\boldsymbol{a}_i \in \mathbb{R}^m$, $i = 1, 2, \dots, n$, son las columnas de la matriz $\boldsymbol{A}$.

La idea es construir una matriz

$$\boldsymbol{Q} = \left(\begin{array}{c|c|c|c} & & & \\ & & & \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \dots & \boldsymbol{q}_n \\ & & & \\ & & & \end{array} \right) \in \mathbb{R}^{m \times n}$$

y una matriz triangular superior

$$\boldsymbol{R} = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ 0 & r_{22} & \dots & r_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & r_{nn} \end{pmatrix} \in \mathbb{R}^{n \times n},$$

tales que

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}.$$

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R} \quad \Longleftrightarrow \quad \left\{ \begin{array}{rcl} \boldsymbol{a}_1 & = & r_{11}\boldsymbol{q}_1, \\ \boldsymbol{a}_2 & = & r_{12}\boldsymbol{q}_1 + r_{22}\boldsymbol{q}_2, \\ & \vdots & \\ \boldsymbol{a}_n & = & r_{1n}\boldsymbol{q}_1 + r_{2n}\boldsymbol{q}_2 + \ldots + r_{nn}\boldsymbol{q}_n. \end{array} \right.$$

$\boldsymbol{Q}$ y $\boldsymbol{R}$ se pueden construir mediante el proceso de **ortogonalización de Gram-Schmidt**:

Para $j = 1, \ldots, n$:

- para $i = 1, \ldots, j-1$:
  - $$r_{ij} = \boldsymbol{q}_i^{\mathrm{t}} \boldsymbol{a}_j,$$
- $$r_{jj} = \left\| \boldsymbol{a}_j - \sum_{i=1}^{j-1} r_{ij}\boldsymbol{q}_i \right\|_2,$$
- $$\boldsymbol{q}_j = \frac{1}{r_{jj}} \left(\boldsymbol{a}_j - \sum_{i=1}^{j-1} r_{ij}\boldsymbol{q}_i \right).$$

Las columnas de $\boldsymbol{Q}$ son vectores ortonormales:

$$\boldsymbol{q}_i^{\mathrm{t}}\boldsymbol{q}_j = \delta_{ij} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j, \end{cases}$$

de donde la matriz $\boldsymbol{Q}$ satisface

$$\boldsymbol{Q}^{\mathrm{t}}\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{q}_1^{\mathrm{t}} \\ \hline \boldsymbol{q}_2^{\mathrm{t}} \\ \hline \vdots \\ \hline \boldsymbol{q}_n^{\mathrm{t}} \end{pmatrix} \left(\begin{array}{c|c|c|c} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \dots & \boldsymbol{q}_n \end{array}\right) = \boldsymbol{I}.$$

Si rango$(\boldsymbol{A}) = n$, entonces el proceso de ortogonalización de Gram-Schmidt genera una matriz $\boldsymbol{R}$ no singular.

Aplicación a la resolución de las ecuaciones normales.

Para resolver el sistema de ecuaciones normales:

$$\boldsymbol{A}^{\mathrm{t}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{t}}\boldsymbol{b},$$

como $\begin{cases} \boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}, \\ \boldsymbol{Q}^{\mathrm{t}}\boldsymbol{Q} = \boldsymbol{I} \quad \text{y} \\ \boldsymbol{R} \text{ es no singular,} \end{cases}$

entonces:

$$\begin{aligned} \boldsymbol{A}^{\mathrm{t}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{t}}\boldsymbol{b} \quad &\Longleftrightarrow \quad \boldsymbol{R}^{\mathrm{t}}\boldsymbol{Q}^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{R}^{\mathrm{t}}\boldsymbol{Q}^{\mathrm{t}}\boldsymbol{b} \\ &\Longleftrightarrow \quad \boldsymbol{R}^{\mathrm{t}}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{R}^{\mathrm{t}}\boldsymbol{Q}^{\mathrm{t}}\boldsymbol{b} \\ &\Longleftrightarrow \quad \boldsymbol{R}\boldsymbol{x} = \boldsymbol{Q}^{\mathrm{t}}\boldsymbol{b}. \end{aligned}$$

Cuando $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ es una matriz rectangular y $\boldsymbol{b} \in \mathbb{R}^{m}$, el comando MATLAB

```
x=A\b;
```

devuelve la solución del sistema rectangular en el sentido de mínimos cuadrados.

MATLAB obtiene esta solución mediante el método **QR**, pero las matrices $\boldsymbol{Q}$ y $\boldsymbol{R}$ no se obtienen mediante Gram-Schmidt, sino mediante **transformaciones de Householder** que propagan menos los errores de redondeo.

Ejemplo 1. Un problema de aproximación polinomial.

Considere la siguiente tabla de valores:

x	y
0.0	10.5000
0.5	5.4844
1.0	0.0000
1.5	-3.6094
2.0	-4.5000
2.5	-2.9531
3.0	0.0000
3.5	2.9531
4.0	4.5000
4.5	3.6094
5.0	0.0000

Se pide ajustar estos datos en el sentido de mínimos cuadrados por un polinomio de grado 3.

Solución.

Nuestro problema se reduce a encontrar constantes a, b, c y d para formar el polinomio

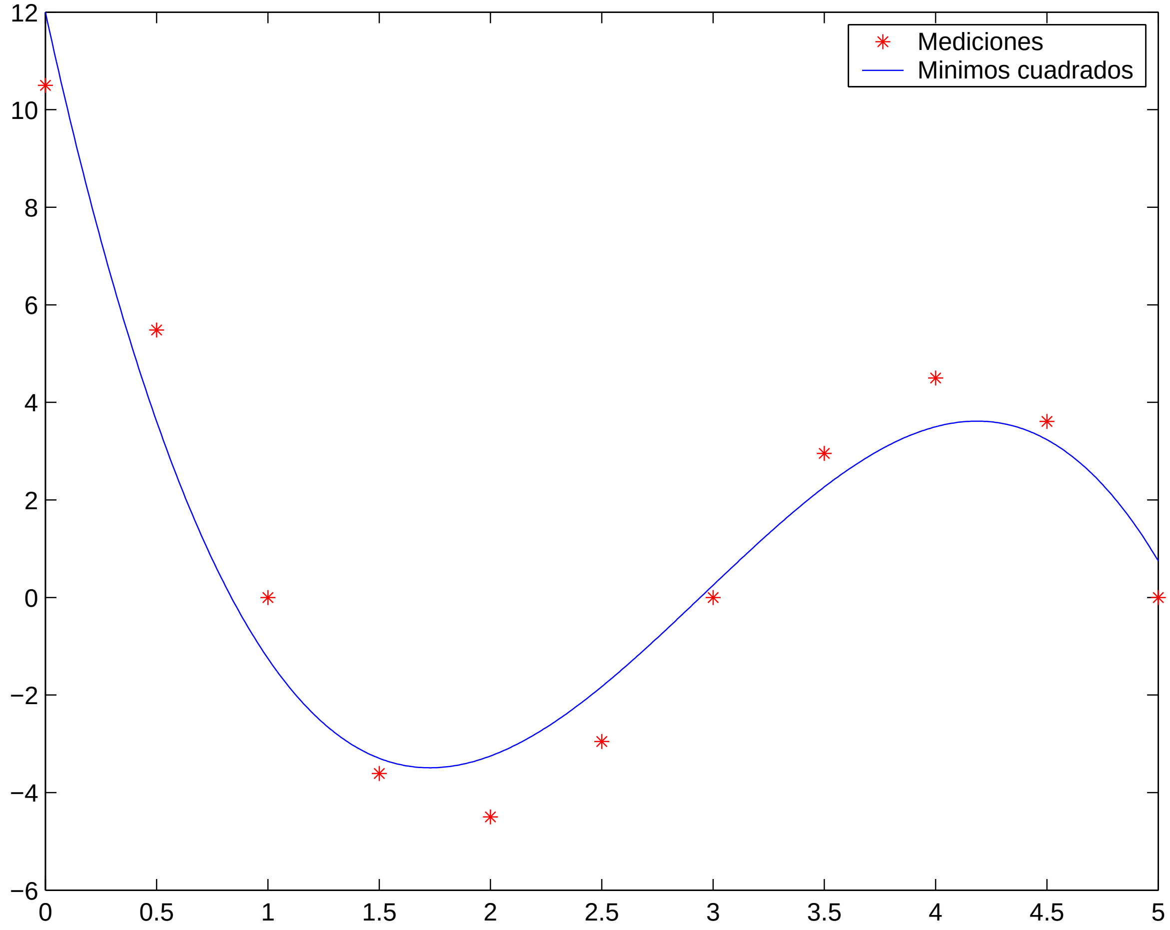
$$p(x) = ax^3 + bx^2 + cx + d.$$

Evaluamos el polinomio p en los diferentes valores x de la tabla, obteniendo así el sistema lineal rectangular con incógnitas a, b, c y d:

$$
\underbrace{\begin{pmatrix}
0.0000 & 0.0000 & 0.0000 & 1.0000 \\
0.1250 & 0.2500 & 0.5000 & 1.0000 \\
1.0000 & 1.0000 & 1.0000 & 1.0000 \\
3.3750 & 2.2500 & 1.5000 & 1.0000 \\
8.0000 & 4.0000 & 2.0000 & 1.0000 \\
15.6250 & 6.2500 & 2.5000 & 1.0000 \\
27.0000 & 9.0000 & 3.0000 & 1.0000 \\
42.8750 & 12.2500 & 3.5000 & 1.0000 \\
64.0000 & 16.0000 & 4.0000 & 1.0000 \\
91.1250 & 20.2500 & 4.5000 & 1.0000 \\
125.0000 & 25.0000 & 5.0000 & 1.0000
\end{pmatrix}}_{\boldsymbol{A}}
\underbrace{\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}}_{\boldsymbol{x}}
=
\underbrace{\begin{pmatrix}
10.5000 \\ 5.4844 \\ 0.0000 \\ -3.6094 \\ -4.5000 \\ -2.9531 \\ 0.0000 \\ 2.9531 \\ 4.5000 \\ 3.6094 \\ 0.0000
\end{pmatrix}}_{\boldsymbol{b}}
$$

Resolviendo el sistema rectangular en el sentido de mínimos cuadrados se obtiene el polinomio:

$$p(x) = -0.9583x^3 + 8.5x^2 - 20.7917x + 12.$$



Ejemplo 2. Un problema no lineal reducible a lineal.

Considere la siguiente tabla de valores

x	0.0000	0.4000	0.8000	1.2000	1.6000	2.0000
y	3.1437	4.4169	6.0203	8.6512	11.0078	16.2161

Se quiere ajustar una función

$$f(x) = ae^{bx}$$

a estos datos en el sentido de mínimos cuadrados.

Solución.

Tomando logaritmos se transforma en un problema lineal de cuadrados mínimos:

$$z = \ln(y) = \ln(f(x)) = \ln(a) + bx.$$

x	0.0000	0.4000	0.8000	1.2000	1.6000	2.0000
$z = \ln(y)$	1.1454	1.4854	1.7951	2.1577	2.3986	2.7860

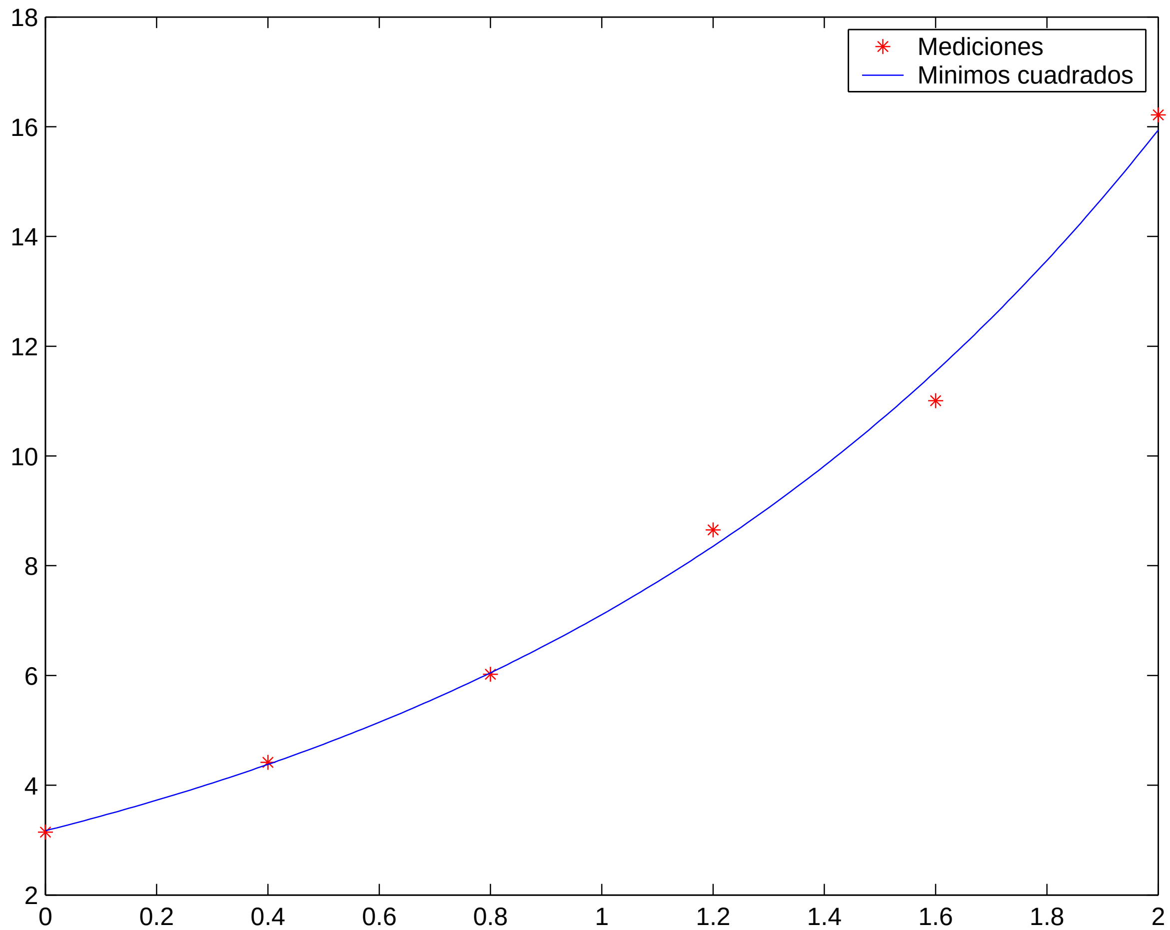
$$\underbrace{\begin{pmatrix} 1 & 0.0000 \\ 1 & 0.4000 \\ 1 & 0.8000 \\ 1 & 1.2000 \\ 1 & 1.6000 \\ 1 & 2.0000 \end{pmatrix}}_{\boldsymbol{A}} \underbrace{\begin{pmatrix} \ln(a) \\ b \end{pmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{pmatrix} 1.1454 \\ 1.4854 \\ 1.7951 \\ 2.1577 \\ 2.3986 \\ 2.7860 \end{pmatrix}}_{\boldsymbol{b}}$$

Resolvemos el problema de cuadrados mínimos y obtenemos:

$$\ln(a) = 1.1539 \quad \Longrightarrow \quad a = e^{1.1539} = 3.1705 \qquad \text{y} \qquad b = 0.8075.$$

Por lo tanto

$$f(x) = 3.1705e^{0.8075x}.$$



Otros ejemplos de modelos no lineales reducibles a lineales

- $f(t) = ce^{at-bt^2}$: en este caso, se aplica logaritmo.

- $f(t) = \dfrac{a}{b+t}$: en este caso se toman los recíprocos.

- $f(t) = \dfrac{k_0}{1+ae^{-ct}}$, donde k_0 es una constante conocida: en este caso se toman recíprocos, se resta 1 y después se aplica logaritmo.