6. Modelos baseados em dados

Objectivo: Após completar este módulo, o aluno deverá ser capaz de formular e resolver problemas simples de estimação de parâmetros com base no método dos mínimos quadrados.

Referência: Ljung e Glad.

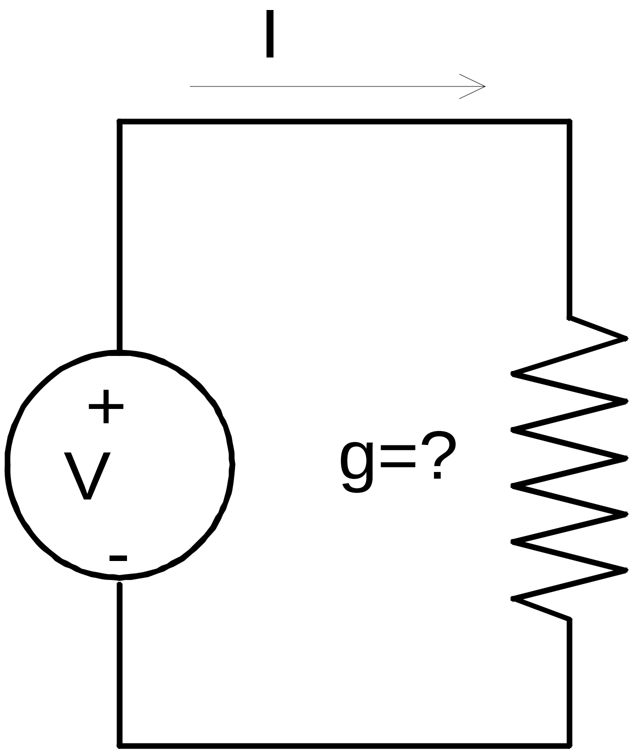
Ajuste de uma recta a dados experimentais

Uma situação experimental:

Pretende-se relacionar a corrente *I* com a tensão *V* no circuito da figura.

Para tal são aplicados diversos valores de tensão à resistência e registados os dados

Tensão [volt]	Corrente [mA]
$V_1=1$	$I_1=2.1$
$V_2=2$	$I_2=3.9$
$V_3=3$	$I_3=6.2$
$V_4=4$	$I_4=7.9$



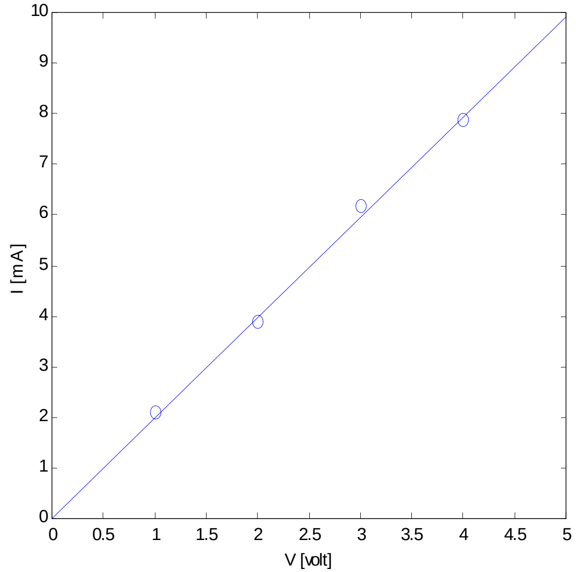
Sob certas condições, a relação teórica (o modelo) existente entre a tensão *V* aplicada à resistência e a corrente *I* é:

$$I = gV$$

em que g é um parâmetro que se pretende estimar a partir dos dados.

Devido aos erros experimentais, os pontos experimentais não se encontram exactamente sobre a recta $I = gV$ mas têm desvios.

Como decidir qual a recta melhor ajustada?



De acordo com o **Princípio dos Mínimos Quadrados** é escolhida a recta que minimiza a soma dos quadrados dos desvios.

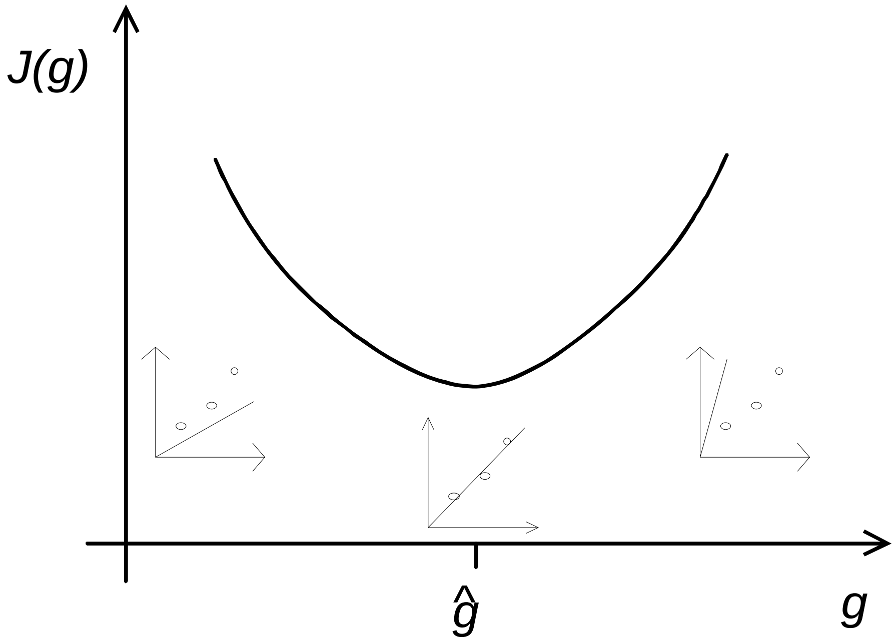
De acordo com este princípio, a estimativa de g é tal que minimiza

$$J(g) = \left(2.1 - g \times 1\right)^2 + \left(3.9 - g \times 2\right)^2 + \left(6.2 - g \times 3\right)^2 + \left(7.9 - g \times 4\right)^2$$

O que efectivamente observamos

O que esperamos que seja a corrente quando a tensão é *2* (Depende da estimativa de *g*)

- **"custo"** $J(g)$ **associado ao critério de mínimos quadrados**



Como $J(g)$ é uma função quadrática de g, a estimativa de mínimos quadrados verifica a equação

$$\frac{1}{2}\frac{dJ}{dg}\bigg|_{g=\hat{g}} = 0$$

ou seja

$$-1\times\left(2.1-\hat{g}\times 1\right)-2\times\left(3.9-\hat{g}\times 2\right)-3\times\left(6.2-\hat{g}\times 3\right)-4\times\left(7.9-\hat{g}\times 4\right)=0$$

Esta equação simplifica-se para $60\hat{g}-120.1=0$ sendo a estimativa de mínimos quadrados dada por

$$\hat{g} = 2.00$$

Ajuste de uma recta a dados experimentais (Caso geral)

Suponhamos que a relação teórica entre duas grandezas X e Y é do tipo

$$Y = \alpha X$$

em que α é um parâmetro desconhecido que se pretende estimar.

Repare-se que, conhecendo uma estimativa de α podemos responder a perguntas do tipo "*Se* X *valer … quanto se espera que valha* Y *?* "

Suponhamos que são observados n pares $\left(X_i, Y_i\right), i = 1, \ldots, n$ correspondentes a outros tantos ensaios experimentais. Dispõe-se da tabela

X	Y
X_1	Y_1
X_2	Y_2
X_3	Y_3
X_4	Y_4
X_5	Y_5

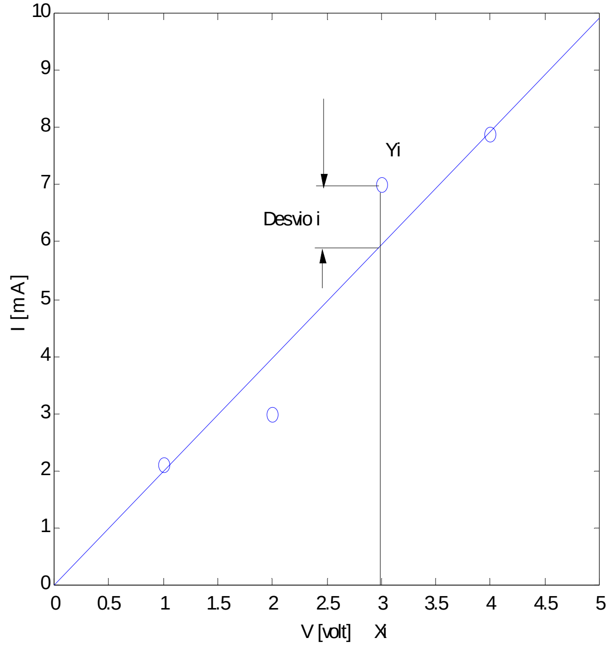
Pretende-se estimar a recta melhor ajustada aos dados experimentais, de acordo com o critério de mínimos quadrados.

De acordo com esta critério, a estimativa é tal que minimiza a soma dos quadrados dos desvios:

$$J(\alpha) = \sum_{i=1}^{n} \left(Y_i - \alpha X_i\right)^2$$

O que efectivamente observamos

O que estamos à espera que seja Y_i



A estimativa de mínimos quadrados verifica a equação (eq. "normal"):

$$\frac{1}{2}\frac{dJ}{d\alpha}\bigg|_{\alpha=\hat{\alpha}} = 0$$

ou seja

$$\sum_{i=1}^{n} X_i\left(Y_i - \hat{\alpha}X_i\right) = 0$$

esta equação tem por solução

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

Mínimo ou não?

A condição $\left.\frac{1}{2}\frac{dJ}{d\alpha}\right|_{\alpha=\hat{\alpha}}=0$ não garante necessariamente que $J(\alpha)$ seja mínimo para $\alpha=\hat{\alpha}$. É necessário impor uma condição na segunda derivada:

$$\frac{d^2J}{d\alpha^2}=\sum_{i=1}^{n}X_i^2>0$$

Neste exemplo, esta condição é verificada se for feita pelo menos uma medida com $X \neq 0$ (o que tem uma interpretação geométrica imediata).

Veremos a seguir que se estimarmos mais do que um parâmetro a segunda derivada deixa de ser um escalar. A condição de mínimo é então a de que os dados sejam tais que a matriz de segundas derivadas seja definida positiva.

Bom, ou apenas óptimo?

A estimativa de mínimos quadrados é "óptima" no sentido em que minimiza um funcional de custo. No entanto, o funcional de custo pode não ser o mais adequado.

Como caricatura, pode dizer-se que os bons relógios são os que estão parados pois dão horas absolutamente certas duas vezes por dia.

Um outro exemplo é o de um caçador que vê dois pombos. Se disparar para o ponto que minimiza a distância média quadrática aos *dois* pombos…

Isto sugere que por vezes são necessários outros critérios.

Outros critérios de Estimação

Ops exemplos anterios sugerem a utilidade de utilizar critérios que ultrapassem as limitações dos Mínimos Quadrados. Um dos mais utilizados em Estimação é o critério de **Máxima Verosimilhança**.

No entanto, quando a motivação é o Controlo Adaptativo, os Mínimos Quadrados gozam (quando integrados num sistema de controlo em cadeia fechada) de propriedades que os tornam suficientes para muitas aplicações. São além disso simples e de convergência robusta.

Carl Frederich Gauss (1777-1855) utilizou pela primeira vez o critério de mínimos quadrados para a estimação de parâmetros em equações.

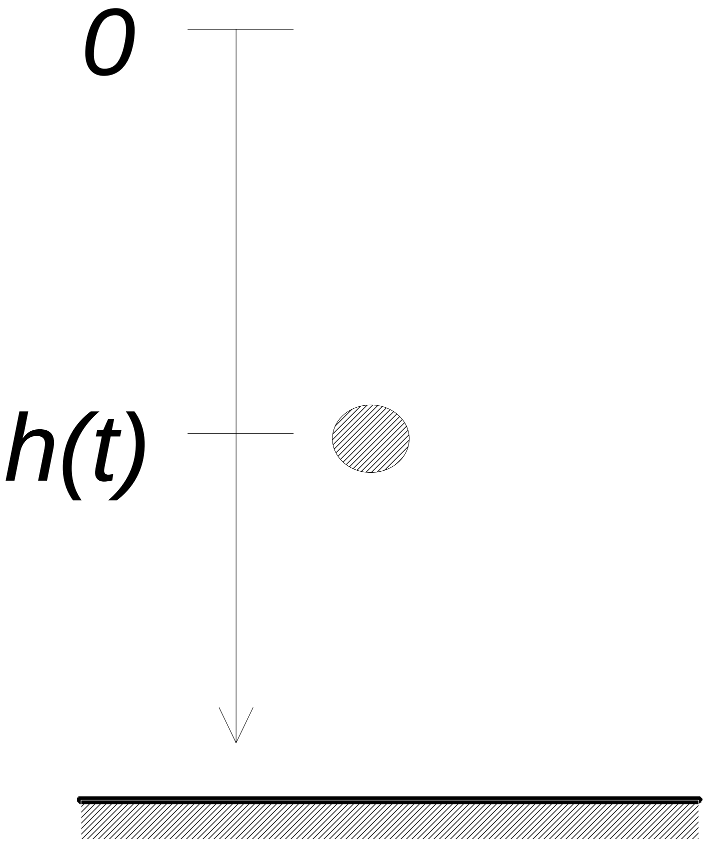
Em 1801, o astronomo italiano Piazzi observou pela primeira vez um pequeno planeta denominado Ceres. Infelizmente, a duração das observações era muito curta devido a Ceres se ser escondido atrás do Sol, pelo que estas eram insuficientes para estimar os parâmetros da sua órbita pelos métodos tradicionais. Recorrendo ao critértio dos mínimos quadrados, Gauss efectuou uma estimativa (bastante diferente das obtidas pelos métodos clássicos) que foi brilhantemente confirmada pelas observações experimentais.

Qual a estimativa de mínimos quadrados

da aceleração da gravidade g?

Modelo: $h = g\dfrac{t^2}{2} + e$

$t\ [s]$	$h\ [m]$
1	8.49
2	20.05
3	50.65
4	72.19
5	129.85
6	171.56

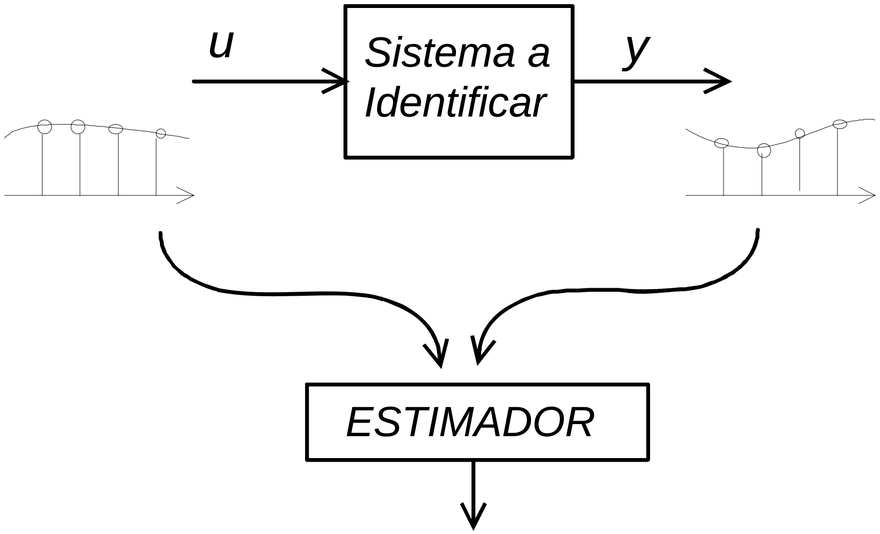


Estimação de parâmetros em equações de diferenças

Modelo:

$$y(t) + a_1 y(t-1) + \ldots + a_n y(t-n) = b_0 u(t-1) + \ldots + b_m u(t-m-1)$$

Problema: A partir das amostras de u e y, estimar os parâmetros a_i, b_j



Exemplo

Considere-se o sistema

$$y(k+1) = ay(k) + bu(k) + e(k+1)$$

Dados recolhidos

$$\sum_{k=1}^{1000} y^2(k) = 30 \qquad \sum_{k=1}^{1000} u^2(k) = 50 \qquad \sum_{k=1}^{1000} y(k+1)y(k) = 1$$

$$\sum_{k=1}^{1000} y(k+1)u(k) = 36 \qquad \sum_{k=1}^{1000} y(k)u(k) = 20$$

Determinar as estimativas de mínimos quadrados dos parâmetros a e b

- Escrever a funcional de mínimos quadrados
- Calcular as derivadas parciais em ordem aos parâmetros e igualar a zero

$$J = \frac{1}{2}\sum_{k=1}^{N}\left[y(k+1) - ay(k) - bu(k)\right]^2$$

$$\frac{\partial J}{\partial a} = -\sum_{k=1}^{N} y(k)\left[y(k+1) - ay(k) - bu(k)\right] = 0$$

$$\frac{\partial J}{\partial b} = -\sum_{k=1}^{N} u(k)\left[y(k+1) - ay(k) - bu(k)\right] = 0$$

$$\hat{a}\sum_{k=1}^{N} y^2(k) + \hat{b}\sum_{k=1}^{N} y(k)u(k) = \sum_{k=1}^{N} y(k)y(k+1) \qquad 30\hat{a} + 20\hat{b} = 1$$

$$\hat{a}\sum_{k=1}^{N} u(k)y(k) + \hat{b}\sum_{k=1}^{N} u^2(k) = \sum_{k=1}^{N} u(k)y(k+1) \qquad 20\hat{a} + 50\hat{b} = 36$$

$$\hat{a} = -0.61 \qquad \hat{b} = 0.964$$

Notação matricial

Se quisermos resolver o problema de estimação para um número arbitrário de parâmetros temos de usar a notação matricial.

$$y(t)+a_1 y(t-1)+\ldots+a_n y(t-n)=b_0 u(t-1)+\ldots+b_m u(t-1-m)+e(t)$$

Define-se o **regressor** , φ ,como

$$\varphi'(t-1)=\left[-y(t-1)\;\ldots\;-y(t-n)\;u(t-1)\;u(t-1-m)\right]$$

e o vector de parâmetros a estimar, θ,como

$$\theta'=\left[a_1\;\ldots\;a_n\;b_1\;\ldots\;b_m\right]$$

O modelo escreve-se:

$$y(t)=\varphi'(t-1)\theta+e(t)$$

Critério de Mínimos Quadrados

Dadas N observações, estimar o vector de parâmetros θ_o por um vector $\hat{\theta}$ por forma a que o funcional seguinte seja mínimo:

$$J(\theta) = \frac{1}{2N}\sum_{t=1}^{N}\left[y(t) - \theta'\varphi(t-1)\right]^2$$

$$\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} \varphi'(0)\theta \\ \varphi'(1)\theta \\ \vdots \\ \varphi'(N-1)\theta \end{bmatrix} + \begin{bmatrix} e(1) \\ e(2) \\ \vdots \\ e(N) \end{bmatrix}$$

$$\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} \varphi'(0) \\ \varphi'(1) \\ \vdots \\ \varphi'(N-1) \end{bmatrix} \theta + \begin{bmatrix} e(1) \\ e(2) \\ \vdots \\ e(N) \end{bmatrix}$$

$$\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} \varphi'(0) \\ \varphi'(1) \\ \vdots \\ \varphi'(N-1) \end{bmatrix} \theta + \begin{bmatrix} e(1) \\ e(2) \\ \vdots \\ e(N) \end{bmatrix}$$

$\bar{y}[N \times 1]$ $\Phi[N \times n_p]$ $[n_p \times 1]$ $\bar{\varepsilon}[N \times 1]$

O conjunto das N observações satisfaz:

$$\bar{y} = \Phi\theta + \bar{\varepsilon}$$

Funcional de mínimos quadrados escrito matricialmente:

$$J(\theta) = \frac{1}{2N}\left\|\bar{\varepsilon}\right\|^2 = \frac{1}{2N}\bar{\varepsilon}'\bar{\varepsilon}$$

Como:

$$\bar{\varepsilon} = \bar{y} - \Phi\hat{\theta}$$

Vem:

$$J(\theta) = \frac{1}{2N}\left(\bar{y}' - \theta'\Phi'\right)\left(\bar{y} - \Phi\theta\right)$$

$$J(\theta) = \frac{1}{2N}\left(\bar{y}'\bar{y} - 2\bar{y}'\Phi\theta + \theta'\Phi'\Phi\theta\right)$$

O funcional de mínimos quadrados é uma **forma quadrática** em θ

A estimativa $\hat{\theta}$ de mínimos quadrados satisfaz

$$\nabla_{\theta} J(\theta)\Big|_{\theta=\hat{\theta}} = 0$$

Gradiente da forma quadrática

$$\nabla_{x}\left(x'Ax\right) = 2x'A$$

Recordando:

$$J(\theta) = \frac{1}{2N}\left(\bar{y}'\bar{y} - 2\bar{y}'\Phi\theta + \theta'\Phi'\Phi\theta\right)$$

vem

$$\nabla_{\theta} J = \frac{1}{2N}\left(-2\bar{y}'\Phi + 2\theta'\Phi'\Phi\right)$$

A estimativa de mínimos quadrados satisfaz pois a equação

$$\nabla_\theta J = \frac{1}{2N}\left(-2\bar{y}'\Phi + 2\theta'\Phi'\Phi\right)$$

ou seja

$$\hat{\theta}'\Phi'\Phi = \bar{y}'\Phi$$

ou, transpondo

$$\Phi'\Phi\hat{\theta} = \Phi'\bar{y}$$

Equação Normal

Em conclusão, a estimativa de mínimos quadrados do vector de parâmetros θ do modelo de regressão linear

$$y(t) = \varphi'(t-1)\theta + e(t)$$

satisfaz a equação matricial (dita *equação normal*)

$$\Phi'\Phi\hat{\theta} = \Phi'\overline{y}$$

Se existir a inversa de $\Phi'\Phi$ a estimativa de mínimos quadrados existe e é única, sendo dada por

$$\hat{\theta} = \left(\Phi'\Phi\right)^{-1}\Phi'\overline{y}$$