Capítulo 5 - O Estimador de Mínimos Quadrados

Eduardo Mendes

emmendes@cpdee.ufmg.br

Departamento de Engenharia Eletrônica
Universidade Federal de Minas Gerais
Av. Antônio Carlos 6627, Belo Horizonte, MG, Brasil

Introdução

1) Gauss se refere a um número de observações *estritamente necessário* à determinação das grandezas desconhecidas;

Introdução

1) Gauss se refere a um número de observações *estritamente necessário* à determinação das grandezas desconhecidas;
2) Gauss deixa claro que, por causa dos erros de medição, mais observações do que o número mínimo são necessárias. O uso dessas observações "redundantes" resultará na redução dos efeitos dos erros;

Introdução

1) Gauss se refere a um número de observações *estritamente necessário* à determinação das grandezas desconhecidas;
2) Gauss deixa claro que, por causa dos erros de medição, mais observações do que o número mínimo são necessárias. O uso dessas observações "redundantes" resultará na redução dos efeitos dos erros;
3) o problema da modelagem está implícito na observação de que o movimento observado satisfaz as leis de Kepler;

Introdução

1) Gauss se refere a um número de observações *estritamente necessário* à determinação das grandezas desconhecidas;

2) Gauss deixa claro que, por causa dos erros de medição, mais observações do que o número mínimo são necessárias. O uso dessas observações "redundantes" resultará na redução dos efeitos dos erros;

3) o problema da modelagem está implícito na observação de que o movimento observado satisfaz as leis de Kepler;

4) os parâmetros estimados devem satisfazer as observações da forma mais exata possível. A diferença entre observação e cálculo (usando os parâmetros estimados) deve ser mínima.

Gerando Sistemas de Equações - Sistemas com solução única

- Seja uma função escalar $y = f(x)$ e várias aplicações da mesma

$$
\begin{aligned}
y_1 &= f(x_1) \\
y_2 &= f(x_2) \\
\vdots &= \vdots \\
y_N &= f(x_N).
\end{aligned}
$$

Gerando Sistemas de Equações - Sistemas com solução única

- Seja uma função escalar $y = f(x)$ e várias aplicações da mesma

$$
\begin{aligned}
y_1 &= f(x_1) \\
y_2 &= f(x_2) \\
\vdots &= \vdots \\
y_N &= f(x_N).
\end{aligned}
$$

- No caso vetorial, $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ depende de um vetor de n parâmetros, $\boldsymbol{\theta}$.

Gerando Sistemas de Equações - Sistemas com solução única

- Seja uma função escalar $y = f(x)$ e várias aplicações da mesma

$$\begin{aligned} y_1 &= f(x_1) \\ y_2 &= f(x_2) \\ \vdots &= \vdots \\ y_N &= f(x_N). \end{aligned}$$

- No caso vetorial, $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ depende de um vetor de n parâmetros, $\boldsymbol{\theta}$.
- Dizemos então que a função $f(\mathbf{x})$ é parametrizada por $\boldsymbol{\theta} \in \mathbb{R}^n$ e pode ser representada da seguinte forma:

$$y = f(\mathbf{x}, \boldsymbol{\theta}).$$

- Semelhantemente, tem-se

$$\begin{aligned} y_1 &= f(\mathbf{x}_1, \boldsymbol{\theta}) \\ y_2 &= f(\mathbf{x}_2, \boldsymbol{\theta}) \\ \vdots &= \vdots \\ y_N &= f(\mathbf{x}_N, \boldsymbol{\theta}). \end{aligned}$$

- Semelhantemente, tem-se

$$\begin{aligned} y_1 &= f(\mathbf{x}_1, \boldsymbol{\theta}) \\ y_2 &= f(\mathbf{x}_2, \boldsymbol{\theta}) \\ \vdots &= \vdots \\ y_N &= f(\mathbf{x}_N, \boldsymbol{\theta}). \end{aligned}$$

- Sendo conhecidos os conjuntos $\{y_1, \ldots, y_N\}$ e $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, é pertinente perguntar se é possível determinar f e $\boldsymbol{\theta}$.

Considerações

- A função f e o vetor $\boldsymbol{\theta}$ não variam de uma restrição para a outra, ou seja, todas as restrições são de fato da mesma equação.

Considerações

- A função f e o vetor $\boldsymbol{\theta}$ não variam de uma restrição para a outra, ou seja, todas as restrições são de fato da mesma equação.
- $y = f(\mathbf{x}, \boldsymbol{\theta})$ pode ser rescrita como

$$y = \mathbf{x}^{\mathrm{T}}\,\boldsymbol{\theta}.$$

Considerações

- A função f e o vetor $\boldsymbol{\theta}$ não variam de uma restrição para a outra, ou seja, todas as restrições são de fato da mesma equação.
- $y = f(\mathbf{x}, \boldsymbol{\theta})$ pode ser rescrita como

$$y = \mathbf{x}^{\mathrm{T}}\,\boldsymbol{\theta}.$$

- Serão tomadas n restrições a fim de se ter n equações para determinar os n elementos de $\boldsymbol{\theta}$, ou seja, nesse caso $N = n$.

- Tendo feito tais considerações, pode-se escrever

$$\begin{aligned} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} &= \begin{bmatrix} \mathbf{x}_1 \ \mathbf{x}_2 \ \ldots \ \mathbf{x}_n \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}, \\ \mathbf{y} &= X\,\boldsymbol{\theta}, \end{aligned}$$

sendo $X \in \mathbb{R}^{n \times n}$.

- Tendo feito tais considerações, pode-se escrever

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \ \mathbf{x}_2 \ \ldots \ \mathbf{x}_n \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix},$$
$$\mathbf{y} = X\boldsymbol{\theta},$$

sendo $X \in \mathbb{R}^{n \times n}$.

- Se X for não singular, é possível determinar o vetor de parâmetros invertendo tal matriz, ou seja,

$$\boldsymbol{\theta} = X^{-1}\mathbf{y}.$$

Sistema de equações Wiener-Hopf com solução única

- Neste exemplo, partindo da equação de Wiener-Hopf, será montado e resolvido um sistema de equações com solução única.

$$\begin{bmatrix} r_{uy}(0) \\ r_{uy}(1) \\ \vdots \\ r_{uy}(N-1) \end{bmatrix} = \begin{bmatrix} r_{uu}(0) & r_{uu}(-1) & \dots & r_{uu}(-N+1) \\ r_{uu}(1) & r_{uu}(0) & \dots & r_{uu}(-N+2) \\ \vdots & \vdots & \vdots & \vdots \\ r_{uu}(N-1) & r_{uu}(N-2) & \dots & r_{uu}(0) \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(N-1) \end{bmatrix}$$

$$\mathbf{r}_{uy} = R_{uu}\,\mathbf{h}.$$

Sistema de equações Wiener-Hopf com solução única

- Neste exemplo, partindo da equação de Wiener-Hopf, será montado e resolvido um sistema de equações com solução única.

$$
\begin{aligned}
\begin{bmatrix} r_{uy}(0) \\ r_{uy}(1) \\ \vdots \\ r_{uy}(N-1) \end{bmatrix} &= \begin{bmatrix} r_{uu}(0) & r_{uu}(-1) & \dots & r_{uu}(-N+1) \\ r_{uu}(1) & r_{uu}(0) & \dots & r_{uu}(-N+2) \\ \vdots & \vdots & \vdots & \vdots \\ r_{uu}(N-1) & r_{uu}(N-2) & \dots & r_{uu}(0) \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(N-1) \end{bmatrix} \\
\mathbf{r}_{uy} &= R_{uu}\,\mathbf{h}.
\end{aligned}
$$

- Por construção, a matriz R_{uu} é quadrada, e se ela tiver inversa, tem-se

$$\mathbf{h} = R_{uu}^{-1}\,\mathbf{r}_{uy}.$$

Sistemas sobredeterminados

Se $N > n$ restrições forem tomadas, tem-se então um sistema de equações sobredeterminado. Como a matriz X não é quadrada, ela não pode ser invertida. Entretanto, pré-multiplicando a equação $\mathbf{y} = X\,\boldsymbol{\theta}$ por X^{T} em ambos os lados tem-se

$$X^{\mathrm{T}}\mathbf{y} = X^{\mathrm{T}}X\,\boldsymbol{\theta},$$

que é chamada de *equação normal* e

$$\boldsymbol{\theta} = [X^{\mathrm{T}}X]^{-1}X^{\mathrm{T}}\mathbf{y},$$

no caso de $X^{\mathrm{T}}X$ não ser singular. A matriz $[X^{\mathrm{T}}X]^{-1}X^{\mathrm{T}}$, na equação acima, é conhecida como a *matriz pseudo-inversa*.

O Método de Mínimos Quadrados

Será assumido que se conhece o valor estimado do vetor de parâmetros, $\hat{\boldsymbol{\theta}}$, e que é cometido um erro ξ, ou seja,

$$y = \mathbf{x}^{\mathrm{T}}\,\hat{\boldsymbol{\theta}} + \xi.$$

$N > n$ aplicações em forma matricial são

$$\mathbf{y} = X\,\hat{\boldsymbol{\theta}} + \boldsymbol{\xi},$$

sendo que $\boldsymbol{\xi} \in \mathbb{R}^N$ é o vetor de erros cometidos ao se tentar explicar **y** como $X\,\hat{\boldsymbol{\theta}}$.

- Intuitivamente, seria interessante que $\hat{\boldsymbol{\theta}}$ fosse tal que reduzisse o somatório do quadrado dos erros

$$J_{\mathrm{MQ}} = \sum_{i=1}^{N} \xi(i)^2 = \boldsymbol{\xi}^{\mathrm{T}} \boldsymbol{\xi} = \| \boldsymbol{\xi} \|^2 \, .$$

$$\begin{aligned} J_{\mathrm{MQ}} &= (\mathbf{y} - X\,\hat{\boldsymbol{\theta}})^{\mathrm{T}} (\mathbf{y} - X\,\hat{\boldsymbol{\theta}}) \\ &= \mathbf{y}^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}} X \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{\mathrm{T}} X^{\mathrm{T}} \mathbf{y} + \hat{\boldsymbol{\theta}}^{\mathrm{T}} X^{\mathrm{T}} X \hat{\boldsymbol{\theta}}. \end{aligned}$$

- Intuitivamente, seria interessante que $\hat{\boldsymbol{\theta}}$ fosse tal que reduzisse o somatório do quadrado dos erros

$$J_{\mathrm{MQ}} = \sum_{i=1}^{N} \xi(i)^2 = \boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{\xi} = \| \boldsymbol{\xi} \|^2 .$$

$$\begin{aligned} J_{\mathrm{MQ}} &= (\mathbf{y} - X\,\hat{\boldsymbol{\theta}})^{\mathrm{T}}(\mathbf{y} - X\,\hat{\boldsymbol{\theta}}) \\ &= \mathbf{y}^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}}X\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{\mathrm{T}}X^{\mathrm{T}}\mathbf{y} + \hat{\boldsymbol{\theta}}^{\mathrm{T}}X^{\mathrm{T}}X\hat{\boldsymbol{\theta}}. \end{aligned}$$

- A fim de minimizar a função de custo J_{MQ} com respeito a $\hat{\boldsymbol{\theta}}$, é necessário resolver $(\partial J_{\mathrm{MQ}}/\partial\hat{\boldsymbol{\theta}}) = 0$. Fazendo-se isso, tem-se

$$\begin{aligned} \frac{\partial J_{\mathrm{MQ}}}{\partial\hat{\boldsymbol{\theta}}} &= -(\mathbf{y}^{\mathrm{T}}X)^{\mathrm{T}} - X^{\mathrm{T}}\mathbf{y} + (X^{\mathrm{T}}X + X^{\mathrm{T}}X)\hat{\boldsymbol{\theta}} \\ &= -X^{\mathrm{T}}\mathbf{y} - X^{\mathrm{T}}\mathbf{y} + 2X^{\mathrm{T}}X\,\hat{\boldsymbol{\theta}}. \end{aligned}$$

- Igualando-se a última equação a zero tem-se

$$\hat{\boldsymbol{\theta}} = [X^{\mathrm{T}}X]^{-1}X^{\mathrm{T}}\mathbf{y}.$$

- Igualando-se a última equação a zero tem-se

$$\hat{\boldsymbol{\theta}} = [X^{\mathrm{T}} X]^{-1} X^{\mathrm{T}} \mathbf{y}.$$

- Para que $\hat{\boldsymbol{\theta}}$ seja o mínimo, é necessário verificar que

$$\frac{\partial^2 J_{\mathrm{MQ}}}{\partial \hat{\boldsymbol{\theta}}^2} = 2X^{\mathrm{T}} X > 0.$$

A equação acima é verdadeira, pois $2X^{\mathrm{T}} X$ é positiva definida por construção.

- Igualando-se a última equação a zero tem-se

$$\hat{\boldsymbol{\theta}} = [X^{\mathrm{T}} X]^{-1} X^{\mathrm{T}} \mathbf{y}.$$

- Para que $\hat{\boldsymbol{\theta}}$ seja o mínimo, é necessário verificar que

$$\frac{\partial^2 J_{\mathrm{MQ}}}{\partial \hat{\boldsymbol{\theta}}^2} = 2X^{\mathrm{T}} X > 0.$$

  A equação acima é verdadeira, pois $2X^{\mathrm{T}}X$ é positiva definida por construção.

- Resumindo,

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\mathrm{MQ}} &= \arg_{\boldsymbol{\theta}} \min J_{\mathrm{MQ}} \\ &= [X^{\mathrm{T}} X]^{-1} X^{\mathrm{T}} \mathbf{y}. \end{aligned}$$

Estimação dos parâmetros de uma reta usando MQ

1) Determinar uma função f parametrizada por $\boldsymbol{\theta}$.
   $f(x) = \theta_1 + \theta_2 x$ é uma boa função candidata.

Estimação dos parâmetros de uma reta usando MQ

1) Determinar uma função f parametrizada por $\boldsymbol{\theta}$.
   $f(x) = \theta_1 + \theta_2 x$ é uma boa função candidata.

2) As vinte e uma aplicações podem ser colocadas em formato matricial, da seguinte forma: $\mathbf{y} = X\boldsymbol{\theta}$, sendo $\mathbf{y} \in \mathbb{R}^{21\times 1}$, $X \in \mathbb{R}^{21\times 2}$ e $\boldsymbol{\theta} \in \mathbb{R}^{2\times 1}$.

Estimação dos parâmetros de uma reta usando MQ

1) Determinar uma função f parametrizada por $\boldsymbol{\theta}$.
   $f(x) = \theta_1 + \theta_2 x$ é uma boa função candidata.

2) As vinte e uma aplicações podem ser colocadas em formato matricial, da seguinte forma: $\mathbf{y} = X\boldsymbol{\theta}$, sendo $\mathbf{y} \in \mathbb{R}^{21\times1}$, $X \in \mathbb{R}^{21\times2}$ e $\boldsymbol{\theta} \in \mathbb{R}^{2\times1}$.

3) Usando-se a equação pseudo-inversa chega-se à reta $y = -0{,}9567\times10^{-3} + 7{,}142\times10^{-4}x$.

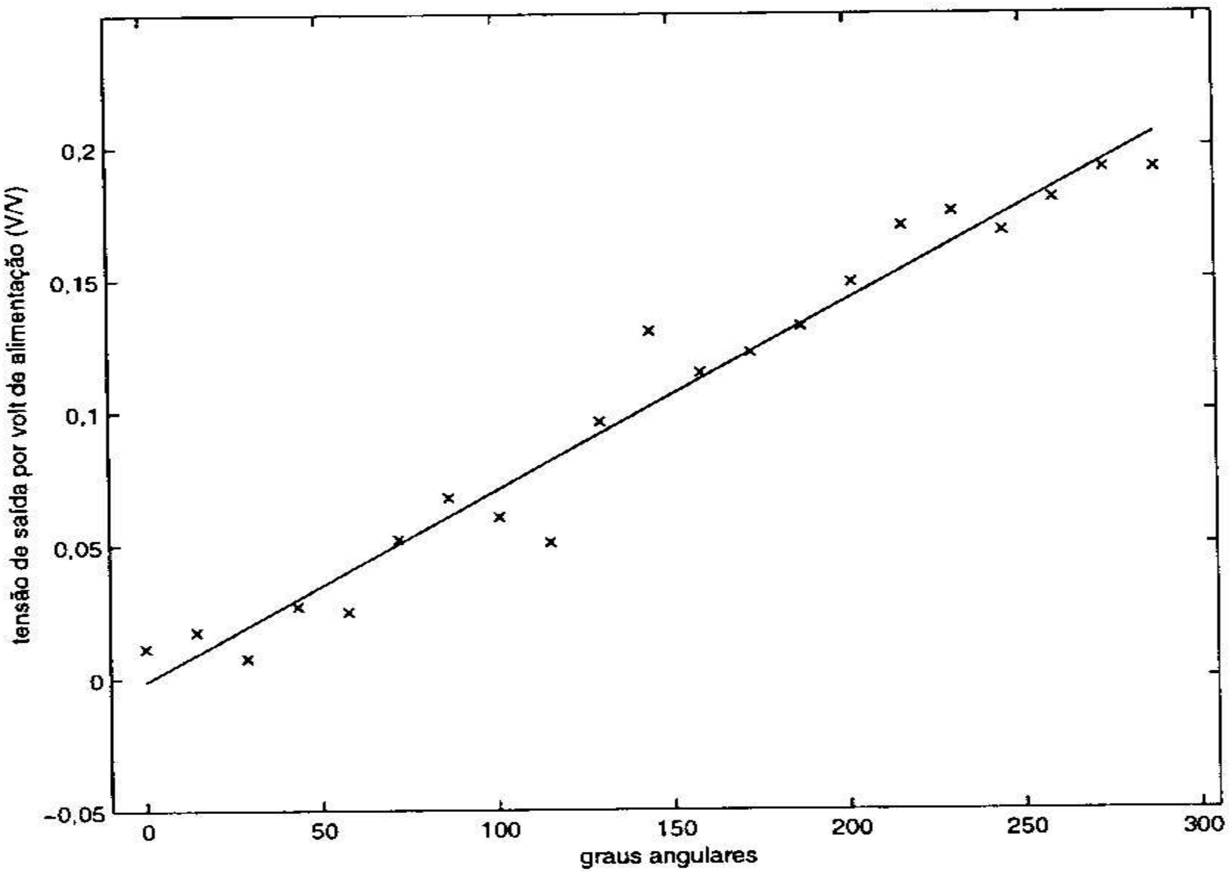

Figura 1: Estimação dos parâmetros de uma reta (x) dados medidos e (–) reta ajustada usando o método MQ .

Relação entre o estimador MQ e funções de correlação

- Portanto, usando-se o somatório de convolução para definir um sistema de equações, tem-se

$$
\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(N-1) \end{bmatrix} = \begin{bmatrix} u(0) & u(-1) & \dots & u(1-n_\mathrm{i}) \\ u(1) & u(0) & \dots & u(2-n_\mathrm{i}) \\ \vdots & \vdots & \dots & \vdots \\ u(N-1) & u(N-2) & \dots & u(N-n_\mathrm{i}) \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(n_\mathrm{i}-1) \end{bmatrix}
$$

$$
\mathbf{y} = U\,\mathbf{h}.
$$

Relação entre o estimador MQ e funções de correlação

- Portanto, usando-se o somatório de convolução para definir um sistema de equações, tem-se

$$\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(N-1) \end{bmatrix} = \begin{bmatrix} u(0) & u(-1) & \dots & u(1-n_{\mathrm{i}}) \\ u(1) & u(0) & \dots & u(2-n_{\mathrm{i}}) \\ \vdots & \vdots & \dots & \vdots \\ u(N-1) & u(N-2) & \dots & u(N-n_{\mathrm{i}}) \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(n_{\mathrm{i}}-1) \end{bmatrix}$$

$$\mathbf{y} = U\,\mathbf{h}.$$

- A estimativa usando-se mínimos quadrados é

$$\hat{\mathbf{h}} = [U^{\mathrm{T}}U]^{-1}U^{\mathrm{T}}\mathbf{y}.$$

Desenvolvendo-se a matriz $U^{\mathrm{T}}U$, tem-se

$$U^{\mathrm{T}}U = \begin{bmatrix} u(0) & u(1) & \dots & u(N-1) \\ u(-1) & u(0) & \dots & u(N-2) \\ \vdots & \vdots & \dots & \vdots \\ u(1-n_{\mathrm{i}}) & u(2-n_{\mathrm{i}}) & \dots & u(N-n_{\mathrm{i}}) \end{bmatrix} \times$$

$$\times \begin{bmatrix} u(0) & u(-1) & \dots & u(1-n_{\mathrm{i}}) \\ u(1) & u(0) & \dots & u(2-n_{\mathrm{i}}) \\ \vdots & \vdots & \dots & \vdots \\ u(N-1) & u(N-2) & \dots & u(N-n_{\mathrm{i}}) \end{bmatrix}$$

$$
U^{\mathrm{T}}U = \begin{bmatrix} \sum_i u(i)^2 & \sum_i u(i)u(i-1) & \dots \\ \sum_i u(i-1)u(i) & \sum_i u(i-1)^2 & \dots \\ \vdots & \vdots & \vdots \\ \sum_i u(i+1-n_{\mathrm{i}})u(i) & \sum_i u(i+1-n_{\mathrm{i}})u(i-1) & \dots \end{bmatrix}
$$

$$
\begin{bmatrix} \dots & \sum_i u(i)u(i+1-n_{\mathrm{i}}) \\ \dots & \sum_i u(i-1)u(i+1-n_{\mathrm{i}}) \\ \vdots & \vdots \\ \dots & \sum_i u(i+1-n_{\mathrm{i}})^2 \end{bmatrix}.
$$

- Lembrando-se de que para o caso estacionário

$$r_{uu}(k) = \frac{1}{N}\sum_{i=0}^{N} u(i)u(i-k) = r_{uu}(-k),$$

$$U^{\mathrm{T}}U \quad = \quad N\begin{bmatrix} r_{uu}(0) & r_{uu}(1) & \dots & r_{uu}(n_{\mathrm{i}}-1) \\ r_{uu}(1) & r_{uu}(0) & \dots & r_{uu}(n_{\mathrm{i}}-2) \\ \dots & \dots & \vdots & \dots \\ r_{uu}(n_{\mathrm{i}}-1) & r_{uu}(n_{\mathrm{i}}-1) & \dots & r_{uu}(0) \end{bmatrix} = NR_{uu}.$$

- Lembrando-se de que para o caso estacionário

$$r_{uu}(k) = \frac{1}{N}\sum_{i=0}^{N} u(i)u(i-k) = r_{uu}(-k),$$

$$U^{\mathrm{T}}U = N\begin{bmatrix} r_{uu}(0) & r_{uu}(1) & \dots & r_{uu}(n_{\mathrm{i}}-1) \\ r_{uu}(1) & r_{uu}(0) & \dots & r_{uu}(n_{\mathrm{i}}-2) \\ \dots & \dots & \vdots & \dots \\ r_{uu}(n_{\mathrm{i}}-1) & r_{uu}(n_{\mathrm{i}}-1) & \dots & r_{uu}(0) \end{bmatrix} = NR_{uu}.$$

- De forma análoga, tem-se

$$U^{\mathrm{T}}\mathbf{y} = N\begin{bmatrix} r_{uy}(1) \\ r_{uy}(2) \\ \vdots \\ r_{uy}(n_{\mathrm{i}}) \end{bmatrix} = N\mathbf{r}_{uy}.$$

- Portanto,

$$
\begin{aligned}
\hat{\mathbf{h}} &= \frac{1}{N} R_{uu}^{-1}\, N\, \mathbf{r}_{uy} \\
\hat{\mathbf{h}} &= R_{uu}^{-1} \mathbf{r}_{uy}.
\end{aligned}
$$

$h(k)$ pode ser estimada de duas formas distintas

1) calculando-se $r_{uu}(k)$ e $r_{uy}(k)$, que têm boa robustez ao ruído, e resolvendo-se $\hat{\mathbf{h}} = R_{uu}^{-1}\,\mathbf{r}_{uy}$;

$h(k)$ pode ser estimada de duas formas distintas

1) calculando-se $r_{uu}(k)$ e $r_{uy}(k)$, que têm boa robustez ao ruído, e resolvendo-se $\hat{\mathbf{h}} = R_{uu}^{-1}\,\mathbf{r}_{uy}$;

2) usando-se os mesmos sinais de entrada e saída *diretamente* no somatório de convolução, montando-se e resolvendo-se o sistema sobredeterminado de equações resultante usando-se o estimador de mínimos quadrados. Ou seja, $\hat{\mathbf{h}} = [U^{\mathrm{T}}U]^{-1}U^{\mathrm{T}}\mathbf{y}$.

$h(k)$ pode ser estimada de duas formas distintas

1) calculando-se $r_{uu}(k)$ e $r_{uy}(k)$, que têm boa robustez ao ruído, e resolvendo-se $\hat{\mathbf{h}} = R_{uu}^{-1}\,\mathbf{r}_{uy}$;
2) usando-se os mesmos sinais de entrada e saída *diretamente* no somatório de convolução, montando-se e resolvendo-se o sistema sobredeterminado de equações resultante usando-se o estimador de mínimos quadrados. Ou seja, $\hat{\mathbf{h}} = [U^{\mathrm{T}}U]^{-1}U^{\mathrm{T}}\mathbf{y}$.

Ou seja

$$\hat{\mathbf{h}} = R_{uu}^{-1}\,\mathbf{r}_{uy} = [U^{\mathrm{T}}U]^{-1}U^{\mathrm{T}}\mathbf{y}.$$

Portanto, o estimador MQ goza das mesmas propriedades de robustez ao ruído que o método de identificação utilizando as funções de correlação.

Ortogonalidade

- Se **x**, **y** e **z** forem vetores com a mesma dimensão e **z** = **x**+**y**, então

$$\begin{aligned} \| \mathbf{z} \|^2 &= \mathbf{z}^{\mathrm{T}}\mathbf{z} \\ &= (\mathbf{x}+\mathbf{y})^{\mathrm{T}}(\mathbf{x}+\mathbf{y}) \\ &= \mathbf{x}^{\mathrm{T}}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{y} + \mathbf{y}^{\mathrm{T}}\mathbf{x} + \mathbf{y}^{\mathrm{T}}\mathbf{y}. \end{aligned}$$

Ortogonalidade

- Se **x**, **y** e **z** forem vetores com a mesma dimensão e **z** = **x**+**y**, então

$$\begin{aligned} \| \mathbf{z} \|^2 &= \mathbf{z}^{\mathrm{T}}\mathbf{z} \\ &= (\mathbf{x}+\mathbf{y})^{\mathrm{T}}(\mathbf{x}+\mathbf{y}) \\ &= \mathbf{x}^{\mathrm{T}}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{y} + \mathbf{y}^{\mathrm{T}}\mathbf{x} + \mathbf{y}^{\mathrm{T}}\mathbf{y}. \end{aligned}$$

- Se **x** e **y** forem ortogonais, então $\mathbf{x}^{\mathrm{T}}\mathbf{y} = \mathbf{y}^{\mathrm{T}}\mathbf{x} = 0$. Portanto, se $\mathbf{x} \perp \mathbf{y}$, então

$$\begin{aligned} \| \mathbf{z} \|^2 &= \mathbf{x}^{\mathrm{T}}\mathbf{x} + \mathbf{y}^{\mathrm{T}}\mathbf{y} \\ &= \| \mathbf{x} \|^2 + \| \mathbf{y} \|^2, \end{aligned}$$

ou seja, **x** não tem componente sobre **y** e vice-versa.

- Será utilizada a seguinte nomenclatura $\hat{\mathbf{y}} = X\hat{\boldsymbol{\theta}}$.

$$
\begin{aligned}
\hat{\mathbf{y}}^{\mathrm{T}}\boldsymbol{\xi} &= \hat{\boldsymbol{\theta}}_{\mathrm{MQ}}^{\mathrm{T}} X^{\mathrm{T}} \left(\mathbf{y} - X\hat{\boldsymbol{\theta}}_{\mathrm{MQ}}^{\mathrm{T}}\right) \\
&= \left\{[X^{\mathrm{T}}X]^{-1}X^{\mathrm{T}}\mathbf{y}\right\}^{\mathrm{T}} X^{\mathrm{T}} \left(\mathbf{y} - X[X^{\mathrm{T}}X]^{-1}X^{\mathrm{T}}\mathbf{y}\right) \\
&= \mathbf{y}^{\mathrm{T}}X[X^{\mathrm{T}}X]^{-1}X^{\mathrm{T}} \left(\mathbf{y} - X[X^{\mathrm{T}}X]^{-1}X^{\mathrm{T}}\mathbf{y}\right) \\
&= \mathbf{y}^{\mathrm{T}}X[X^{\mathrm{T}}X]^{-1}X^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}}X[X^{\mathrm{T}}X]^{-1}X^{\mathrm{T}}X[X^{\mathrm{T}}X]^{-1}X^{\mathrm{T}}\mathbf{y} \\
&= 0.
\end{aligned}
$$

Os vetores $\hat{\mathbf{y}}$ e $\boldsymbol{\xi}$ são ortogonais quando o correspondente vetor de parâmetros é estimado, usando-se o estimador MQ.

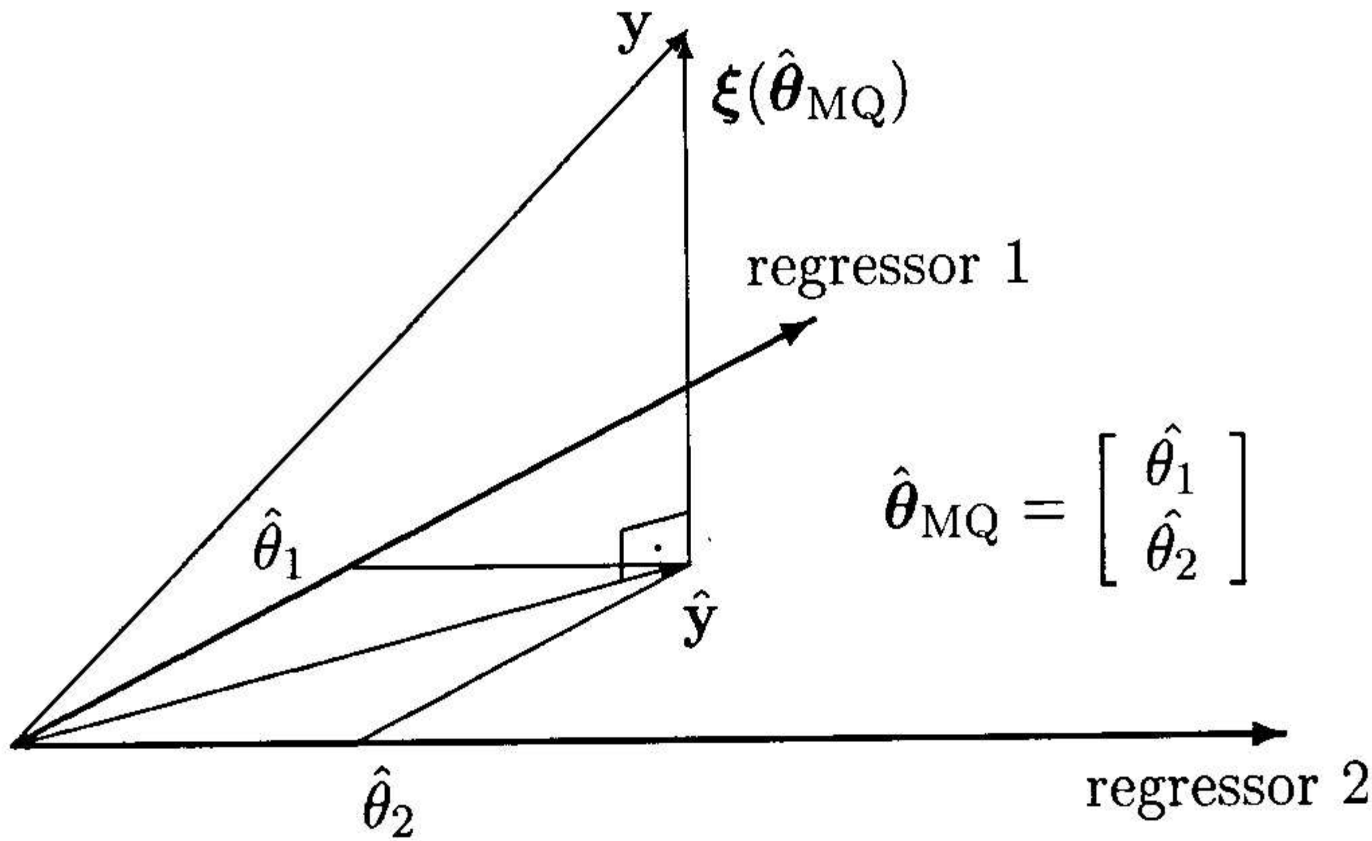

Figura 2: Propriedade de ortogonalidade do estimador de MQ Representação gráfica da propriedade de ortogonalidade do estimador MQ, para uma superfície de ordem dois.

- A interpretação da Figura anterior é a seguinte. Dado um vetor de observações **y** e uma matriz de regressores, X, o estimador MQ determinará parâmetros que definem $\hat{\mathbf{y}}$ de tal forma que $\hat{\mathbf{y}}$ é a projeção ortogonal de **y** sobre a hipersuperfície gerada pelas colunas de X, sendo a matriz de projeção $P(X)$ tal que

$$\hat{\mathbf{y}} = X\hat{\boldsymbol{\theta}} = X[X^{\mathrm{T}}X]^{-1}X^{\mathrm{T}}\mathbf{y} = P(X)\mathbf{y}.$$

- A interpretação da Figura anterior é a seguinte. Dado um vetor de observações **y** e uma matriz de regressores, X, o estimador MQ determinará parâmetros que definem $\hat{\mathbf{y}}$ de tal forma que $\hat{\mathbf{y}}$ é a projeção ortogonal de **y** sobre a hipersuperfície gerada pelas colunas de X, sendo a matriz de projeção $P(X)$ tal que

$$\hat{\mathbf{y}} = X\hat{\boldsymbol{\theta}} = X[X^{\mathrm{T}}X]^{-1}X^{\mathrm{T}}\mathbf{y} = P(X)\mathbf{y}.$$

- Conseqüentemente, parâmetros são estimados de forma que a diferença (distância) entre os dados **y** e $\hat{\mathbf{y}}$ é mínima. A equação acima mostra que tanto $\hat{\boldsymbol{\theta}}$ quanto $\hat{\mathbf{y}}$ são lineares nas obsevações **y**.

Ortogonalidade de estimadores do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$

- O objetivo do presente exemplo é derivar as condições de ortogonalidade para estimadores do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$.

Ortogonalidade de estimadores do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$

- O objetivo do presente exemplo é derivar as condições de ortogonalidade para estimadores do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$.
- A condição para que cada regressor seja ortogonal a $\boldsymbol{\xi}$, é

$$\begin{aligned} X^{\mathrm{T}}\boldsymbol{\xi} &= X^{\mathrm{T}}(\mathbf{y} - \hat{\mathbf{y}}) \\ &= X^{\mathrm{T}}(\mathbf{y} - X\hat{\boldsymbol{\theta}}) \\ &= X^{\mathrm{T}}(\mathbf{y} - XA\mathbf{y}) = 0. \end{aligned}$$

Ortogonalidade de estimadores do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$

- O objetivo do presente exemplo é derivar as condições de ortogonalidade para estimadores do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$.
- A condição para que cada regressor seja ortogonal a $\boldsymbol{\xi}$, é

$$\begin{aligned} X^{\mathrm{T}}\boldsymbol{\xi} &= X^{\mathrm{T}}(\mathbf{y} - \hat{\mathbf{y}}) \\ &= X^{\mathrm{T}}(\mathbf{y} - X\hat{\boldsymbol{\theta}}) \\ &= X^{\mathrm{T}}(\mathbf{y} - XA\mathbf{y}) = 0. \end{aligned}$$

- Portanto,

$$X^{\mathrm{T}}XA\mathbf{y} = X^{\mathrm{T}}\mathbf{y}.$$

O estimador de mínimos quadrados ponderados

- Considere a seguinte função

$$J_{\mathrm{MQP}} = \sum_{i=1}^{N} \xi(i) w_i \xi(i) = \boldsymbol{\xi}^{\mathrm{T}} W \boldsymbol{\xi},$$

sendo que $W \in \mathbb{R}^{N\times N}$ é uma matriz diagonal cujos elementos são os pesos w_i, ou seja, $W = \mathrm{diag}\{w_1 \; w_2 \; \ldots \; w_N\}$.

$$\frac{\partial J_{\mathrm{MQP}}}{\partial \hat{\boldsymbol{\theta}}} = 2(X^{\mathrm{T}} W X \hat{\boldsymbol{\theta}} - X^{\mathrm{T}} W \mathbf{y}).$$

O estimador de mínimos quadrados ponderados

- Considere a seguinte função

$$J_{\mathrm{MQP}} = \sum_{i=1}^{N} \xi(i) w_i \xi(i) = \boldsymbol{\xi}^{\mathrm{T}} W \boldsymbol{\xi},$$

  sendo que $W \in \mathbb{R}^{N \times N}$ é uma matriz diagonal cujos elementos são os pesos w_i, ou seja, $W = \mathrm{diag}\{w_1 \; w_2 \; \ldots \; w_N\}$.

$$\frac{\partial J_{\mathrm{MQP}}}{\partial \hat{\boldsymbol{\theta}}} = 2(X^{\mathrm{T}} W X \hat{\boldsymbol{\theta}} - X^{\mathrm{T}} W \mathbf{y}).$$

- Igualando-se a expressão acima a zero e resolvendo-se para $\hat{\boldsymbol{\theta}}$, tem-se

$$\hat{\boldsymbol{\theta}}_{\mathrm{MQP}} = [X^{\mathrm{T}} W X]^{-1} X^{\mathrm{T}} W \mathbf{y}.$$

O estimador de mínimos quadrados ponderados

- Considere a seguinte função

$$J_{\mathrm{MQP}} = \sum_{i=1}^{N} \xi(i) w_i \xi(i) = \boldsymbol{\xi}^{\mathrm{T}} W \boldsymbol{\xi},$$

sendo que $W \in \mathbb{R}^{N\times N}$ é uma matriz diagonal cujos elementos são os pesos w_i, ou seja, $W = \mathrm{diag}\{w_1 \; w_2 \; \ldots \; w_N\}$.

$$\frac{\partial J_{\mathrm{MQP}}}{\partial \hat{\boldsymbol{\theta}}} = 2(X^{\mathrm{T}} W X \hat{\boldsymbol{\theta}} - X^{\mathrm{T}} W \mathbf{y}).$$

- Igualando-se a expressão acima a zero e resolvendo-se para $\hat{\boldsymbol{\theta}}$, tem-se

$$\hat{\boldsymbol{\theta}}_{\mathrm{MQP}} = [X^{\mathrm{T}} W X]^{-1} X^{\mathrm{T}} W \mathbf{y}.$$

- Se W for semidefinida positiva, então $X^{\mathrm{T}} W X$ também o será e existirá a inversa.

Estimação dos parâmetros de uma reta usando MQP

No presente exemplo será considerado o mesmo problema do exemplo da reta e a mesma massa de dados. Serão atribuídos pesos diferentes às observações 9, 10 e 11.

$$W = \mathrm{diag}\left[w_1 \ldots w_{21}\right],$$

sendo $w_i = a,\ i = 1, \ldots, 8, 12, \ldots, 21$ e $w_i = b,\ i = 9, 10, 11$. Dois casos foram calculados. No primeiro $b = 100a$ e no segundo $b = 1000a$.

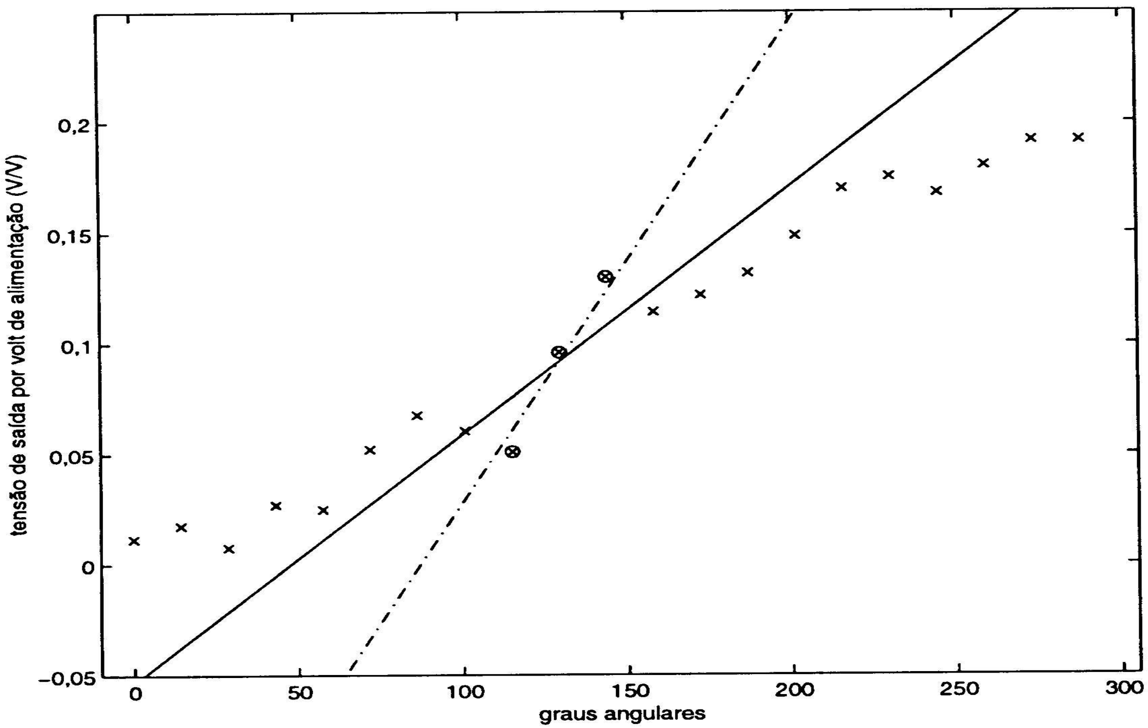

Figura 3: Estimação dos parâmetros de uma reta usando MQP (x) dados medidos, e retas ajustadas usando o estimador MQP com (—) $b = 100a$ e $(- \cdot -)$ $b = 1000a$. As três observações que receberam peso diferenciado estão indicadas $\otimes$.

O estimador MQP sem a propriedade de ortogonalidade

- É interessante notar que o estimador MQP é um estimador linear do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$, com $A = [X^{\mathrm{T}}WX]^{-1}X^{\mathrm{T}}W$. Para esse estimador tem-se

$$\begin{aligned} X^{\mathrm{T}}XA\mathbf{y} &\stackrel{?}{=} X^{\mathrm{T}}\mathbf{y} \\ X^{\mathrm{T}}X[X^{\mathrm{T}}WX]^{-1}X^{\mathrm{T}}W\mathbf{y} &\stackrel{?}{=} X^{\mathrm{T}}\mathbf{y}, \end{aligned} \tag{-39}$$

sendo que a condição não é satisfeita para o estimador MQP.

O estimador MQP sem a propriedade de ortogonalidade

- É interessante notar que o estimador MQP é um estimador linear do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$, com $A = [X^{\mathrm{T}}WX]^{-1}X^{\mathrm{T}}W$. Para esse estimador tem-se

$$\begin{aligned} X^{\mathrm{T}}XA\mathbf{y} &\overset{?}{=} X^{\mathrm{T}}\mathbf{y} \\ X^{\mathrm{T}}X[X^{\mathrm{T}}WX]^{-1}X^{\mathrm{T}}W\mathbf{y} &\overset{?}{=} X^{\mathrm{T}}\mathbf{y}, \end{aligned} \qquad (-38)$$

sendo que a condição não é satisfeita para o estimador MQP.

- Percebe-se que na equação acima seria errado fazer o seguinte:

$$\cancel{X}^{\mathrm{T}}X[X^{\mathrm{T}}WX]^{-1}X^{\mathrm{T}}W\mathbf{y} = \cancel{X}^{\mathrm{T}}\mathbf{y}, \qquad (-39)$$

uma vez que isso equivaleria a dividir ambos os lados da equação por X^{T}, o que é uma impossibilidade, pois X não é uma matriz quadrada.

Estimação de Parâmetros de Modelos ARX Usando MQ

- Será comum representar o modelo como

$$y(k \mid \hat{\boldsymbol{\theta}}) = \boldsymbol{\psi}^{\mathrm{T}}(k-1)\,\hat{\boldsymbol{\theta}} + \boldsymbol{\xi}(k),$$

sendo que k indica o instante considerado e $\boldsymbol{\psi}(k-1)$ é um vetor de $n_\theta = \dim[\hat{\boldsymbol{\theta}}]$ variáveis regressoras,

$$\boldsymbol{\psi}(k-1) = [\psi_1 \;\; \psi_2 \;\; \ldots \;\; \psi_{n_\theta}]^{\mathrm{T}},$$

tomadas até o instante $k-1$.

Estimação de Parâmetros de Modelos ARX Usando MQ

- Será comum representar o modelo como

$$y(k \mid \hat{\boldsymbol{\theta}}) = \boldsymbol{\psi}^{\mathrm{T}}(k-1)\,\hat{\boldsymbol{\theta}} + \boldsymbol{\xi}(k),$$

  sendo que k indica o instante considerado e $\boldsymbol{\psi}(k-1)$ é um vetor de $n_\theta = \dim[\hat{\boldsymbol{\theta}}]$ variáveis regressoras,

$$\boldsymbol{\psi}(k-1) = [\psi_1 \;\; \psi_2 \;\; \ldots \;\; \psi_{n_\theta}]^{\mathrm{T}},$$

  tomadas até o instante $k-1$.

- Será comum representar um conjunto de equações gerado a partir de um modelo dinâmico como

$$\mathbf{y} = \Psi\hat{\boldsymbol{\theta}} + \boldsymbol{\xi}.$$

- Portanto, a função custo para modelos ARX pode ser expressa de forma detalhada como se segue:

$$J_{\mathrm{MQ}} = \sum_{k=1}^{N} \xi(k \mid k-1, \hat{\boldsymbol{\theta}})^2 = \boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{\xi} = \parallel \boldsymbol{\xi} \parallel^2 .$$

- Portanto, a função custo para modelos ARX pode ser expressa de forma detalhada como se segue:

$$J_{\mathrm{MQ}} = \sum_{k=1}^{N} \xi(k \mid k-1, \hat{\boldsymbol{\theta}})^2 = \boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{\xi} = \parallel \boldsymbol{\xi} \parallel^2 .$$

- Os estimadores MQ e MQP podem ser representados, respectivamente, como

$$\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} = [\Psi^{\mathrm{T}}\Psi]^{-1}\Psi^{\mathrm{T}}\mathbf{y}$$

e

$$\hat{\boldsymbol{\theta}}_{\mathrm{MQP}} = [\Psi^{\mathrm{T}}W\Psi]^{-1}\Psi^{\mathrm{T}}W\mathbf{y}.$$

- Essas equações podem ser rescritas da seguinte forma:

$$\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} = \left[\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\psi}(k-1)\boldsymbol{\psi}^{\mathrm{T}}(k-1)\right]^{-1}\left[\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\psi}(k-1)y(k)\right],$$

- Essas equações podem ser rescritas da seguinte forma:

$$\hat{\boldsymbol{\theta}}_{\text{MQ}} = \left[\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\psi}(k-1)\boldsymbol{\psi}^{\text{T}}(k-1)\right]^{-1}\left[\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\psi}(k-1)y(k)\right],$$

- e

$$\hat{\boldsymbol{\theta}}_{\text{MQP}} = \left[\frac{1}{N}\sum_{k=1}^{N}w_k\boldsymbol{\psi}(k-1)\boldsymbol{\psi}^{\text{T}}(k-1)\right]^{-1}\left[\frac{1}{N}\sum_{k=1}^{N}w_k\boldsymbol{\psi}(k-1)y(k)\right]$$

para $W = \text{diag}[w_1 \ \ldots \ w_N]$ diagonal.

Estimação de parâmetros de um modelo ARX

- Seja o modelo

$$
\begin{aligned}
y(k) &= 1,7649y(k-1) - 0,8027y(k-2) + 0,8661u(k-3) \\
&\quad -0,73578u(k-1) + 0,07513u(k-2) + \xi(k).
\end{aligned}
$$

Estimação de parâmetros de um modelo ARX

- Seja o modelo

$$\begin{aligned} y(k) &= 1,7649y(k-1) - 0,8027y(k-2) + 0,8661u(k-3) \\ &\quad -0,73578u(k-1) + 0,07513u(k-2) + \xi(k). \end{aligned}$$

- Sejam os dados

$$\begin{aligned} y(k) &= [12,2\ \ 11,8\ \ 11,6\ \ 11,6\ \ 11,8\ \ 12,2\ \ 13,0\ \ 14,4\ \ 16,2\ \ 15,8] \\ u(k) &= [2,50\ \ 2,50\ \ 2,50\ \ 2,50\ \ 2,50\ \ 2,23\ \ 2,20\ \ 2,20\ \ 2,21\ \ 2,20]. \end{aligned}$$

Estimação de parâmetros de um modelo ARX

- Seja o modelo

$$\begin{aligned} y(k) &= 1,7649y(k-1) - 0,8027y(k-2) + 0,8661u(k-3) \\ &\quad -0,73578u(k-1) + 0,07513u(k-2) + \xi(k). \end{aligned}$$

- Sejam os dados

$$\begin{aligned} y(k) &= [12,2\ \ 11,8\ \ 11,6\ \ 11,6\ \ 11,8\ \ 12,2\ \ 13,0\ \ 14,4\ \ 16,2\ \ 15,8] \\ u(k) &= [2,50\ \ 2,50\ \ 2,50\ \ 2,50\ \ 2,50\ \ 2,23\ \ 2,20\ \ 2,20\ \ 2,21\ \ 2,20]. \end{aligned}$$

- O vetor de regressores é

$$\boldsymbol{\psi}(k-1) = [y(k-1)\ \ y(k-2)\ \ u(k-3)\ \ u(k-1)\ \ u(k-2)]^{\mathrm{T}},$$

- e gera o seguinte conjunto de restrições ao longo dos dados acima:

$$\begin{bmatrix} 11,6 \\ 11,8 \\ 12,2 \\ 13,0 \\ 14,4 \\ 16,2 \\ 15,8 \end{bmatrix} = \begin{bmatrix} 11,6 & 11,8 & 2,50 & 2,50 & 2,50 \\ 11,6 & 11,6 & 2,50 & 2,50 & 2,50 \\ 11,8 & 11,6 & 2,50 & 2,50 & 2,50 \\ 12,2 & 11,8 & 2,50 & 2,23 & 2,50 \\ 13,0 & 12,2 & 2,50 & 2,20 & 2,23 \\ 14,4 & 13,0 & 2,23 & 2,20 & 2,20 \\ 16,2 & 14,4 & 2,20 & 2,21 & 2,20 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{bmatrix}.$$

- Usando-se a pseudo-inversa, obtém-se

$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \\ \hat{\theta}_4 \\ \hat{\theta}_5 \end{bmatrix} = \begin{bmatrix} 3,4167 \\ -3,4799 \\ 4,7104 \\ 1,0410 \\ -0,6832 \end{bmatrix}.$$

- Usando-se a pseudo-inversa, obtém-se

$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \\ \hat{\theta}_4 \\ \hat{\theta}_5 \end{bmatrix} = \begin{bmatrix} 3,4167 \\ -3,4799 \\ 4,7104 \\ 1,0410 \\ -0,6832 \end{bmatrix}.$$

-

$$\begin{aligned} \boldsymbol{\xi} &= \mathbf{y} - \Psi\hat{\boldsymbol{\theta}} \\ &= [0,3584 \;\; -0,1376 \;\; -0,4209 \;\; -0,0106 \;\; -0,1052 \;\; 0,9467 \;\; -0,6006]^{\mathrm{T}} \end{aligned}$$

$$\begin{aligned} \Psi^{\mathrm{T}}\boldsymbol{\xi} &= [\boldsymbol{\psi}_1^{\mathrm{T}}\boldsymbol{\xi} \;\; \boldsymbol{\psi}_2^{\mathrm{T}}\boldsymbol{\xi} \;\; \boldsymbol{\psi}_3^{\mathrm{T}}\boldsymbol{\xi} \;\; \boldsymbol{\psi}_4^{\mathrm{T}}\boldsymbol{\xi} \;\; \boldsymbol{\psi}_5^{\mathrm{T}}\boldsymbol{\xi}]^{\mathrm{T}} \\ &= [0,1661 \;\; 0,1574 \;\; 0,0306 \;\; 0,0296 \;\; 0,0301]^{\mathrm{T}} \times 10^{-8}. \end{aligned}$$

Estimação de parâmetros de um modelo contínuo no tempo

- Este exemplo é análogo ao anterior, porém o modelo usado é contínuo no tempo.

Estimação de parâmetros de um modelo contínuo no tempo

- Este exemplo é análogo ao anterior, porém o modelo usado é contínuo no tempo.
- Considere o circuito elétrico descrito por:

$$\begin{aligned} v(t) &= LC\ddot{v}_c(t) + RC\dot{v}_c(t) + v_c(t) \\ \frac{1}{C}v_c(t) &= \frac{1}{C}v(t) - L\ddot{v}_c(t) - R\dot{v}_c(t) \end{aligned} \qquad \text{(-56)}$$

Estimação de parâmetros de um modelo contínuo no tempo

- Este exemplo é análogo ao anterior, porém o modelo usado é contínuo no tempo.
- Considere o circuito elétrico descrito por:

$$\begin{aligned} v(t) &= LC\ddot{v}_c(t) + RC\dot{v}_c(t) + v_c(t) \\ \frac{1}{C}v_c(t) &= \frac{1}{C}v(t) - L\ddot{v}_c(t) - R\dot{v}_c(t) \end{aligned} \tag{-57}$$

- o modelo foi simulado $C = 0,8$, $L = 0,5$, $R = 0,2$ e o sinal de entrada é um PRBS.

- O vetor de regressores e o vetor de parâmetros do modelo são, respectivamente

$$\phi(t) = [v(t) - \ddot{v}_c(t) - \dot{v}_c(t)]^T$$

e

$$\theta = \begin{bmatrix} \frac{1}{C} \\ L \\ R \end{bmatrix}$$

- O vetor de regressores e o vetor de parâmetros do modelo são, respectivamente

$$\phi(t) = [v(t) - \ddot{v}_c(t) - \dot{v}_c(t)]^T$$

e

$$\theta = \begin{bmatrix} \frac{1}{C} \\ L \\ R \end{bmatrix}$$

- Obtém-se, então

$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{bmatrix} = \begin{bmatrix} 1,25 \\ 0,5 \\ 0,2 \end{bmatrix}$$

A qualidade dos resultados obtidos é devido a dois fatores:

- Nenhum ruído foi adicionado aos dados

A qualidade dos resultados obtidos é devido a dois fatores:

- Nenhum ruído foi adicionado aos dados
- As derivadas do sinal $v_c(t)$ não foram estimadas, mas calculadas utilizando-se as seguintes equações do circuito:

$$\begin{aligned}
\dot{v}_c(t) &= \frac{1}{C} i(t) \\
\ddot{v}_c(t) &= \frac{1}{C} \frac{di(t)}{dt} \\
\frac{di(t)}{dt} &= -\frac{R}{L} i(t) - \frac{1}{L} v_c(t) + \frac{1}{L} v(t)
\end{aligned}$$

A qualidade dos resultados obtidos é devido a dois fatores:

- Nenhum ruído foi adicionado aos dados
- As derivadas do sinal $v_c(t)$ não foram estimadas, mas calculadas utilizando-se as seguintes equações do circuito:

$$
\begin{aligned}
\dot{v}_c(t) &= \frac{1}{C} i(t) \\
\ddot{v}_c(t) &= \frac{1}{C} \frac{di(t)}{dt} \\
\frac{di(t)}{dt} &= -\frac{R}{L} i(t) - \frac{1}{L} v_c(t) + \frac{1}{L} v(t)
\end{aligned}
$$

A estimação das derivadas é normalmente numérica.