

# 5

## Identificação de modelos

*“All identification is for control and ownership..”*

– Bryant McGill

Identificação, no contexto desta obra, é o processo de construir um modelo para um dado sistema a partir dos registros de pares entrada-saída,  $\{u(t), y(t)\}_{t_0 \leq t \leq t_f}$  ou  $\{u[k], y[k]\}_{0 \leq k_0 \leq k_f}$ .

Em geral, podem ser agrupadas em duas categorias.

1. Identificação paramétrica: tendo-se escolhido a estrutura do modelo, os seus parâmetros são estimados a partir dos dados disponíveis. Por exemplo, se a estrutura escolhida é um sistema LTI de segunda ordem, busca-se estimar  $\xi$  e  $\omega_n$ .
2. Identificação não paramétrica: a partir dos pares entrada-saída, busca-se diretamente um mapa que os relacione. Por exemplo, um algoritmo de aprendizado pode ser utilizado para ajustar uma rede neural artificial que represente esse mapa entrada-saída.

### 5.1 Métodos elementares

Sistemas simples, por exemplo, do tipo LTI de uma entrada e uma saída, podem ser identificados por experimentos simples, como:

- Excitar o processo com sinais  $u(t)$  do tipo degrau e observar a resposta. Caso a resposta se aproxime àquela característica de  $2^a$  ordem, o sobressinal, o tempo de subida e o erro em regime permanente permite determinar  $K$ ,  $\xi$  e  $\omega_n$  de

$$T(s) = \frac{K}{s^2 + 2\xi\omega_n s + \omega_n^2} \quad (5.1)$$

- Excitar o processo com sinais senoidais em uma ampla faixa de frequências e ajustar assíntotas às curvas de Bode.
- Excitar o processo com um ruído branco e obter o gráfico da função de correlação cruzada entrada-saída. Estando  $u(t)$  e  $y(t)$  na condição de estacionaridade no sentido amplo, a resposta impulso do processo é dada por  $h(t) = R_{uy}(t)$ , em que  $R_{uy}(t) = E[u(\tau)y(\tau - t)] \approx \frac{1}{T} \int_0^T u(\tau)y(\tau - t) d\tau$  para  $T$  suficientemente grande. O detalhamento deste método foge ao escopo deste livro e os interessados são referidos aos textos (AGUIRRE, 2015b), (EYKOFF, 1974), (GOODWIN; PAYNE, 2012) e (LJUNG, 1998).

## 5.2 Identificação de modelos de tempo discreto

Neste capítulo são tratados apenas os métodos paramétricos, ou seja, a estrutura do modelo é assumida conhecida e os seus parâmetros são estimados através das medidas das entradas e as correspondentes saídas.

Com a intensificação do uso de equipamentos digitais em controle, a identificação de modelos de tempo discreto tem ganhado cada vez mais espaço na teoria e aplicação de sistemas de controle, competindo com métodos de discretização de modelos obtidos invocando leis físicas.

A leitura deste capítulo requer alguns conhecimentos básicos da teoria de probabilidade e de processos estocásticos que pode ser encontrado em livros textuais como (PAPOULIS, 1984), (GNEDENKO; KHINCHIN, 2013) e (DURRETT, 2019)

1. Dado um espaço de probabilidade  $(\Omega, \mathcal{F}, P)$ , uma variável aleatória real é uma função  $X : \Omega \rightarrow R$  ou  $X(\omega)$  com  $\omega \in \Omega$ . Porém, aqui e como em muitos textos, o argumento  $\omega$  é omitido.
2. O valor esperado de uma variável aleatória  $X$  é definido como

$$E[X] = \int_{\Omega} X(\omega)P(d\omega) \quad (5.2)$$

$$= \int_R \xi f_X(\xi)d\xi \quad (5.3)$$

em que  $f_X$  é a função densidade de probabilidade de  $X$  ( $\xi$  é apenas uma variável de integração).

3. Variáveis aleatórias  $X$  normalmente distribuídas, denotadas por  $X \sim N(\mu, \sigma^2)$ , são as que possuem densidade de probabilidade  $f_X$  expressa
-

por

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (5.4)$$

4. Uma sequência i.i.d. de variáveis aleatórias  $\{X[0], X[1], \dots\}$  é aquela em que  $X[i]$  e  $X[j]$  possuem distribuições idênticas e, além disso,  $X[i]$  e  $X[j]$  são variáveis aleatórias independentes,  $\forall i \neq j$ .
5. Quando as variáveis  $X$  e  $Y$  possuem distribuição normal, a não correlação ( $\text{corr}(X, Y) = 0$ ) implica independência.
6. A notação utilizada para a matriz de covariância é  $\text{cov}(X) = E[(X - E[X])(X - E[X])^T]$
7. Quando duas variáveis aleatórias  $X$  e  $Y$  são correlacionadas, o conhecimento de  $Y$  afeta a probabilidade de  $X$ . Assim, a esperança condicional de  $X$  dado  $Y$  é denotada  $E[X|Y]$  (que é uma nova variável aleatória).

### 5.2.1 Estimador de mínimos quadrados

Neste módulo será apresentado apenas o caso de estimação de parâmetros de uma classe de modelos lineares utilizando o método dos mínimos quadrados.

Considere um sistema SISO descrito por

$$A(q^{-1})y[k] = B(q^{-1})u[k] + w[k] \quad (5.5)$$

em que  $A$  e  $B$  são polinômios em  $q^{-1}$

$$A(q^{-1}) = 1 + a_1q^{-1} + \dots + a_nq^{-n} \quad (5.6)$$

$$B(q^{-1}) = b_0 + b_1q^{-1} + \dots + b_mq^{-m} \quad (5.7)$$

e  $q^{-1}$  representa o operador atrasador unitário,  $y[k]$  é a saída,  $u[k]$  é o sinal de entrada e  $w[k] \sim N(0, 1)$  são ruídos independentes identicamente distribuídos (*i.i.d.*).

Essa expressão pode ser colocada em uma forma mais conveniente

$$y[k] = \begin{bmatrix} y[k-1] & \dots & y[k-n] & u[k] & \dots & u[k-m] \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_0 \\ \vdots \\ b_m \end{bmatrix} + w[k] \quad (5.8)$$

Suponha que os registros dos sinais de saída e da entrada, respectivamente,  $\{y[0], \dots, y[N]\}$  e  $\{u[0], \dots, u[N]\}$  com  $N \gg n$  são fornecidos.

Então, empilhando-se as expressões 5.8 com  $y[N]$  no topo e decrescendo até  $y[n]$ , tem-se que

$$\mathbf{Y} = \begin{bmatrix} y[N] & y[N-1] & \cdots & y[n] \end{bmatrix}^T$$

$$\boldsymbol{\theta} = \begin{bmatrix} a_1 & \cdots & a_n & b_0 & \cdots & b_m \end{bmatrix}^T \quad (5.9)$$

$$\mathbf{W} = \begin{bmatrix} w[N] & w[N-1] & \cdots & w[n] \end{bmatrix}^T \quad (5.10)$$

$$\mathbf{A} = \begin{bmatrix} -y[N-1] & \cdots & -y[N-n] & \left| \begin{array}{ccc} u[N] & \cdots & u[N-m] \\ u[N-1] & \cdots & u[N-m-1] \\ \vdots & & \vdots \\ u[n] & \cdots & u[n-m] \end{array} \right. \\ -y[N-2] & \cdots & -y[N-n-1] & \\ \vdots & & \vdots & \\ -y[n-1] & \cdots & -y[0] & \end{bmatrix} \quad (5.11)$$

tem-se a representação compacta

$$\mathbf{Y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{W}. \quad (5.12)$$

A matriz  $\mathbf{A}$  é  $(N-n+1) \times (n+m+1)$  com  $N-n+1 \gg n+m+1$ , ou seja, o sistema possui muito mais equações do que incógnitas e não haveria como satisfazer todas elas simultaneamente.

As inconsistências são, em geral, devidas aos ruídos  $w[k]$  desconhecidos.

A ideia, portanto, é utilizar o método dos mínimos quadrados, de modo que, quando o comprimento dos registros  $N$  é grande, o efeito do ruído é reduzido.

O valor de  $\boldsymbol{\theta}$  que minimiza  $\|\mathbf{Y} - \mathbf{A}\boldsymbol{\theta}\|^2$ , denotado  $\hat{\boldsymbol{\theta}}$ , deve satisfazer a condição

$$\left. \frac{d}{d\boldsymbol{\theta}} \|\mathbf{Y} - \mathbf{A}\boldsymbol{\theta}\|^2 \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0 \quad (5.13)$$

Desenvolvendo o lado esquerdo dessa condição, obtém-se que

$$\left. \frac{d}{d\boldsymbol{\theta}} (\mathbf{Y} - \mathbf{A}\boldsymbol{\theta})^T (\mathbf{Y} - \mathbf{A}\boldsymbol{\theta}) \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0 \quad (5.14)$$

$$\left. \frac{d}{d\boldsymbol{\theta}} (\mathbf{Y}^T \mathbf{Y} - 2\mathbf{Y}^T \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\theta}^T \mathbf{A}^T \mathbf{A}\boldsymbol{\theta}) \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0 \quad (5.15)$$

$$\left. (-2\mathbf{A}^T \mathbf{Y} + 2\mathbf{A}^T \mathbf{A}\boldsymbol{\theta}) \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0 \quad (5.16)$$

A expressão de  $\hat{\boldsymbol{\theta}}$  que minimiza o erro quadrático é, portanto,

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y} \quad (5.17)$$

Essa é a expressão do EMQ (Estimador de Mínimos Quadrados ou *Least Squares Estimator* - LSE) e a matriz  $\mathbf{A}^\dagger = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$  é conhecida como a pseudoinversa de Moore-Penrose de  $\mathbf{A}$ .

Algumas propriedades do EMQ (LSE):

- Uma estimativa  $\hat{\boldsymbol{\theta}}$  de  $\boldsymbol{\theta}$  é dita ser não polarizada se  $E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$ .

$$E[\hat{\boldsymbol{\theta}}] = E\left[(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y}\right] \quad (5.18)$$

$$= E\left[(\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{A} \boldsymbol{\theta} + \mathbf{W})\right] \quad (5.19)$$

$$= E\left[(\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{A}) \boldsymbol{\theta} + (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{W}\right] \quad (5.20)$$

$$= E\left[(\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{A})\right] \boldsymbol{\theta} + (\mathbf{A}^T \mathbf{A})^{-1} E[\mathbf{W}] \quad (5.21)$$

$$= \boldsymbol{\theta} \quad (5.22)$$

- $cov(\hat{\boldsymbol{\theta}}) = (\mathbf{A}^T \mathbf{A})^{-1}$

Lembrando que  $\hat{\boldsymbol{\theta}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y}$  e  $\mathbf{Y} = \mathbf{A} \boldsymbol{\theta} + \mathbf{W}$ , pode-se escrever

$$\begin{aligned} E[(\hat{\boldsymbol{\theta}} - E[\hat{\boldsymbol{\theta}}])(\hat{\boldsymbol{\theta}} - E[\hat{\boldsymbol{\theta}}])^T] &= E[(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{W}^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1}] \\ &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T E[\mathbf{W} \mathbf{W}^T] \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \\ &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \\ &= (\mathbf{A}^T \mathbf{A})^{-1} \end{aligned} \quad (5.23)$$

- Teorema de Gauss-Markov:

A estimativa obtida pelo MQE,  $\hat{\boldsymbol{\theta}}$ , é ótima na classe de estimadores lineares não polarizados (ou se  $\check{\boldsymbol{\theta}} = \check{\mathbf{A}} \mathbf{Y}$  tal que  $E[\check{\boldsymbol{\theta}}] = \boldsymbol{\theta}, \forall \boldsymbol{\theta}$ , então  $cov(\hat{\boldsymbol{\theta}}) \leq cov(\check{\boldsymbol{\theta}})$ ).

Note, inicialmente, que

$$\boldsymbol{\theta} = E[\check{\boldsymbol{\theta}}] = E[\check{\mathbf{A}} \mathbf{Y}] \quad (5.24)$$

$$= E[\check{\mathbf{A}} (\mathbf{A} \boldsymbol{\theta} + \mathbf{W})] \quad (5.25)$$

$$= \check{\mathbf{A}} \mathbf{A} \boldsymbol{\theta} + E[\mathbf{W}] \quad (5.26)$$

$$= \check{\mathbf{A}} \mathbf{A} \boldsymbol{\theta} \quad (5.27)$$

Em vista de  $\boldsymbol{\theta} = \check{\mathbf{A}}\mathbf{A}\boldsymbol{\theta}$ , conclui-se que  $\check{\mathbf{A}}\mathbf{A} = \mathbf{I}$  e, por outro lado,

$$\text{cov}(\check{\mathbf{A}}) = E[(\check{\boldsymbol{\theta}} - E[\check{\boldsymbol{\theta}}])(\check{\boldsymbol{\theta}} - E[\check{\boldsymbol{\theta}}])^T] \quad (5.28)$$

$$= E[(\check{\mathbf{A}}\mathbf{Y} - \boldsymbol{\theta})(\check{\mathbf{A}}\mathbf{Y} - \boldsymbol{\theta})^T] \quad (5.29)$$

$$= E[\check{\mathbf{A}}\mathbf{W}\mathbf{W}^T\check{\mathbf{A}}^T] \quad (5.30)$$

$$= \check{\mathbf{A}}\check{\mathbf{A}}^T \quad (5.31)$$

uma vez que

$$\check{\mathbf{A}}\mathbf{Y} - \boldsymbol{\theta} = \check{\mathbf{A}}(\mathbf{A}\boldsymbol{\theta} + \mathbf{W}) - \boldsymbol{\theta} \quad (5.32)$$

$$= (\check{\mathbf{A}}\mathbf{A})\boldsymbol{\theta} + \check{\mathbf{A}}\mathbf{W} - \boldsymbol{\theta} \quad (5.33)$$

$$= \boldsymbol{\theta} + \check{\mathbf{A}}\mathbf{W} - \boldsymbol{\theta} \quad (\text{usando } \check{\mathbf{A}}\mathbf{A} = \mathbf{I}) \quad (5.34)$$

$$= \check{\mathbf{A}}\mathbf{W}. \quad (5.35)$$

Seja agora  $\tilde{\mathbf{A}}$  definido por  $\tilde{\mathbf{A}} = \check{\mathbf{A}} - (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$  e note que  $\tilde{\mathbf{A}}\mathbf{A} = \mathbf{0}$  pois  $\check{\mathbf{A}}\mathbf{A} = \mathbf{I}$ .

Lembrando que havia sido mostrado que  $\text{cov}(\check{\mathbf{A}}) = \check{\mathbf{A}}\check{\mathbf{A}}^T$ , tem-se que

$$\text{cov}(\check{\mathbf{A}}) = \check{\mathbf{A}}\check{\mathbf{A}}^T \quad (5.36)$$

$$= [\tilde{\mathbf{A}} + (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T][\tilde{\mathbf{A}} + (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T]^T \quad (5.37)$$

$$= \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T + (\mathbf{A}^T\mathbf{A})^{-1} \quad (5.38)$$

$$= \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T + \text{cov}(\hat{\boldsymbol{\theta}}) \quad (5.39)$$

Logo,  $\text{cov}(\check{\boldsymbol{\theta}}) \geq \text{cov}(\hat{\boldsymbol{\theta}})$ , pois  $\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T \geq \mathbf{0}$ .

- O teorema de Gauss-Markov permite concluir que a estimativa  $\hat{\boldsymbol{\theta}}$  é BLUE - *Best Linear Unbiased Estimator*.

### 5.2.2 Regiões de confiança para EMQ

A identificação paramétrica de modelos resulta em incertezas quando há presença de ruído.

Assim, é importante que se estime a região de confiança das estimativas com algum grau de significância.

Por exemplo, saber que um parâmetro possui dispersão de  $10.0 \pm 0.1$  permite a realização de um projeto melhor do que no caso de  $10.0 \pm 1.0$ .

Assuma que se está utilizando o método do EMQ, caracterizado por

$$\mathbf{Y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{W} \quad (5.40)$$

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y} \quad (5.41)$$

$$\widehat{\mathbf{W}} = \mathbf{Y} - \mathbf{A}\hat{\boldsymbol{\theta}} \quad (5.42)$$

em que  $\mathbf{W} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ .

### Caso com $\sigma$ conhecido

Sabendo-se que  $\text{cov}(\hat{\theta}) = (\mathbf{A}^T \mathbf{A})^{-1}$ , tem-se

$$(\hat{\theta} - \theta) \sim N(\mathbf{0}, \sigma^2 (\mathbf{A}^T \mathbf{A})^{-1}) \quad (5.43)$$

Logo, adotando se a faixa de significância de 0.95, e denotando  $\mathbf{P} = (\mathbf{A}^T \mathbf{A})^{-1}$ , tem-se

$$\theta_i \in (\hat{\theta}_i - 2\sigma\sqrt{p_{ii}}, \hat{\theta}_i + 2\sigma\sqrt{p_{ii}}) \quad (5.44)$$

### Caso com $\sigma$ desconhecido

Agora, não somente o  $\theta$ , mas também  $\sigma^2$  deve ser estimado:

$$\hat{\sigma}^2 = \frac{1}{N-m} \sum_{k=1}^{N-m+1} \hat{w}[k]^2 \quad (5.45)$$

Necessita-se, ainda, dos seguintes resultados auxiliares ([WITTE; WITTE, 2017](#)):

- “Se  $\{u[k]\}_{k=1,2,\dots,\bar{N}}$  é tal que  $u[k] \sim N(0, 1)$ , então  $\sum_{k=1}^{\bar{N}} u[k] \sim \chi_{\bar{N}}^2$ ”
- “Se  $u \sim N(0, 1)$  e  $v \sim \chi_{\bar{N}}^2$ , então

$$\frac{u}{\sqrt{\frac{v}{\bar{N}}}} \quad (5.46)$$

tem distribuição  $t$  de estudante com  $N$  graus de liberdade ( $t_N$ )”

Faça-se a associação

$$u = \frac{\hat{\theta}_i - \theta_i}{\sqrt{p_{ii}}} \quad \text{e} \quad v = \frac{1}{\sigma^2} \sum_{k=1}^{N-m+1} \hat{w}[k]^2 \quad (5.47)$$

Nessas condições,  $u \sim N(0, 1)$  e  $v \sim \chi_{N-m}^2$  e

$$\frac{u}{\sqrt{\frac{v}{N-m}}} = \frac{\hat{\theta}_i - \theta_i}{\sigma\sqrt{p_{ii}}} \frac{1}{\sqrt{\frac{1}{\sigma^2} \sum_{k=1}^{N-m+1} \hat{w}[k]^2}} \quad (5.48)$$

$$= \frac{\hat{\theta}_i - \theta_i}{\sqrt{p_{ii}}} \frac{1}{\sqrt{\sum_{k=1}^{N-m+1} \hat{w}[k]^2}} \quad (5.49)$$

$$= \frac{\hat{\theta}_i - \theta_i}{\sqrt{p_{ii}\hat{\sigma}^2}} \sim t_{N-m} \quad (5.50)$$

Consultando-se tabelas estatísticas, pode-se obter  $k_\alpha$  correspondente à região de confiança com nível de significância  $\alpha$  da distribuição  $t_{N-m}$ .

O valor verdadeiro estaria com significância  $\alpha$  em

$$\theta_i = \left( \hat{\theta}_i - k_\alpha \sqrt{p_{ii} \hat{\sigma}}, \hat{\theta}_i + k_\alpha \sqrt{p_{ii} \hat{\sigma}} \right) \quad (5.51)$$

Observação: Existem outros métodos de estimação de parâmetros, tais como o da máxima verosimilhança (*maximum likelihood*) e MAP (máximo a priori).

### 5.2.3 Mínimos quadrados recursivos

A expressão  $(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$  não é conveniente pois a estimativa  $\hat{\theta}$  no instante  $k+1$  teria que ser recalculada usando todos os pares  $(y[i], u[i])$  para  $i = 0, \dots, k+1$ .

Se  $\hat{\theta}[k]$  é a estimativa com as informações  $\{y[0], \dots, y[k]\}$  e  $\{u[0], \dots, u[k]\}$ , a ideia é obter  $\hat{\theta}[k+1]$  a partir de  $\hat{\theta}[k]$ ,  $y[k+1]$ ,  $u[k+1]$ .

Seja a matriz  $\mathbf{P}_k$  definida por

$$\mathbf{P}_k = \left( \mathbf{A}_k^T \mathbf{A}_k \right)^{-1} \quad (5.52)$$

em que  $\mathbf{A}_k$  é construída a partir de  $\{y[0], \dots, y[k]\}$  e  $\{u[0], \dots, u[k]\}$ .

A matriz  $\mathbf{A}_{k+1}$  possui uma linha a mais do que  $\mathbf{A}_k$ , ou seja,

$$\mathbf{A}_{k+1} = \begin{bmatrix} \mathbf{A}_k \\ \mathbf{a}_{k+1}^T \end{bmatrix} \quad (5.53)$$

Analogamente, agrega-se a nova medida  $y[k+1]$  ao vetor  $\mathbf{Y}_k$

$$\mathbf{Y}_{k+1} = \begin{bmatrix} \mathbf{Y}_k \\ y[k+1] \end{bmatrix} \quad (5.54)$$

#### Lema de inversão de matrizes:

Sejam  $\mathbf{M}_{n \times n}$ ,  $\mathbf{b}_{n \times 1}$  e  $\mathbf{c}_{n \times 1}$  tais que  $\det(\mathbf{M} + \mathbf{b}\mathbf{c}^T) \neq 0$  e  $1 + \mathbf{c}^T \mathbf{A}^{-1} \mathbf{b} \neq 0$ .

Nestas condições pode-se mostrar que

$$\left( \mathbf{M} + \mathbf{b}\mathbf{c}^T \right)^{-1} = \mathbf{M}^{-1} - \left( 1 + \mathbf{c}^T \mathbf{M}^{-1} \mathbf{b} \right)^{-1} \mathbf{M}^{-1} \mathbf{b}\mathbf{c}^T \mathbf{M}^{-1} \quad (5.55)$$

A ideia é utilizar esse lema para reescrever  $\left( \mathbf{A}_{k+1}^T \mathbf{A}_{k+1} \right)^{-1}$

A matriz de covariância do erro é dada por

$$\mathbf{P}_{k+1} = \left( \mathbf{A}_{k+1}^T \mathbf{A}_{k+1} \right)^{-1} \quad (5.56)$$

$$= \left( \begin{bmatrix} \mathbf{A}_k^T & \mathbf{a}_{k+1} \end{bmatrix} \begin{bmatrix} \mathbf{A}_k \\ \mathbf{a}_{k+1}^T \end{bmatrix} \right)^{-1} \quad (5.57)$$

$$= \left( \mathbf{A}_k^T \mathbf{A}_k + \mathbf{a}_{k+1} \mathbf{a}_{k+1}^T \right)^{-1} \quad (5.58)$$

$$= \left[ \mathbf{I} - \frac{\mathbf{P}_k \mathbf{a}_{k+1} \mathbf{a}_{k+1}^T}{\left( 1 + \mathbf{a}_{k+1}^T \mathbf{P}_k \mathbf{a}_{k+1} \right)} \right] \mathbf{P}_k \quad (5.59)$$

uma vez que  $\mathbf{a}_{k+1}^T \mathbf{P}_k \mathbf{a}_{k+1}$  é um escalar.

Por outro lado,

$$\hat{\theta}[k+1] = \left( \mathbf{A}_{k+1}^T \mathbf{A}_{k+1} \right)^{-1} \mathbf{A}_{k+1}^T \mathbf{Y}_{k+1} \quad (5.60)$$

$$= \left( \mathbf{A}_{k+1}^T \mathbf{A}_{k+1} \right)^{-1} \begin{bmatrix} \mathbf{A}_k^T & \mathbf{a}_{k+1} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_k \\ y[k+1] \end{bmatrix} \quad (5.61)$$

$$= \left( \mathbf{A}_k^T \mathbf{A}_k + \mathbf{a}_{k+1} \mathbf{a}_{k+1}^T \right)^{-1} \begin{bmatrix} \mathbf{A}_k^T & \mathbf{a}_{k+1} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_k \\ y[k+1] \end{bmatrix} \quad (5.62)$$

$$= \left[ \mathbf{I} - \frac{\mathbf{P}_k \mathbf{a}_{k+1} \mathbf{a}_{k+1}^T}{\left( 1 + \mathbf{a}_{k+1}^T \mathbf{P}_k \mathbf{a}_{k+1} \right)} \right] \mathbf{P}_k \left( \mathbf{A}_k^T \mathbf{Y}_k + \mathbf{a}_{k+1} y[k+1] \right) \quad (5.63)$$

$$= \left[ \mathbf{I} - \frac{\mathbf{P}_k \mathbf{a}_{k+1} \mathbf{a}_{k+1}^T}{\left( 1 + \mathbf{a}_{k+1}^T \mathbf{P}_k \mathbf{a}_{k+1} \right)} \right] \left( \hat{\theta}[k] + \mathbf{P}_k \mathbf{a}_{k+1} y[k+1] \right) \quad (5.64)$$

$$= \hat{\theta}[k] + \frac{\mathbf{P}_k \mathbf{a}_{k+1}}{\left( 1 + \mathbf{a}_{k+1}^T \mathbf{P}_k \mathbf{a}_{k+1} \right)} \left( y[k+1] - \mathbf{a}_{k+1}^T \hat{\theta}[k] \right) \quad (5.65)$$

Para a última passagem, usou-se a seguinte simplificação

$$\left[ \mathbf{I} - \frac{\mathbf{P}_k \mathbf{a}_{k+1} \mathbf{a}_{k+1}^T}{\left( 1 + \mathbf{a}_{k+1}^T \mathbf{P}_k \mathbf{a}_{k+1} \right)} \right] \mathbf{P}_k \mathbf{a}_{k+1} y[k+1] = \mathbf{P}_k \mathbf{a}_{k+1} y[k+1] \quad (5.66)$$

já que

$$\overbrace{\mathbf{a}_{k+1}^T \mathbf{P}_k \mathbf{a}_{k+1}}^{\text{escalar}} \mathbf{P}_k \mathbf{a}_{k+1} y[k+1] = \mathbf{P}_k \mathbf{a}_{k+1} \overbrace{\mathbf{a}_{k+1}^T \mathbf{P}_k \mathbf{a}_{k+1}}^{\text{escalar}} y[k+1] \quad (5.67)$$

Definindo-se

$$\mathbf{K}_{k+1} = \mathbf{P}_k \mathbf{A}_{k+1} \left( 1 + \mathbf{a}_{k+1}^T \mathbf{P}_k \mathbf{a}_{k+1} \right)^{-1} \quad (5.68)$$

nota-se que a expressão do EMQ recursivo é a mesma do filtro de Kalman.

$$\hat{\boldsymbol{\theta}}[k+1] = \hat{\boldsymbol{\theta}}[k] + \mathbf{K}_{k+1} \overbrace{\left( y[k+1] - \mathbf{a}_{k+1}^T \hat{\boldsymbol{\theta}}[k] \right)}^{\text{inovação}} \quad (5.69)$$

A matriz de covariância do erro é dada recursivamente por

$$\mathbf{P}_{k+1} = \left[ \mathbf{I} - \mathbf{K}_{k+1} \mathbf{A}_{k+1}^T \right] \mathbf{P}_k \quad (5.70)$$

Observação:

- Modelos tendem a variar com o tempo, mesmo que lentamente, de modo que dados “antigos” podem não contribuir para melhorar a estimação.
- Uma solução é “esquecer” as medidas antigas utilizando o *forgetting factor*  $0 < \lambda < 1$ :

$$\hat{\boldsymbol{\theta}}[k+1] = \hat{\boldsymbol{\theta}}[k] + \frac{\mathbf{P}_k \mathbf{a}_{k+1}}{\left( 1 + \lambda \mathbf{a}_{k+1}^T \mathbf{P}_k \mathbf{a}_{k+1} \right)} \left( y[k+1] - \mathbf{a}_{k+1}^T \hat{\boldsymbol{\theta}}[k] \right) \quad (5.71)$$

### Exemplo

Considere o modelo de tempo discreto

$$\mathbf{x}[k+1] = \begin{bmatrix} 0 & 0 & 0.12 \\ 1 & 0 & -0.74 \\ 0 & 1 & 1.5 \end{bmatrix} \mathbf{x}[k] + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u \quad (5.72)$$

$$y[k] = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \mathbf{x}[k] \quad (5.73)$$

cuja representação na forma ARMA é

$$y[k+1] = -1.50y[k] + 0.74y[k-1] - 0.12y[k-2] + 1.00u[k-2] \quad (5.74)$$

Aplicando-se o método de identificação por mínimos quadrados recursivos, obtém-se o gráfico da figura 5.1.

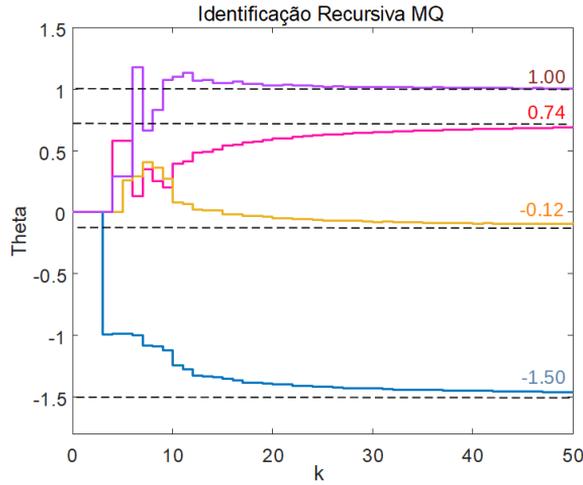


Figura 5.1: Identificação pelo método de mínimos quadrados recursivos.

### 5.3 Identificação pelo método do subespaço

A partir de medidas de sinais de entrada  $u[k]$  e de saída  $y[k]$ , deseja-se estimar o modelo SISO no espaço de estados  $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$

$$\mathbf{x}[k+1] = \mathbf{A}\mathbf{x}[k] + \mathbf{B}u[k] \quad (5.75)$$

$$y[k] = \mathbf{C}\mathbf{x}[k] + \mathbf{D}u[k] \quad (5.76)$$

em que  $\mathbf{x}[k]$  é  $n \times 1$ .

Supondo  $\mathbf{x}[k]$  e  $\{u[k]\}$  conhecidos, pode-se escrever

$$\mathbf{x}[k+1] = \mathbf{A}\mathbf{x}[k] + \mathbf{B}u[k] \quad (5.77)$$

$$\begin{aligned} \mathbf{x}[k+2] &= \mathbf{A}\mathbf{x}[k+1] + \mathbf{B}u[k+1] \\ &= \mathbf{A}^2\mathbf{x}[k] + \mathbf{A}\mathbf{B}u[k] + \mathbf{B}u[k+1] \end{aligned} \quad (5.78)$$

$$\begin{aligned} \mathbf{x}[k+3] &= \mathbf{A}\mathbf{x}[k+2] + \mathbf{B}u[k+2] \\ &= \mathbf{A}^3\mathbf{x}[k] + \mathbf{A}^2\mathbf{B}u[k] + \mathbf{A}\mathbf{B}u[k+1] + \mathbf{B}u[k+2] \end{aligned} \quad (5.79)$$

$\vdots$

$$\mathbf{x}[k+N] = \mathbf{A}^N\mathbf{x}[k] + \sum_{i=0}^{N-1} \mathbf{A}^{N-1-i}\mathbf{B}u[k+i] \quad (5.80)$$

Portanto, multiplicando-se por  $\mathbf{C}$  à esquerda, tem-se que

$$y[k] = \mathbf{C}\mathbf{x}[k] + \mathbf{D}u[k] \quad (5.81)$$

$$y[k+1] = \mathbf{C}\mathbf{A}\mathbf{x}[k] + \mathbf{C}\mathbf{B}u[k] + \mathbf{D}u[k+1] \quad (5.82)$$

$$y[k+2] = \mathbf{C}\mathbf{A}^2\mathbf{x}[k] + \mathbf{C}\mathbf{A}\mathbf{B}u[k] + \mathbf{C}\mathbf{B}u[k+1] + \mathbf{D}u[k+2] \quad (5.83)$$

$$y[k+3] = \mathbf{C}\mathbf{A}^3\mathbf{x}[k] + \mathbf{C}\mathbf{A}^2\mathbf{B}u[k] + \mathbf{C}\mathbf{A}\mathbf{B}u[k+1] + \mathbf{C}\mathbf{B}u[k+2] + \mathbf{D}u[k+3] \quad (5.84)$$

⋮

$$y[k+N] = \mathbf{C}\mathbf{A}^N\mathbf{x}[k] + \sum_{i=0}^{N-1} \mathbf{C}\mathbf{A}^{N-1-i}\mathbf{B}u[k+i] + \mathbf{D}u[k+N] \quad (5.85)$$

Para simplificar a notação definem-se  $\mathbf{Y}_N$ ,  $\mathbf{V}_N$ ,  $\mathbf{H}_N$  e  $\mathbf{U}_N$

$$\underbrace{\begin{bmatrix} y[k] \\ y[k+1] \\ y[k+2] \\ \vdots \\ y[k+N] \end{bmatrix}}_{\mathbf{Y}_N} = \underbrace{\begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \mathbf{C}\mathbf{A}^2 \\ \vdots \\ \mathbf{C}\mathbf{A}^N \end{bmatrix}}_{\mathbf{V}_N} \mathbf{x}[k] + \underbrace{\begin{bmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{C}\mathbf{B} & \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{C}\mathbf{A}\mathbf{B} & \mathbf{C}\mathbf{B} & \mathbf{D} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}\mathbf{A}^{N-1}\mathbf{B} & \mathbf{C}\mathbf{A}^{N-2}\mathbf{B} & \mathbf{C}\mathbf{A}^{N-3}\mathbf{B} & \cdots & \mathbf{D} \end{bmatrix}}_{\mathbf{H}_N} \underbrace{\begin{bmatrix} u[k] \\ u[k+1] \\ u[k+2] \\ \vdots \\ u[k+N] \end{bmatrix}}_{\mathbf{U}_N} \quad (5.86)$$

de modo que em notação compacta, para cada instante  $k$ , tem-se

$$\mathbf{Y}_N = \mathbf{V}_N \mathbf{x}[k] + \mathbf{H}_N \mathbf{U}_N \quad (5.87)$$

A matriz  $\mathbf{V}_N$  possui uma estrutura que lembra a matriz de observabilidade e os elementos da matriz  $\mathbf{H}_N$  contém respostas impulso  $\mathbf{C}\mathbf{A}\mathbf{B}$ .

Os parâmetros a serem estimados são  $\mathbf{V}_N$  e  $\mathbf{H}_N$ .

A ideia é coletar dados desde o instante  $k$  até  $k+N$ , com  $N$  suficientemente grande para que se possa aplicar, por exemplo, o método dos mínimos quadrados.

Agrupando-se lado a lado os  $\mathbf{Y}_N$  correspondentes a  $\mathbf{x}[k]$  e  $\mathbf{U}_N$  para vários instantes  $k$ , podem ser construídas as matrizes de dados  $\mathbf{Y}$  e  $\mathbf{U}$ , na forma

$$\mathbf{Y} = \begin{bmatrix} y[k] & y[k+1] & \cdots & y[k+M] \\ y[k+1] & y[k+2] & \cdots & y[k+M+1] \\ y[k+2] & y[k+3] & \cdots & y[k+M+2] \\ \vdots & \vdots & \ddots & \vdots \\ y[k+N] & y[k+N+1] & \cdots & y[k+M+N] \end{bmatrix} \quad (5.88)$$

$$\mathbf{U} = \begin{bmatrix} u[k] & u[k+1] & \cdots & u[k+M] \\ u[k+1] & u[k+2] & \cdots & u[k+M+1] \\ u[k+2] & u[k+3] & \cdots & u[k+M+2] \\ \vdots & \vdots & \ddots & \vdots \\ u[k+N] & u[k+N+1] & \cdots & u[k+M+N] \end{bmatrix} \quad (5.89)$$

Definindo-se

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}[k] & \mathbf{x}[k+1] & \cdots & \mathbf{x}[k+M] \end{bmatrix} \quad (5.90)$$

pode-se escrever utilizando-se as definições de  $\mathbf{Y}$  e  $\mathbf{U}$

$$\mathbf{Y} = \mathbf{V}_N \mathbf{x} + \mathbf{H}_N \mathbf{U} \quad (5.91)$$

Se  $\mathbf{x}$  é medido, então a obtenção de  $\mathbf{V}_N$  e  $\mathbf{H}_N$  a partir de 5.91 poderia ser feita diretamente pelo método de mínimos quadrados.

Porém, se  $\mathbf{x}$  não é disponível, então o problema de estimar  $\mathbf{V}_N$  e  $\mathbf{H}_N$  a partir de 5.91 é mais intrincado e foge ao escopo deste livro.

Um tratamento sistemático desse caso pode ser encontrado em Qin, S. J. An Overview of Subspace Identification. *Comp. and Chemical Engineering.*, v. 30, 2006, p. 1502-1513.

**Embora existam métodos não paramétricos, como o dos casos que empregam redes neurais artificiais, que permitem casar modelos (*model matching*) à relação entre os registros de entradas e saídas, esta seção buscou a identificação de parâmetros de modelos na forma de equações a diferenças. É importante ressaltar que uma identificação acurada e precisa requer adequada quantidade de informações também acuradas e precisas.**

## 5.4 Algumas notas históricas

Em 1809, Johann Carl Friedrich Gauss (1777-1855) publicou um trabalho que apresentava um tratamento sobre o método dos mínimos quadrados sob a hipó-

tese de erros normalmente distribuídos. (Gauss, J. C. F. *Teoria de movimento de corpos celestes movendo em secções cônicas ao redor do sol* (em latim). Friedrich Perthes e I. H. Besser, Hamburgo, 1809).

Porém, em 1805, Adrien-Marie Legendre (1752-1833) já tinha publicado um livro em que o método de ajustar curvas a uma série de dados era apresentado no apêndice (Legendre, A. M. *Novos métodos para a determinação de órbitas de cometas* (em francês). F. Didot, 1805).

Um fato curioso é que o retrato de Legendre que tem circulado em variados meios de publicação é, em verdade, do político Louis Legendre (1752-1797), não sendo do matemático Adrien-Marie Legendre.

Em 1821, Gauss apresentou uma prova do teorema de Gauss-Markov que mostra a otimalidade do método de mínimos quadrados, conforme a primeira parte da sua obra: Gauss, C. F. *A teoria da combinação menos afetada por erros de observação* (em latim). Societas Regia Scientiarum Gottingensis, 1821.

O teorema foi redescoberto em 1900 por Andrey Andreyevich Markov (1856-1922), famoso pelos processos de Markov (Markov, A. A. *Extensão da lei dos grandes números para quantidades dependentes* (em russo). *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete*, v.15, 1906, pp. 135-156).

Em 1933, Andrey Nikolaevich Kolmogorov (1903-1987) publicou o livro que estabeleceu uma formulação axiomática para a Teoria de Probabilidades (Kolmogorov, A. *Foundations of the probability theory* (em alemão). Julius Springer, 1933).

Uma excelente referência na língua portuguesa é (AGUIRRE, 2015b) e, na língua inglesa, (LJUNG; SÖDERSTROM, 1984).



Markov



Kolmogorov

Gauss

---

## 5.5 Exercícios

### 5.5.1 Exercício: Ajuste de retas usando o método de mínimos quadrados

Assuma que em algum experimento de laboratório é sabido que

$$y = ax + b \quad (5.92)$$

em que  $x \in \mathbb{R}$  e  $y \in \mathbb{R}$ .

Repetindo-se o experimento  $N$  vezes, obtém-se um conjunto de pares entrada-saída

$$\mathcal{C} = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\} \quad (5.93)$$

Devido a erros de medida, cada par  $(x_i, y_i)$  possui uma incerteza aditiva  $w_i$ .

Ao invés de marcar os pontos em um papel milimetrado e ajustar a reta que mais aproxima esses pontos usando uma régua, mostrar como o método dos mínimos quadrados pode ser utilizado para obter as estimativas  $\hat{a}$  e  $\hat{b}$ .

Observação: O conjunto de medidas satisfaz:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}}_{\mathbf{Y}} = \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} \quad (5.94)$$

### 5.5.2 Exercício: Identificação paramétrica pela resposta degrau

Um certo sistema linear invariante no tempo, de  $2^a$  ordem, sem zeros, foi ensaiado no laboratório injetando-se na entrada um sinal  $u(t)$  do tipo degrau unitário. A saída  $y(t)$  correspondente encontra-se na figura 5.2. Esboçar a saída  $y(t)$ , em regime permanente ( $t \uparrow \infty$ ), quando o sinal de excitação  $u(t)$  é agora de forma senoidal,  $u(t) = 5 \sin(2t)$ .

### 5.5.3 Exercício: Identificação paramétrica pelas curvas de Bode

Excitou-se um sistema LTI de segunda ordem com um gerador de sinais senoidais e com o auxílio de um osciloscópio, e traçaram-se as curvas de Bode da figura 5.3. Apresentar modelos desses sistemas na forma de uma função de transferência.

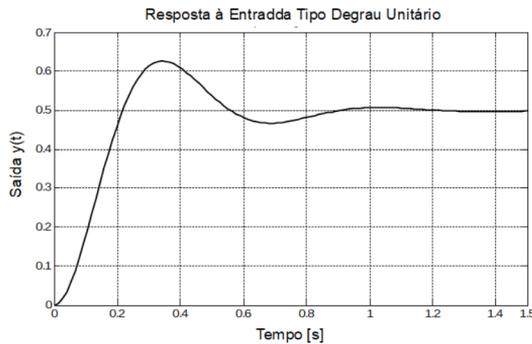


Figura 5.2: Resposta degrau de um sistema de segunda ordem.

#### 5.5.4 Exercício: Método dos mínimos quadrados

Aplicou-se um sinal  $u[k]$  em um sistema de tempo discreto LTI de segunda ordem e registrou-se o sinal de saída  $y[k]$ . O registro dos sinais encontra-se na tabela a seguir.

Tabela 5.1: Registros de pares entrada-saída de um sistema LTI de segunda ordem

k	$u[k]$	$y[k]$
0	1.000000	0.700000
1	0.830258	-0.439742
2	0.933023	0.791610
3	1.110557	-0.094803
4	1.207818	0.519340
5	0.776472	-0.370270
6	1.252132	1.115390
7	1.132029	-0.609886
8	0.986427	0.756670
9	0.960956	-0.181629
10	0.956479	0.448792
11	0.939378	0.021957
12	1.004609	0.376130

Identificar os parâmetros do modelo na forma ARMAX utilizando o método da batelada (usando a pseudoinversa de Moore-Penrose).

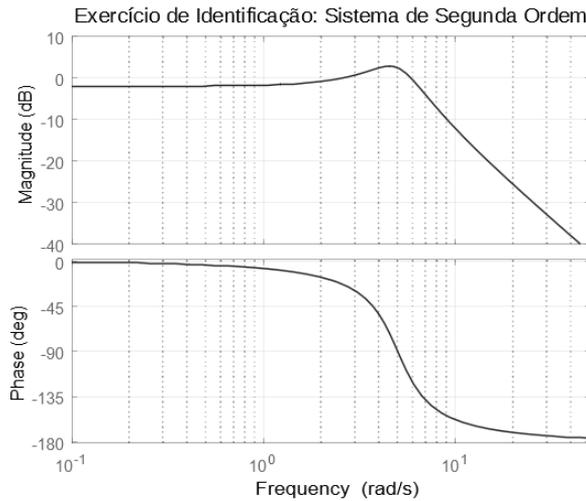


Figura 5.3: Curvas de Bode - Sistema LTI de 2ª ordem.

### 5.5.5 Exercício: Identificação

Excitou-se um sistema LTI de segunda ordem com um gerador de sinais senoidais e com o auxílio de um osciloscópio, traçou-se as curvas de Bode da figura 5.4. Apresentar modelos desses sistemas na forma de uma função de transferência.

### 5.5.6 Exercício: Método do EMQ recursivo

Repetir o exercício anterior, utilizando-se, porém, o método dos mínimos quadrados recursivo.

### 5.5.7 Exercício: Região de confiança da estimativa

Estimar a região de confiança das estimativas obtidas no exercício anterior.

### 5.5.8 Algumas observações

O campo de estudos de métodos de identificação é extenso e vários conceitos e técnicas importantes estão omitidas neste capítulo.

Por outro lado é comum que, na prática, a modelagem e identificação ocupem recursos e esforços substanciais em um projeto de sistema de controle.

Em alguns projetos de maior porte, o dispêndio na obtenção de um modelo

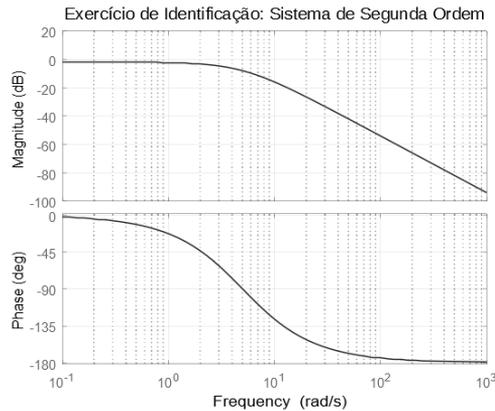


Figura 5.4: Curvas de Bode - Sistema LTI de 2<sup>a</sup> ordem.

confiável pode vir a significar uma fatia substancial do orçamento.

Por exemplo, no projeto de aeronaves, há eventuais gastos com testes em túneis de vento ou CFD extensivo, construção de *ironbirds*, construção de modelos em escala, construção de protótipos instrumentados e ensaios em voo, entre outras possibilidades.

Recomenda-se, portanto, àqueles que se interessam por esse tópico, a leitura de obras tais como (AGUIRRE, 2015b), (COELHO; COELHO, 2016), (GODWIN; PAYNE, 2012) e (LJUNG, 1998).