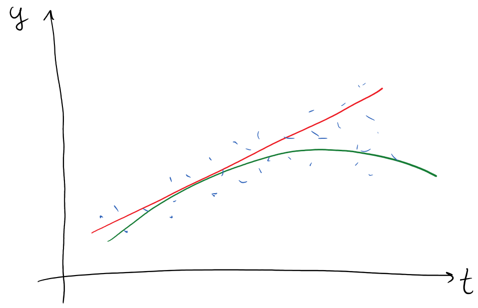
Repaso.



$\{(t_i, y_i)\}_{i=1}^{n}$ → muy grande

⊗ Espacio de funciones que modelan los datos

⊗ Una noción de distancia.

Ejemplos: 1) Polinomios

$$f(t) = \sum_{j=0}^{p} a_j t^j \rightarrow \text{polinomios de grado } p.$$

→ $p+1$ parámetros.

$$f(t) = \sum_{j=0}^{p} a_j \underbrace{\phi_j(t)}$$

↳ forman una base del espacio

2) Exponenciales

$$f(t) = \sum_{j=1}^{p} a_j e^{-b_j t}$$

$2p$ parámetros → p lineales, p no lineales.

$$f(t) = \sum_{j=1}^{p} a_j \phi_{b_j}(t)$$

3) Racionales:

$$f(t) = \sum_{j=1}^{p} \frac{a_j t^j}{\sum_{j=1}^{r} t^r b_r} \rightsquigarrow p+r \text{ parámetros}$$

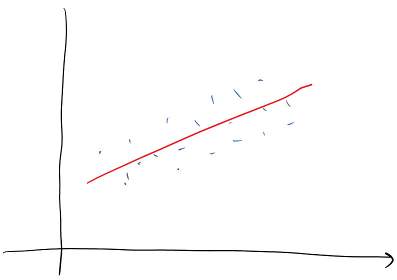
↙ lineales ↘ no lineales.

4) Gaussianas:

$$f(t) = \sum_{j=1}^{p} a_j e^{-\frac{(x-\mu_j)^2}{\sigma_j^2}} \quad 3p \rightarrow \text{parámetros.}$$

Idea: $f(t_i) = \bar{y}_i \neq y_i$

Sistema en general incompatible $\boxed{f(t_i) = y_i}$



Ejemplo si queremos ajustar por funciones lineales

$$f(t) = a\,1 + b\,t$$

$1 \downarrow \phi_1(t)$, $t \downarrow \phi_2(t)$

$$\begin{cases} a\,\phi_1(t_1) + b\,\phi_2(t_1) = y_1 \\ a\,\phi_1(t_2) + b\,\phi_2(t_2) = y_2 \end{cases}$$

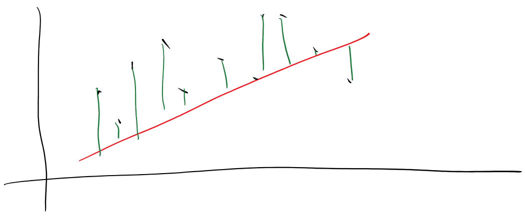
$$\begin{cases} \cdots \\ a\,\phi_1(t_2) + b\,\phi_2(t_2) = y_2 \\ \cdots\cdots \\ a\,\phi_1(t_n) + b\,\phi_2(t_n) = y_n \end{cases}$$

$$\begin{pmatrix} \phi_1(t_1) & \phi_2(t_1) \\ \phi_1(t_2) & \phi_2(t_2) \\ \cdots & \cdots \\ \phi_1(t_n) & \phi_2(t_n) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Descomposición QR.

$A\,x = y$

$x \to$ parámetros del espacio de funciones



$$\Phi(x) = \sum_{i=1}^{n} \Big(\underbrace{f(t_i)}_{\shortparallel} - y_i \Big)$$

$$\Big(\underbrace{A_i}_{\text{fila } i}(x) - y_i \Big)$$

$$f = \sum_{j=1}^{m} a_j \phi_j$$

$$\Phi(x) = \sum_{i=1}^{n} \Big(\sum_{j=1}^{m} a_j\,\phi_j(t_i) - y_i \Big)$$

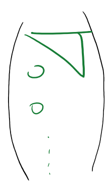
Minimizando la función distancia, encontramos $\boxed{A\,x = y}$

$x = (a_1, \dots, a_m)$

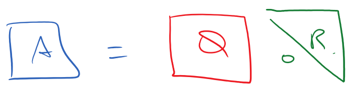
Descomposición QR.

$\underbrace{A}_{n\times m} = \underbrace{Q}_{n\times n} \cdot \underbrace{R}_{n\times m} \to$ matriz triangular superior

Q es matriz ortogonal (las columnas forma una bon. de $\mathbb{R}^n$)

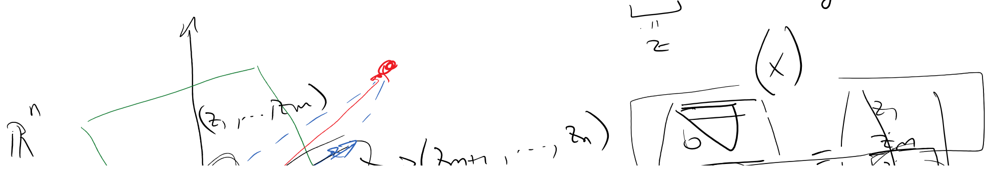
$R =$ 

Si $n=m$,

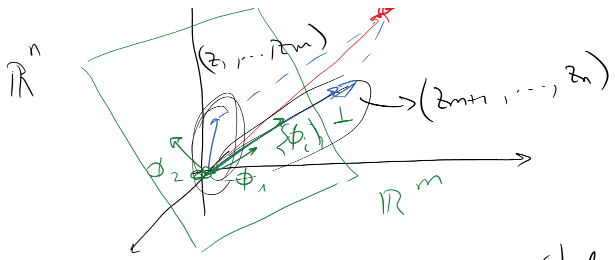


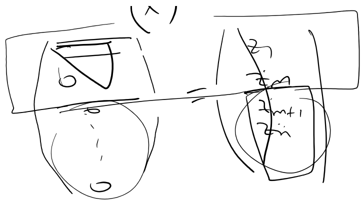
$A\,x = y \implies Q\,R\,x = y$

$\underbrace{R\,x}_{z} = Q^{-1} y = \boxed{Q^t y}$



$\mathbb{R}^n$ $(z_1, \dots, z_m)$ $(z_{m+1}, \dots, z_n)$ (x) ... ...posible !!





→ Incompatible !!

módulo $(z_{m+1}, \dots, z_n)$

$$\boxed{\underbrace{R(1:m,\ 1:m)}_{\substack{m \times m \\ \text{det no nulo.}}}\ x = z(1:m)}$$

Ejercicio 1 (Un ajuste polinomial). Se dispone de un conjunto de observaciones $\{(x_i, y_i) : i = 1, \ldots, n\}$, que se quieren aproximar por una función $f(x) = ax^3 + bx + c$.

a) Deducir las ecuaciones normales para este modelo.

b) Escribir un algoritmo que utilice la descomposición QR[1] para resolver el problema anterior y aplicarlo a los datos $\{(-1, 7/2), (0, 3/2), (1, 2), (2, 11/5), (3, 3)\}$.

a) Un espacio de funciones de dimensión $\underline{3}$

$$f(t) = a\underbrace{t^3}_{\phi_3(t)} + b\underbrace{t}_{\phi_2(t)} + c\cdot\underbrace{1}_{\phi_1(t)}$$

$$A = \begin{pmatrix} \phi_1(t_1) & \cdots & \phi_3(t_1) \\ - & - & - \\ \phi_1(t_n) & \cdots & \phi_3(t_n) \end{pmatrix} = \begin{pmatrix} 1 & t & t_1^3 \\ 1 & t_2 & t_2^3 \\ & \cdots & \\ 1 & t_n & t_n^3 \end{pmatrix}$$

$$\underbrace{A}_{n\times 3}\underbrace{\begin{pmatrix} c \\ b \\ a \end{pmatrix}}_{3\times 1} = \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{n\times 1}$$

Ejercicio 2 (Masas atómicas). El nitrógeno y el oxígeno tienen pesos atómicos (expresados en la unidad llamada de *masas atómicas unificadas*) $N \approx 14$ y $O \approx 16$. Se quieren usar los pesos moleculares de los seis óxidos de nitrógeno que se muestran abajo para determinar los pesos atómicos del nitrógeno y el oxígeno con cuatro dígitos decimales:

$$NO = 30{,}006, \qquad N_2O = 44{,}013, \qquad NO_2 = 46{,}006,$$
$$N_2O_3 = 76{,}012, \qquad N_2O_5 = 108{,}010, \qquad N_2O_4 = 92{,}011.$$

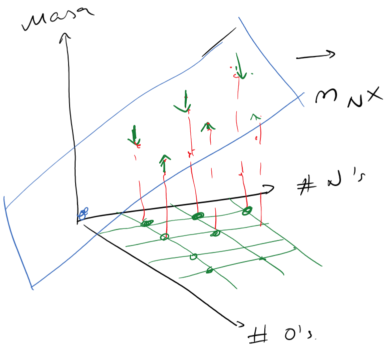
Determinar el sistema sobredeterminado de ecuaciones y computar la solución de mínimos cuadrados.

$$f(x, y) = m_N x + m_O y \longrightarrow \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 2 \\ 2 & 3 \\ 2 & 5 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} m_N \\ m_O \end{pmatrix} = \begin{pmatrix} 30{,}006 \\ 44{,}013 \\ \vdots \\ 92{,}011 \end{pmatrix}$$

x: # N's $\quad$ y: # O's

$$QRx = y \longrightarrow \underbrace{Rx}_{z} = Q^t y$$

$x = \begin{pmatrix} m_N \\ m_O \end{pmatrix}$

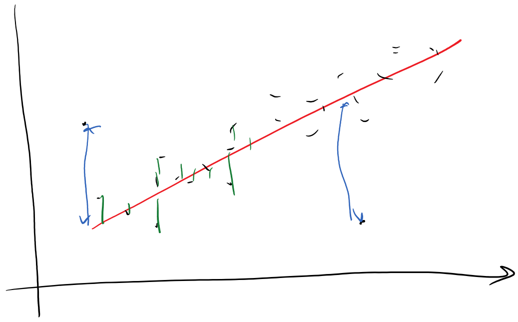


$\longrightarrow m_N x + m_O y = z \longrightarrow$

$$\boxed{R(1:2, 1:2)\, x = z(1:2)}$$

Ejercicio 3 (*Outliers*). Aquí tenemos 25 observaciones, $\{y_k\}_{k=1}^{25}$, tomados en valores de t equiespaciados:

```
t = (1:25)'
y = [ 5.0291  6.5099  5.3666  4.1272  4.2948
      6.1261  12.5140 10.0502 9.1614  7.5677
      7.2920  10.0357 11.0708 13.4045 12.8415
      11.9666 11.0765 11.7774 14.5701 17.0440
      17.0398 15.9069 15.4850 15.5112 17.6572]
y = y';
y = y(:);
```



a) Ajustar los datos con una función lineal $y(t) = \beta_1 + \beta_2 t$ y graficar las componentes del residuo $y(t_k) - y_k$. Al hacerlo, se debería observar que uno de los datos presenta un residuo mucho mayor que los otros. Ese sea seguramente un *valor atípico* (también llamado *outlier*).

b) Descartar el valor atípico y ajustar nuevamente los datos con una recta. Graficar los residuos nuevamente. ¿Se ve algún patrón en los residuos?

c) Ajustar los datos, excluyendo al valor atípico, con un modelo de la forma $y(t) = \beta_1 + \beta_2 t + \beta_3 \operatorname{sen} t$.

d) Evaluar el ajuste de la parte anterior en una grilla más fina sobre el intervalo $[0, 26]$. Graficar la curva ajustada, los datos, y marque al valor atípico usando un marcador diferente (por ejemplo, usando el estilo `'o'` para los datos y `'*'` para el valor atípico).