# 3 – Identificação Recursiva.

***Objectivo:*** *Introdução às técnicas de identificação Recursiva com Aplicação no Controlo Adaptativo: O Método dos Mínimos Quadrados. Limitações.*

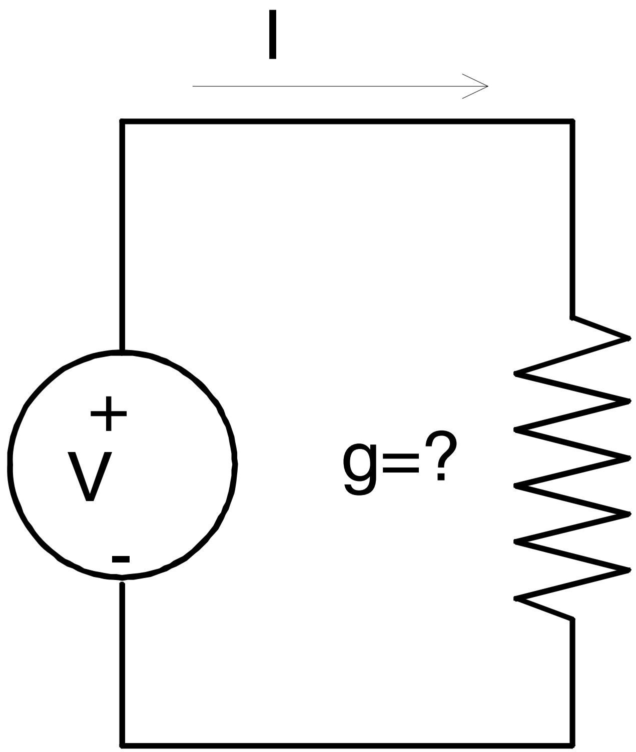
Referência: AW cap. 13

# 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] |
|---|---|
| $V_1=1$ | $I_1=2.1$ |
| $V_2=2$ | $I_2=3.9$ |
| $V_3=3$ | $I_3=6.2$ |
| $V_4=4$ | $I_4=7.9$ |

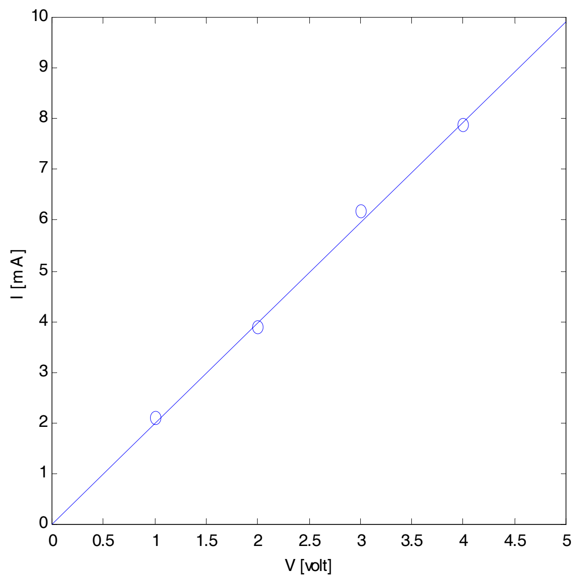
Sob certas condições, a relação teórica (o modelo) existente entre a tensão *V* aplicada à resistência e a corrente *I* é:

$$I = gV$$

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.

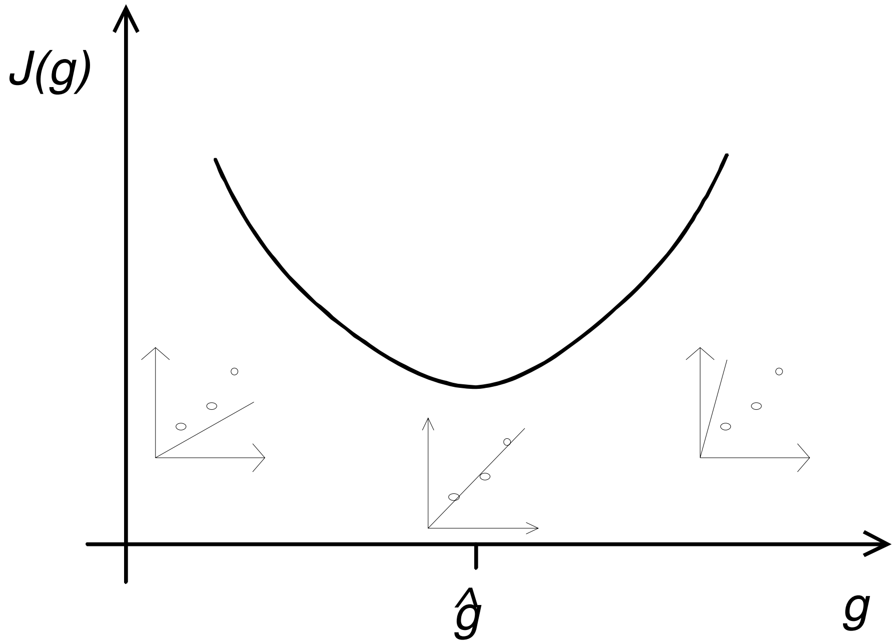
De acordo com este princípio, a estimativa de $g$ é tal que minimiza

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

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

$$\left.\frac{1}{2}\frac{dJ}{dg}\right|_{g=\hat{g}} = 0$$

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 $\alpha$ é um parâmetro desconhecido que se pretende estimar.

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

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

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

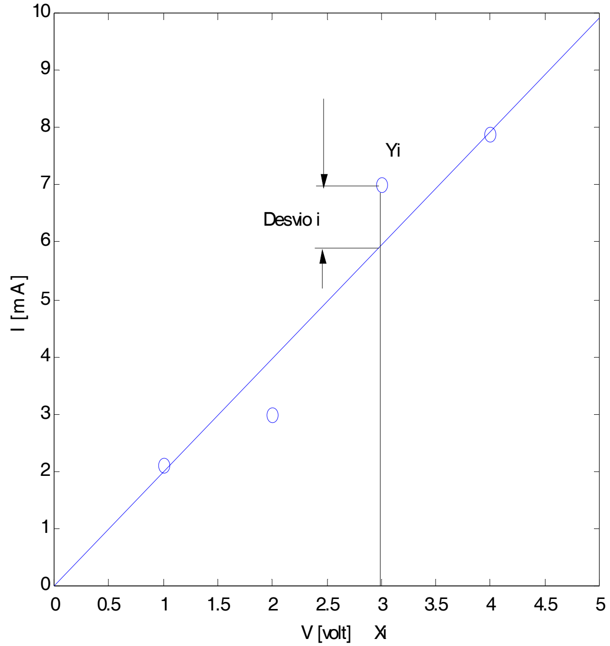
Pretende-se estimar a recta melhor ajustada aos dados experimentais, de acordo com o critério de mínimos quadrados.

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"):

$$\left.\frac{1}{2}\frac{dJ}{d\alpha}\right|_{\alpha=\hat{\alpha}} = 0$$

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

Os 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.

São assim o único algoritmo de estimação que será estudado nesta disciplina (o que, repete-se, não implica limitações em termos de controlo adaptativo).

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

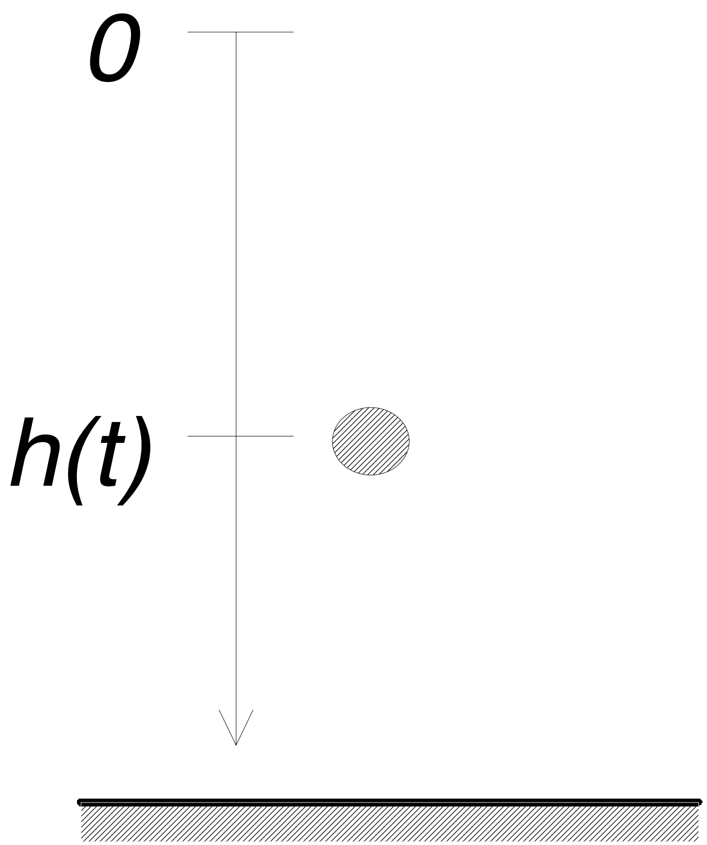
Em 1801, o astronomo italiano Piazzi observou pela primeira vez um pequeno planeta denominado Ceres. Infelizmente, a duração das observações era muito curta devido a Ceres se ser escondido atrás do Sol, pelo que estas eram insuficientes para estimar os parâmetros da sua órbita pelos métodos tradicionais. Recorrendo ao critértio dos mínimos quadrados, Gauss efectuou um a 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 |

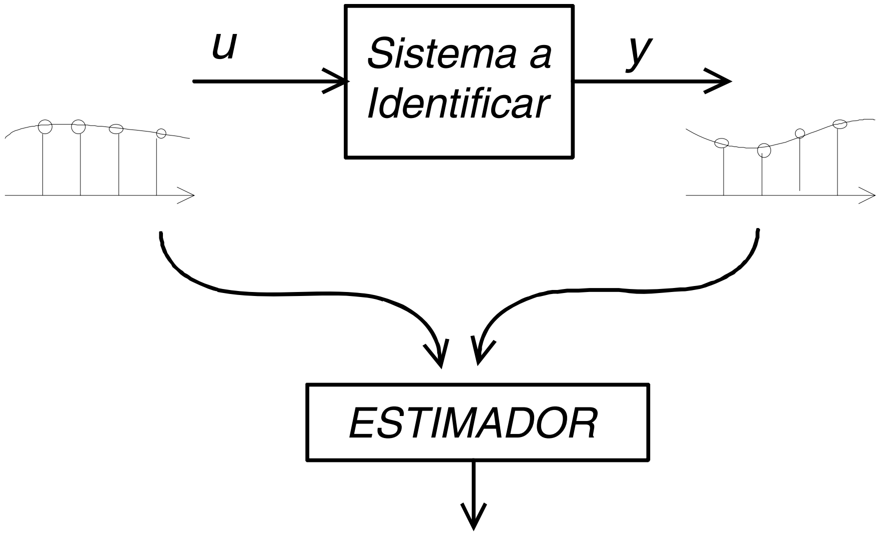


# Estimação de parâmetros em equações de diferenças

## 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

Ruído branco de média nula.

Traduz erros no modelo

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.

$$y(t)+a_1y(t-1)+\ldots+a_ny(t-n)=b_0u(t-1)+\ldots+b_mu(t-1-m)+e(t)$$

Define-se o **regressor** , $\varphi$ ,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, $\theta$ ,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 $\theta_o$ por um vector $\hat{\theta}$ por forma a que o funcional seguinte seja mínimo:

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

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

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

$$\underbrace{\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix}}_{\bar{y}[N \times 1]} = \underbrace{\begin{bmatrix} \varphi'(0) \\ \varphi'(1) \\ \vdots \\ \varphi'(N-1) \end{bmatrix}}_{\Phi[N \times n_p]} \underbrace{\theta}_{[n_p \times 1]} + \underbrace{\begin{bmatrix} e(1) \\ e(2) \\ \vdots \\ e(N) \end{bmatrix}}_{\bar{\varepsilon}[N \times 1]}$$

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 $\theta$

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 $\theta$ 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

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

## Exemplo - Invertibilidade da matriz $\Phi'\Phi$

Considere o sistema

Ruído de média nula

$$y(k) = b_0 u(k-1) + b_1 u(k-2) + e(k)$$

a) Diga se é possível determinar estimativas dos parâmetros $b_0$ e $b_1$ quando a entrada é sempre $u(k) = 1 \quad \forall_k$ ?

b) E se se souber que $b_1 = 0$ ?

c) E se $u(0) = 0$ e $u(k) = 1 \quad \forall_{k \geq 1}$?

Sugestão: *Escreva a matriz* $\Phi'\Phi$ *quando N=4 e depois considere o que acontece para valores de N superiores.*

# Condições de Excitação Persistente

Para que a estimativa de mínimos quadrados exista e seja única é necessário que a matriz $\Phi'\Phi$ seja definida positiva. Caso contrário o funcional de mínimos quadrados não tem um mínimo.

A matriz $\Phi'\Phi$ é definida positiva se os dados forem suficientemenbte ricos, o que depende da entrada do sistema.

As condições na entrada do sistema que levam a que $\Phi'\Phi > 0$ dizem-se condições de excitação persistente.

## Matriz de Covariância do Erro de Estimação e Matriz de Informação

Define-se a matriz de covariância do erro de estimação como

$$P = E\left[\left(\theta - \hat{\theta}\right)\left(\theta - \hat{\theta}\right)'\right]$$

No caso em que os resíduos formam uma sequência branca, tem-se:

*i)*A estimativa de mínimos quadrados é **centrada** (a média do erro é zero):

$$E\left[\hat{\theta} - \theta\right] = 0$$

*ii)*Sendo a variância dos resíduos unitária:

$$P = \left(\Phi'\Phi\right)^{-1}$$

Este último facto motiva que se denomine matriz de informação a matriz dada por

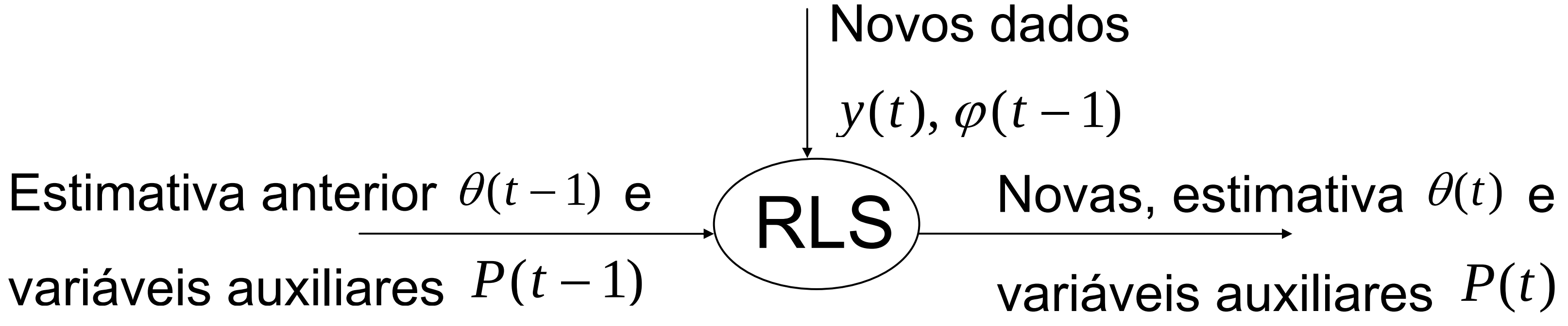
$$\Lambda = P^{-1} = \Phi'\Phi$$

Note-se que o facto de as estimativas de mínimos quadrados não serem centradas quando o ruído é colorido não é uma limitação em Controlo Adaptativo (embora implique o recurs o a outros métodos quando a motivação é outra, por exemplo identificação).

Isto deve-se ao facto de o Controlo Adapt ativo identificar o sistema em cadeia fechada. Posteriormente voltaremos a este aspecto.

## Estimador Recursivo

Objectivo: Obter a estimativa combinando uma estimativa anterior com novos dados sem ter que reescrever a equação normal.




Em inglês: *RLS = Recursive Least Squares*

Estimador não recursivo dadas $t$ observações:

$$\hat{\theta}(t) = \left(\Phi'\Phi\right)^{-1}\Phi'\overline{y}$$

Pode ser escrita como (mostre!):

$$\hat{\theta}(t) = \Lambda^{-1}(t)\sum_{k=1}^{t} y(k)\varphi(k-1)$$

em que a matriz de informação é dada por

$$\Lambda(t) = \sum_{k=1}^{t} \varphi(k-1)\varphi'(k-1)$$

Matriz de informação verifica

$$\Lambda(t) = \Lambda(t-1) + \varphi(t-1)\varphi'(t-1) \qquad (1)$$

Tem-se ainda

$$\hat{\theta}(t) = \Lambda^{-1}(t)\sum_{k=1}^{t} y(k)\varphi(k-1) \qquad (2) \qquad \text{e} \qquad \Lambda(t)\hat{\theta}(t) = \sum_{k=1}^{t} y(k)\varphi(k-1) \qquad (3)$$

Pretende-se: Escreva $\hat{\theta}(t)$ como função de $\hat{\theta}(t-1)$, de $\Lambda(t)$, de $y(t)$ e de $\varphi(t-1)$

Sugestão: Isole o último termo do somatório em (2).

Escreva (3) para $t-1$;

use (1);

$$\hat{\theta}(t) = \Lambda^{-1}(t)\sum_{k=1}^{t} y(k)\varphi(k-1)$$

$$\hat{\theta}(t) = \Lambda^{-1}(t)\left[\sum_{k=1}^{t-1} y(k)\varphi(k-1) + y(t)\varphi(t-1)\right]$$

$$\Lambda(t-1)\hat{\theta}(t-1) = \sum_{k=1}^{t-1} y(k)\varphi(k-1)$$

$$\hat{\theta}(t) = \Lambda^{-1}(t)\left[\Lambda(t-1)\hat{\theta}(t-1) + y(t)\varphi(t-1)\right]$$

$$\hat{\theta}(t) = \Lambda^{-1}(t)\left[\Lambda(t-1)\hat{\theta}(t-1) + y(t)\varphi(t-1)\right]$$

$$\Lambda(t-1) = \Lambda(t) - \varphi(t-1)\varphi'(t-1)$$

Assim:

$$\hat{\theta}(t) = \Lambda^{-1}(t)\left[\Lambda(t)\hat{\theta}(t-1) - \varphi(t-1)\varphi'(t-1)\hat{\theta}(t-1) + y(t)\varphi(t-1)\right]$$

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \Lambda^{-1}(t)\varphi(t-1)\left[y(t) - \varphi'(t-1)\hat{\theta}(t-1)\right]$$

As equações do estimador recursivo são:

É necessário inverter uma
matriz em cada iteração

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \Lambda^{-1}(t)\varphi(t-1)\left[y(t) - \varphi'(t-1)\hat{\theta}(t-1)\right]$$

Nova
estimativa

Estimativa
anterior

Ganho
vectorial

Diferença entre o que esperamos que seja y(t)
dada a estimativa e o qure observamos

$$\Lambda(t) = \Lambda(t-1) + \varphi(t-1)\varphi'(t-1)$$

## Será que podemos alterar o algoritmo para evitar inversão de matrizes?

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \Lambda^{-1}(t)\varphi(t-1)\left[y(t) - \varphi'(t-1)\hat{\theta}(t-1)\right]$$

$$\Lambda(t) = \Lambda(t-1) + \varphi(t-1)\varphi'(t-1)$$

Se trabalharmos com a matriz de covariância as equações ficam:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + P(t)\varphi(t-1)\left[y(t) - \varphi'(t-1)\hat{\theta}(t-1)\right]$$

$$P(t) = \left[\Lambda(t-1) + \varphi(t-1)\varphi'(t-1)\right]^{-1}$$

$$P(t) = \left[\Lambda(t-1) + \varphi(t-1)\varphi'(t-1)\right]^{-1}$$

## Lema de inversão de matrizes

$$\left[A + BCD\right]^{-1} = A^{-1} - A^{-1}B\left[DA^{-1}B + C^{-1}\right]^{-1} DA^{-1}$$

Aplique-se este lema com

$A^{-1} = \Lambda^{-1}(t-1) = P(t-1)$, $B = \varphi(t-1)$, $C = 1$, $D = \varphi'(t-1)$

Obtém-se:

$$P(t) = P(t-1) - P(t-1)\varphi(t-1)\left[\varphi'(t-1)P(t-1) + 1\right]^{-1} \varphi'(t-1)P(t-1)$$

Escalar

# Algoritmo de Mínimos Quadrados Recursivo (*RLS*)

Modelo:

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

Estimador:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\left[y(t) - \hat{\theta}(t-1)\varphi(t-1)\right]$$

"Ganho de Kalman"

"Equação de Riccati"

## Exercício

$$K(t) = P(t)\varphi(t-1)$$

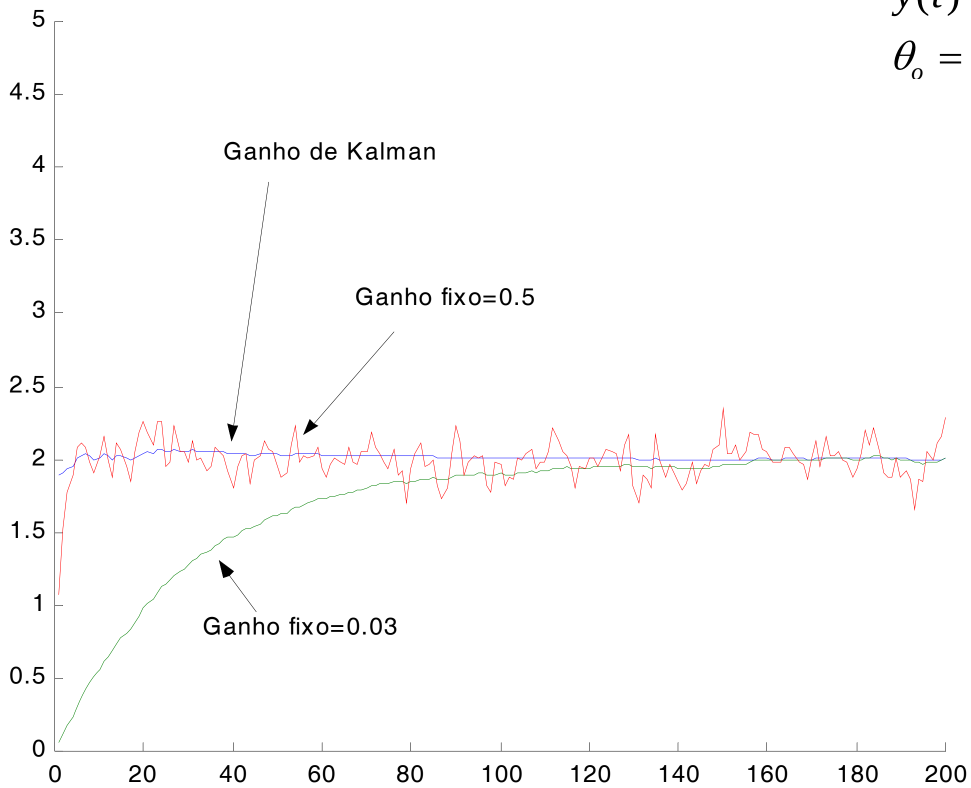
$$P(t) = P(t-1) - \frac{P(t-1)\varphi(t-1)\varphi'(t-1)P(t-1)}{1+\varphi'(t-1)P(t-1)\varphi(t-1)}$$

Mostre que o ganho de Kalman e a equação de Riccati podem ser escritos como:

$$K(t) = \frac{P(t-1)\varphi(t-1)}{1+\varphi'(t-1)P(t-1)\varphi(t-1)}$$

$$P(t) = \left[I - K(t)\varphi'(t-1)\right]P(t-1)$$

# Exemplo: Estimação recursiva de um parâmetro




$$y(t) = \varphi(t-1)\theta_o + e(t)$$

$$\theta_o = 2$$

Ganho do estimador:

**Alto** leva a convergência rápida mas a grande variância das estimativas em regime estacionário.

**Baixo** leva a convergência lenta.

O **ganho de Kalman** é alto no início e baixo no fim, variando no tempo.

# Programa MATLAB usado no exemplo

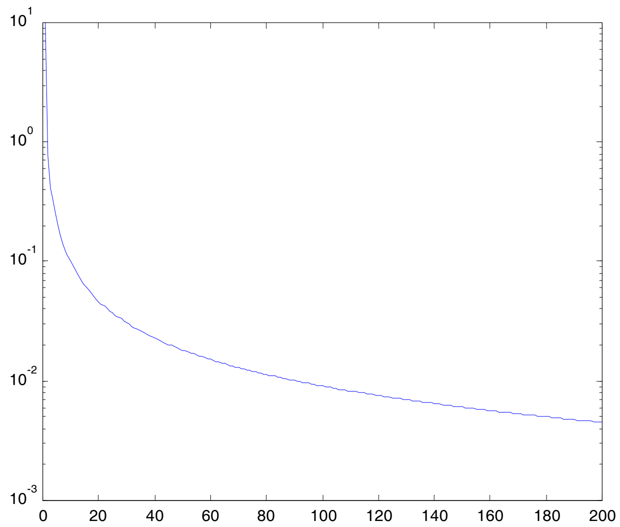
```
thrls=0;
thsm=0;
thbig=0;

p=10;
theta0=2;
tfinal=200;

for t=1:tfinal
   pp(t)=p;

  fi=1+0.1*rand;
  y=theta0*fi+0.2*randn;

  p=p-p*fi*fi*p/(1+fi*p*fi);

  kalm=p*fi;
  thrls=thrls+kalm*(y-thrls*fi);

  thsm=thsm+0.03*(y-thsm*fi);

  thbig=thbig+0.5*(y-thbig*fi);

  sth(t,1)=thrls;
  sth(t,2)=thsm;
  sth(t,3)=thbig;
end;

axis([0 tfinal 0 5])
hold on
plot(sth)
hold off
```

Este exemplo ilustra a convergência dos mínimos quadrados recursivos.

Inicialmente, como a nossa incerteza sobre o valor verdadeiro do parâmetro a estimar é grande, escolhemos um valor elevado para a covariância.

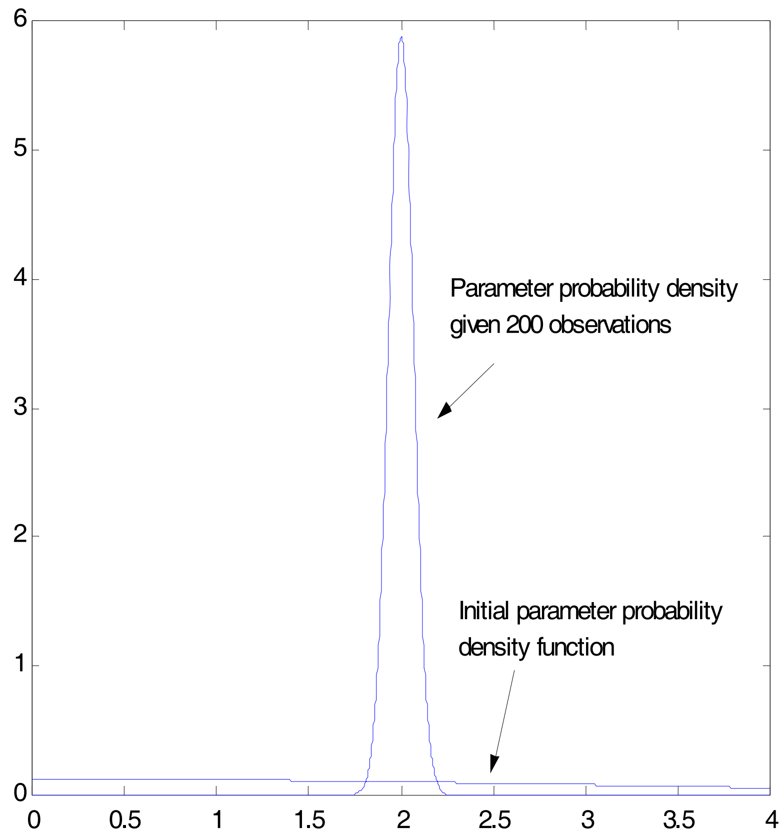
Neste caso, $P = 10$. O ganho de Kalman é elevado inicialmente pelo que a convergência é rápida.

À medida que o tempo passa e vamos recebendo dados, $P$ diminui e o ganho de Kalman também. Isto leva à convergência da estimativa pois o termo de correcção fica progressivamentre menor.



O estimador pode ser encarado como um mecanismo que reduz a nossa incerteza sobre o valor verdadeiro do parâmetro através das observações.

A incerteza é medida por uma função densidade de probabilidade do erro na estimativa. Neste exemplo esta incerteza é gaussiana e com variância proporcional a $P$.

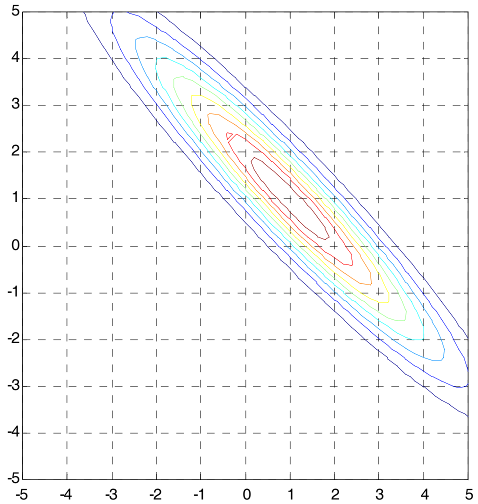


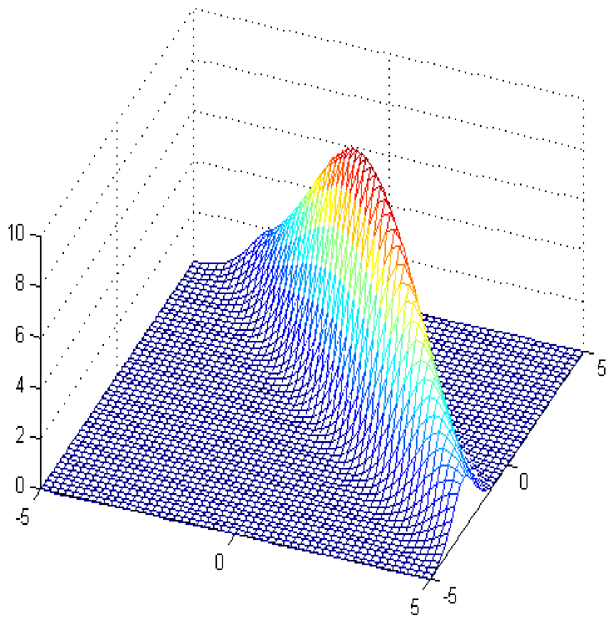
# Um exemplo com parâmetros não identificáveis

$$y(k) = b_0 u(k-1) + b_1 u(k-2) + e(k)$$

Apenas a soma pode ser estimada

Densidade de probabilidade dos parâmetros





Aumentando o número de observações, r eduz-se a incerteza da soma mas mantém-se a incerteza na direcção perpendicular (assume-se $u$ constante).

# Um exemplo de não identificabilidade de parâmetros em cadeia fechada

Modelo do processo:

$$y(t) = ay(t-1) + bu(t-1) + e(t) \qquad (*)$$

A este sistema está acoplado o controlador:

$$u(t) = -Ky(t)$$

Seja $\beta$ uma constante arbitrária. Da equação do controlador

$$\beta u(t) + \beta K y(t) = 0$$

Como a quantidade $\beta u(t) + \beta K y(t) = 0$, podemos adicioná-la a (*) e obter:

$$y(t) = (a + \beta K) y(t-1) + (b + \beta) u(t-1) + e(t)$$

Assim, o sistema, em conjunto com o controlador, é descrito pelo modelo

$$y(t) = (a + \beta K) y(t-1) + (b + \beta) u(t-1) + e(t)$$

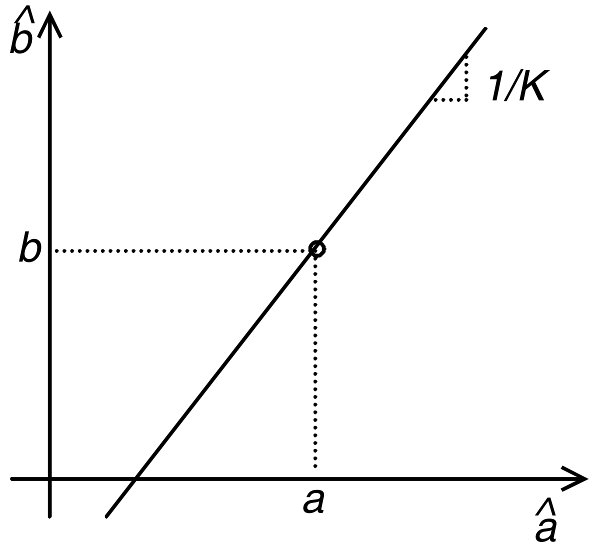
isto mostra que parâmetros tais que

$$\hat{a} = a + \beta K \qquad \hat{b} = b + \beta$$

conduzem à mesma relação entrada/saída. Eliminando $\beta$, obtém-se a seguinte relação entre as estimativas e os parâmetros a estimar

$$\hat{b} = b + \frac{1}{K}\left(\hat{a} - a\right)$$

Qualquer estimativa $(\hat{a}, \hat{b})$ que verifique esta condição descreve igualmente bem o comportamento entrada/saída.




As estimativas pertencem à recta, mas não são necessariamente proxímas dos valores verdadeiros. Para que o sejam, pode:

- Fazer variar-se o ganho *K*
- Adicionar um sinal externo ao controlo

## Adormecimento dos mínimos quadrados recursivos

Se os dados forem adequados, os elementos do ganho de Kalman diminuem à medida que o tempo passa, tornando-se eventualmente muito reduzidos se o ruído for baixo.

A partir daí, as estimativas tendem a tornar-se constantes. Se o sistema a identificar sofrer alguma alteração, será necessário muito tempo para que as estimativas convirjam para o novo valor.

Diz-se que o algoritmo *adormeceu*.

# Esquecimento exponencial

O *adormecimento* dos mínimos quadrados recursivos acontece porque o algoritmo pesa igualmente os dados recentes e os do passado remoto.

Para o evitar, pode modificar-se o funcional de custo por forma a pesar menos os pontos do passado:

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

Peso exponencial, menor nos dados mais antigos

A $\lambda$ dá-se o nome de *factor de esquecimento*. Tem-se $0 < \lambda < 1$

# Memória assimptótica

A memória assimptótica $N$ dá-nos uma ideia do número de dados que influenciam a estimativa actual.

$$N = \frac{1}{1-\lambda}$$

| $\lambda$ | $N$ |
|---|---|
| 1 | $\infty$ |
| 0.99 100 | |
| 0.98 50 | |
| 0.95 20 | |

- $\lambda$ pequeno -> Poucos dados retidos; algoritmo "ágil" a seguir alterações
- $\lambda$ grande-> Muitos dados retidos; algoritmo lento a seguir alterações mas mais preciso

# RLS com Esquecimento Exponencial

Minimiza o custo com esquecimento exponencial.

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\left[y(t) - \hat{\theta}(t-1)\varphi(t-1)\right]$$

$$K(t) = \frac{P(t-1)\varphi(t-1)}{\lambda + \varphi'(t-1)P(t-1)\varphi(t-1)}$$

$$P(t) = \left[I - K(t)\varphi'(t-1)\right]P(t-1) / \lambda$$

Demonstre que estas equações minimizam o custo exponencial.

Observe que a matriz de informação é actualizada por

$$\Lambda(t) = \lambda\Lambda(t-1) + \varphi(t-1)\varphi'(t-1)$$

## **Explosão da covariância** (*covariance windup*)

Pretende-se estimar os parâmetros $a$ e $b$ no modelo:

$$y(t+1) = ay(t) + bu(t) + 0.1e(t+1)$$

O valor verdadeiro dos parâmetros é

$$a = 0.6 \quad b = 0.1$$

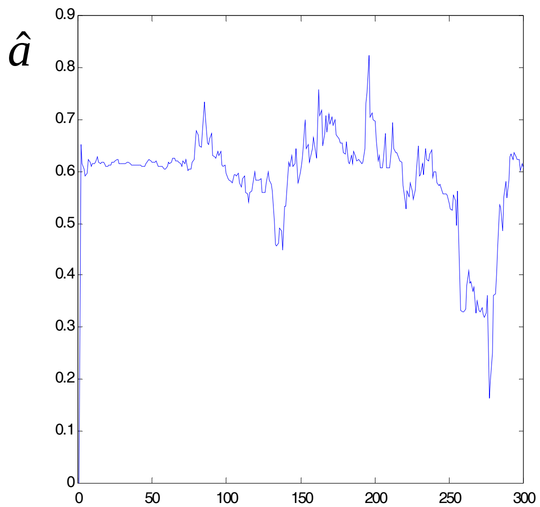
(este valor não se sabe na prática; aqui é usado apenas para comparação!)

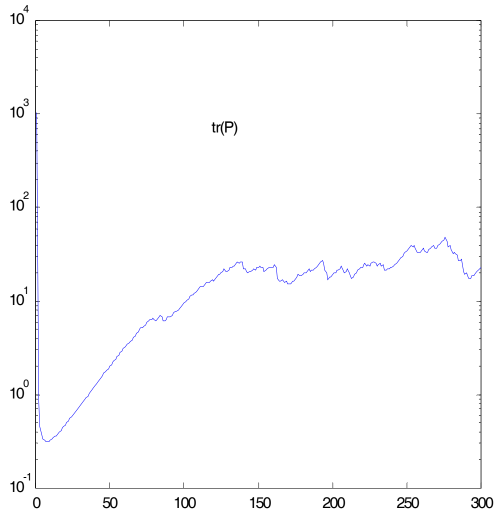
Usa-se RLS com factor de esquecimento exponencial e $\lambda = 0.95$

Consideram-se duas situações para a entrada:

- Entrada constante
- Entrada constante somada com ruído branco

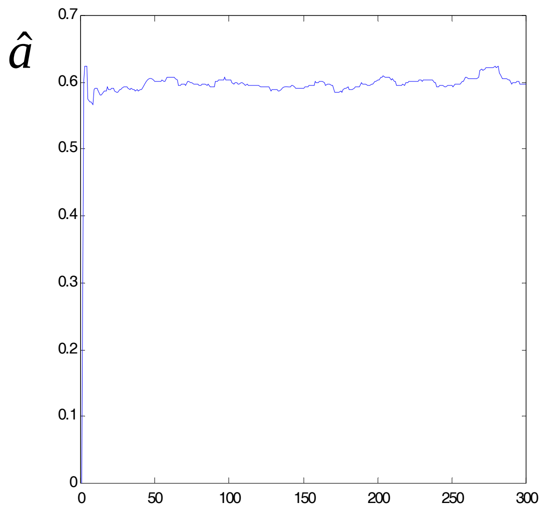
# Resultados com a entrada do sistema constante

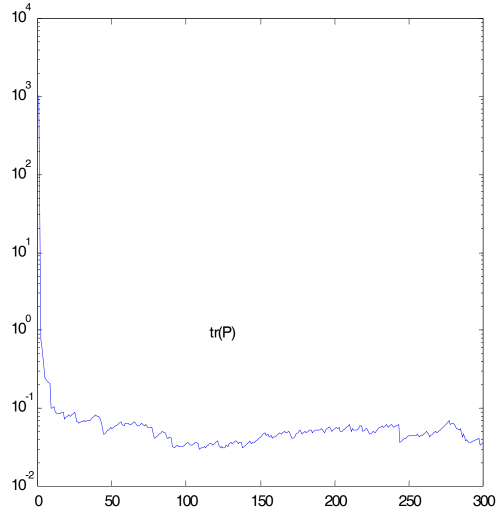



- traço de *P* dimunui inicialmente e a estimativa aproxima-se do valor verdadeiro $a = 0.6$. No entanto, devido a o sistema não ser excitado e a usar-se o algoritmo de esquecimento, *P* aumenta o causa um aumento do ganho de Kalman e leva a fortes variações da estimativa.

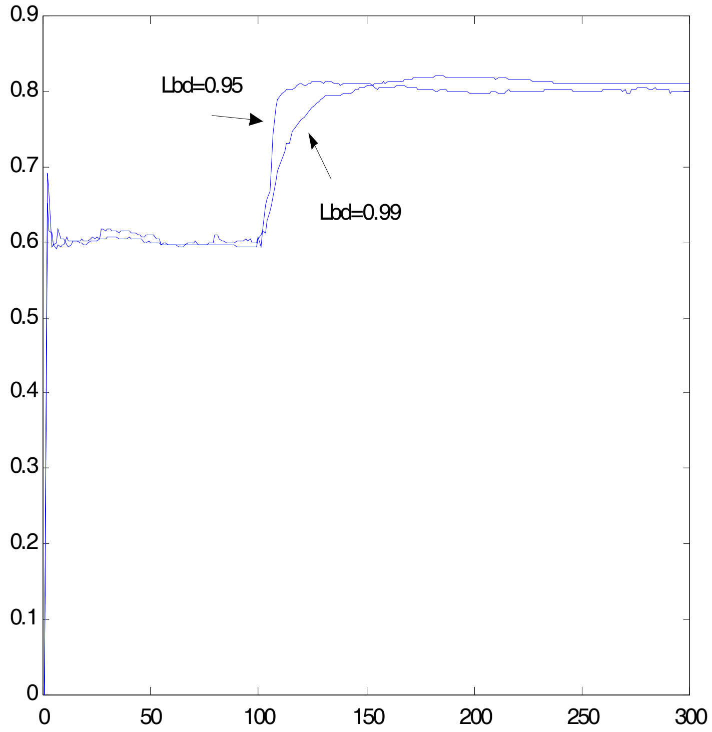
# Resultados com a entrada do sistema variável




Quando a entrada é suficientemente excitante, o traço de $P$ não volta a subir devido ao esquecimento e a estimativa mantém-se próximo do valor verdadeiro.

# Sistemas variáveis no tempo




O parâmetro $a$ muda em *t=100* de 0.6 para 0.8.

A figura mostra os resultados que se obtêm com dois valores diferentes do factor de esquecimento. Quando este é mais baixo, a transição da estimativa para o novo valor é mais rápida, mas em regime estacionário as fluctuações são maiores.

## Outro tipo de algoritmos de esquecimento

Para evitar os problemas com o esquecimento exponencial utilizam-se outros algoritmos. Dois exemplos importantes:

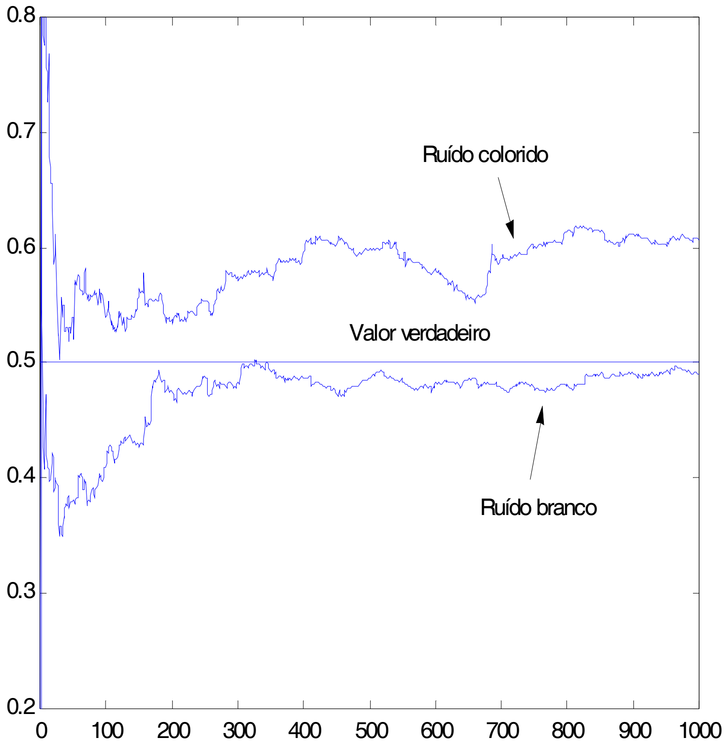
- **Esquecimento variável no tempo**. O factor de esquecimento é 1 quando a potência dos resíduos está abaixo de um dado limiar, sendo reduzido quando a potência aumenta. Deste modo eliminam-se os problemas com a explosão da covariância;
- **Esquecimento direccional**. O factor de esquecimento é diferente em diversas direcções do espaço de parâmetros, o que permite reduzir problemas devidos à não identificabilidade.

# Algoritmos numericamente robustos

As equações do algoritmo de RLS, tal como foram apresentadas, podem apresentar problemas numéricos (embora não no MATLAB e em simulações de pequena duração). Há várias formas de evitar estes problemas. Uma é o chamado algoritmo UD em que se propaga não a matriz *P* mas duas matrizes, U (triangular superior com 1's na diagonal) e *D* (diagonal), tal que $P = UDU^T$. Deste modo é possível trabalhar com uma gama de números menor e garantir o carácter definido positivo de *P*.

No capítulo 13 de AW pode ser visto um procedimento em PASCAL para este algoritmo.

# Efeito do ruído colorido




Em presença de ruído branco a estimativa de mínimos quadrados é centrada, *i. e.*, a média do erro é nula. Isto deixa de ser verdade quando o ruído é colorido. Na figura mostra-se a estimativa do parâmetro $a$ em

$$y(t+1) = ay(t) + bu(t) + e(t+1) + ce(t)$$

quando $c = 0$ e quando $c = 0.95$

O facto de os mínimos quadrados fornecerem estimativas polarizadas em presença de ruído colorido pode parecer à partida uma limitação importante.

Em classes importantes de controladores adaptativos de facto não o é.

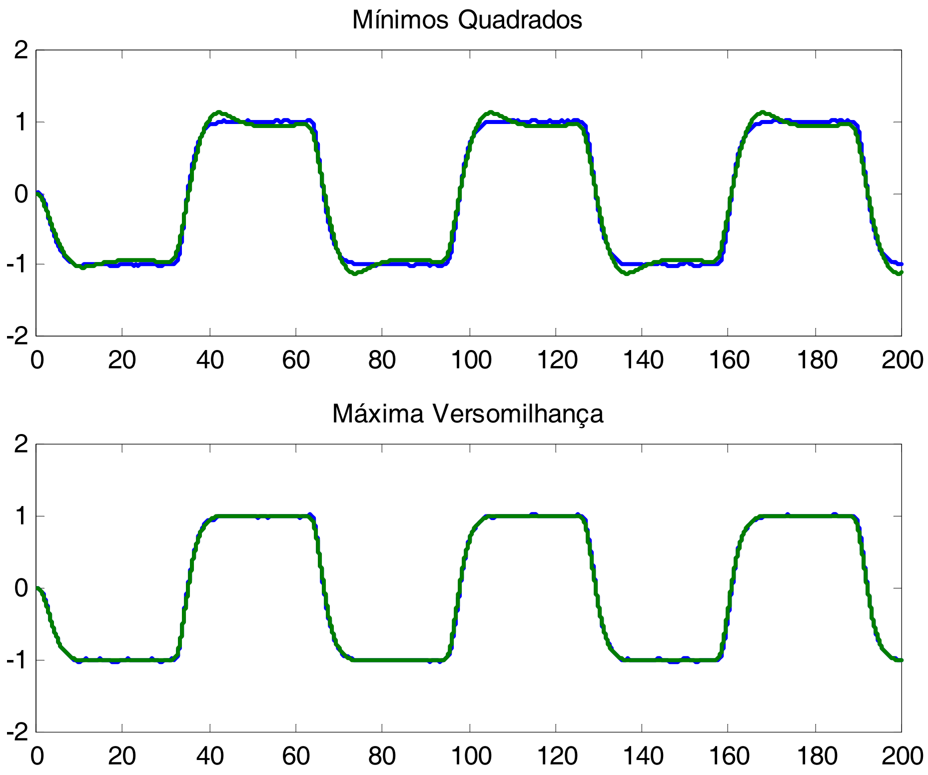
Isto é devido a, em **cadeia fechada** e sob certas condições, sistemas com ruído colorido admitirem um outro modelo com ruído branco. Como se verá, esta é a base do **controlo adaptativo auto-sintonizável** (*self-tuning*) descoberto no início dos anos 70 por Astrom e Wittenmark.

# Estimação de Máxima Verosimilhança com o MATLAB

```
z=[y u];
na=4;
nb=1;
nc=4;
nk=1;
nn=[na nb nc nk];
th=armax(z,nn);
yh=idsim(u,th);
plot(t,[y yh]);
[Phi,Gamma,C,D,K,X0]=th2ss(th);
T=0.05;
[A,B]=d2c(Phi,Gamma,T);
[num,den]=ss2tf(A,B,C,D);
```

A estimação de Máxima verosimilhança faz-se com a função *armax*

A estimação de Mínimos Quadrados faz-se com a função *arx*


Mínimos Quadrados
2
1
0
-1
-2
0 20 40 60 80 100 120 140 160 180 200
Máxima Versomilhança
2
1
0
-1
-2
0 20 40 60 80 100 120 140 160 180 200