

# CAPÍTULO 10

## ESTIMAÇÃO DE PARÂMETROS PELO MÉTODO DOS MÍNIMOS QUADRADOS

Suponha que um conjunto de modelos candidatos tenha sido escolhido e tenha sido parametrizado como uma estrutura de modelo, usando um vetor de parâmetros  $\hat{\theta}$ . A busca pelo melhor modelo dentro do conjunto se torna um problema de determinar ou estimar  $\hat{\theta}$ .

### 10.1 COLOCAÇÃO DO PROBLEMA DE ESTIMAR PARÂMETROS

Suponha que já se tenha selecionado uma certa estrutura de modelo  $M$ , com modelos particulares  $M(\hat{\theta})$  parametrizados usando o vetor de parâmetros  $\hat{\theta}$ . O conjunto de modelos assim definido é dado por (LJUNG, 1999):

$$M^* = \{M(\hat{\theta})/\hat{\theta}\}$$

Lembre-se que cada modelo representa uma forma de prever futuras saídas. O preditor um passo à frente poderia ser um filtro linear:

$$M(\hat{\theta}) = \hat{y}(t/t-1, \hat{\theta}) = W_y(q, \hat{\theta}) \cdot y(t) + W_u(q, \hat{\theta}) \cdot u(t) \quad (10.1)$$

A Expressão (10.1) corresponde à predição um passo à frente para um sistema descrito por:

$$y(t) = G(q, \hat{\theta}) \cdot u(t) + H(q, \hat{\theta}) \cdot e(t) \quad (10.2)$$

sendo que:

$$W_y(q, \hat{\theta}) = [1 - H^{-1}(q, \hat{\theta})] \quad W_u(q, \hat{\theta}) = H^{-1}(q, \hat{\theta}) \cdot G(q, \hat{\theta})$$

O modelo  $M(\hat{\theta})$  pode também conter hipóteses acerca do caráter dos erros de predição associados, tais como suas variâncias  $\sigma^2(\hat{\theta})$  ou sua função densidade de probabilidade  $f_e(x, \hat{\theta})$ .

Supõe-se dispor de uma batelada de dados do sistema:

$$Z^N = [y(1), u(1), y(2), u(2), \dots, y(N), u(N)]$$

O problema que se enfrenta agora é decidir como usar as informações contidas em  $Z^N$  para selecionar um valor adequado para o vetor de parâmetros  $\hat{\theta}_N$  e daí um membro adequado  $M(\hat{\theta}_N)$  no conjunto  $M^*$ . Essa busca corresponde ao **método de estimação de parâmetros** (LJUNG, 1999).

## 10.2 DESCRIÇÃO DO MÉTODO DO ERRO DE PREDIÇÃO

A ideia aqui é encontrar um teste através do qual a habilidade dos diferentes modelos para “descrever” os dados observados possa ser avaliada. A essência de um modelo é seu aspecto preditivo e pode-se julgar seu desempenho nesse sentido. Seja então o erro de predição 1 passo à frente gerado por um certo modelo  $M(\hat{\theta}^*)$  dado por (LJUNG, 1999):

$$\varepsilon(t, \hat{\theta}^*) = y(t) - \hat{y}(t/t-1; \hat{\theta}^*) \quad (10.3)$$

Em (10.3)  $\hat{y}(t/t-1; \hat{\theta}^*)$  denota uma predição um passo à frente de  $y(t)$ , fornecidos os dados até o instante  $(t-1)$ , isto é,  $y(t-1)$ ,  $u(t-1)$ ,  $y(t-2)$ ,  $u(t-2)$ , ... e baseado no vetor de parâmetros do modelo  $\hat{\theta}$ . Quando o conjunto de dados  $Z^N$  é conhecido, esses erros podem ser computados para  $t = 1, 2, \dots, N$ .

A Figura 10.1 mostra um diagrama de blocos do método do erro de predição (LJUNG, 1999).

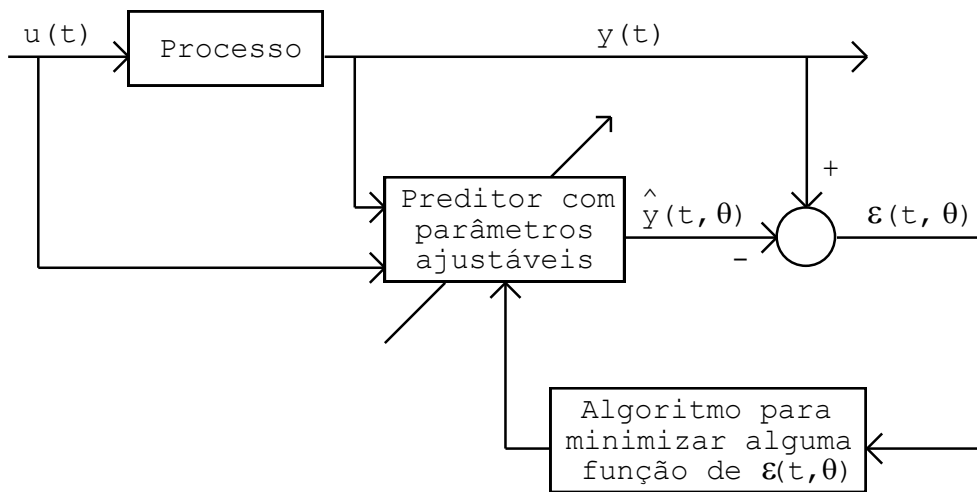


Fig. 10.1 Diagrama de blocos do método do erro de predição.

Diz-se que um “bom” modelo é aquele que seja bom para prever, isto é, um que produza erros de predição pequenos, quando aplicado aos dados observados.

Um princípio básico na estimação de parâmetros é (LJUNG, 1999):

“Baseado em  $Z^t$  pode-se computar o erro de predição  $\varepsilon(t, \hat{\theta})$  usando (10.3). No instante  $t = N$  selecione  $\hat{\theta}_N$  de forma que os erros de predição  $\varepsilon(t, \hat{\theta}_N)$ ,  $t = 1, 2, \dots, N$  se tornem tão pequenos quanto possível”. A questão é como qualificar o que “pequeno” significa. Na próxima subseção são descritas algumas possibilidades, que correspondem a formar uma norma de valor escalar ou uma função de mérito que meça o tamanho de  $\varepsilon$ .

### 10.2.1 Critérios para definir o melhor ajuste para o vetor de parâmetros $\hat{\theta}$

Esta subseção expõe os possíveis critérios usados para definir qual resposta do modelo provê o “melhor” ajuste à resposta experimental. Caso se possa especificar um critério quantitativo que defina o que é o “melhor”, então a representação do modelo pode ser me-

lhorada ajustando sua forma para melhorar o valor do critério. O melhor modelo exibe, em princípio, o menor erro entre os dados reais e a resposta simulada pelo modelo. Caso se faça  $\varepsilon(t)$  ser a diferença entre a resposta do modelo  $\hat{y}(t)$  e a resposta experimental  $y(t)$  no tempo  $t$  ou  $\varepsilon_i$  ser a diferença na  $i$ -ésima observação ( $\varepsilon_i = y_i - \hat{y}_i$ ,  $i = 1, 2, \dots, N$ , onde  $N$ =número total de pontos obtidos), as seguintes possibilidades existem:

- Integral ou somatório dos erros:  $\int \varepsilon(t) dt$  ou  $\sum_{i=1}^N \varepsilon_i$

Este critério ou função-objetivo tem o inconveniente de erros positivos se cancelarem com erros negativos.

- Integral ou somatório dos valores absolutos dos erros:  $\int |\varepsilon(t)| dt$  ou  $\sum_{i=1}^N |\varepsilon_i|$

Este critério penaliza igualmente desvios grandes e pequenos.

- Integral ou somatório do quadrado dos erros:  $\int [\varepsilon(t)]^2 dt$  ou  $\sum_{i=1}^N (\varepsilon_i)^2$

Este critério penaliza mais os grandes desvios que os pequenos, tratando-se do critério mais comumente utilizado. Além disso, esta função-objetivo, quando minimizada, conduz a uma solução analítica dos coeficientes desconhecidos. Se pesos são incluídos no somatório, de forma inversamente proporcional à incerteza no erro, chega-se a:

$$\int w_i \cdot [\varepsilon(t)]^2 dt \quad \text{ou} \quad \sum_{i=1}^N w_i \cdot (\varepsilon_i)^2$$

- Minimizar o máximo desvio:  $\min\{\max[\varepsilon(t)]\}$  ou  $\min\{\max[\varepsilon_i]\}$

Como uma pequena região de ajuste pobre pode ser encoberta por regiões grandes de ajuste bom quando se usa um somatório, este critério é utilizado somente no ponto em que o modelo mais se desvia da resposta experimental.

A seleção do tipo do critério certamente afeta os valores finais dos parâmetros do modelo, muito embora os valores não serão provavelmente muito diferentes se o modelo for razoavelmente bom.

Um aspecto importante a ser considerado ao criar modelos no domínio do tempo é que os parâmetros serão diferentes para diferentes entradas do processo (mesmo se o processo for linear), a menos que o modelo seja exato. A resposta em frequência, por outro lado, não sofre essa desvantagem se o processo for linear (SMITH, 1972).

A identificação de sistemas é inquestionavelmente mais fácil se o modelo for linear. Mas visto que a maioria dos processos é não linear, eles não podem ser representados em faixas extensas por modelos lineares. Isso significa que para um modelo linear ser adequado, ele deve ser determinado no ponto de operação normal (ou próximo a ele) do processo não linear. Se o ponto de operação muda apreciavelmente, é necessário reavaliar os parâmetros do modelo. Isso torna as técnicas de identificação *on line* muito desejáveis.

Quando o modelo é linear nos coeficientes, estes podem ser estimados através de um procedimento chamado "regressão linear". Se os coeficientes aparecem na função em uma forma não linear, a estimação dos mesmos é intitulada "regressão não linear".

### 10.2.2 Minimizando erros de predição

A sequência de erros de predição em (10.3) pode ser vista como um vetor em  $\mathbf{R}^N$ . O "tamanho" desse vetor pode ser medido usando qualquer norma no  $\mathbf{R}^N$ , quadrática ou não.

Considere que a sequência de erros de predição seja filtrada através de um filtro linear estável  $L(q)$  (LJUNG, 1999):

$$\varepsilon_F(t, \hat{\theta}) = L(q) \varepsilon(t, \hat{\theta}) \quad 1 \leq t \leq N \quad (10.4)$$

Seja então a seguinte norma:

$$V_N(\hat{\theta}, Z^N) = \frac{1}{N} \sum_{t=1}^N l(\varepsilon_F(t, \hat{\theta})) \quad (10.5)$$

onde  $l(\varepsilon)$  é tipicamente uma norma quadrática da forma:

$$l(\varepsilon) = \frac{1}{2} \varepsilon^2$$

Essa é uma escolha padrão, conveniente tanto para a computação quanto para a análise.

A função  $V_N(\hat{\theta}, Z^N)$  é, para um dado  $Z^N$ , uma função de valor escalar bem definida do vetor de parâmetros  $\theta$ . A função  $V_N(\hat{\theta})$  é intitulada função perda ou função custo (*loss function*). A estimativa  $\hat{\theta}_N$  é então definida pela minimização de (10.5) (LJUNG, 1999):

$$\hat{\theta}_N = \hat{\theta}_N(Z^N) = \arg \min_{\theta} V_N(\hat{\theta}, Z^N) \quad (10.6)$$

Aqui  $\arg \min$  significa "o argumento que minimiza a função". Se o mínimo não é único,  $\arg \min$  denota o conjunto de argumentos minimizantes.

Esta forma de estimar  $\theta$  contém muitos procedimentos bem conhecidos. O termo **métodos de identificação por erro de predição** (PEM - *prediction error methods*) é usado para a família de alternativas que corresponda a (10.6). Métodos específicos são obtidos como casos especiais de (10.6), dependendo da escolha de  $l(\varepsilon)$ , do pré-filtro  $L(q)$ , da estrutura do modelo e, em alguns casos, do método pelo qual a minimização é realizada. É dada especial atenção a um membro da família (10.6) na seção subsequente. Antes disso, no entanto, discutem-se alguns aspectos da escolha de  $L(q)$  em (10.4).

O efeito do filtro  $L$  é permitir liberdade adicional ao lidar com propriedades não transitórias dos erros de predição. Obviamente, se o preditor é linear e invariante no tempo e  $y$  e  $u$  são escalares, então o resultado de filtrar  $\varepsilon$  é o mesmo que primeiro filtrar os dados de entrada e saída e então aplicar os preditores.

O efeito de  $L$  é melhor compreendido através de uma interpretação no domínio da frequência de (10.6). Pelo uso de  $L$ , os efeitos de perturbações de alta frequência, não

essenciais ao problema de modelagem ou termos que derivem lentamente podem ser removidos.  $L$  assim age como um ponderador de frequências (LJUNG, 1999).

O seguinte aspecto particular da filtragem (10.4) deve ser notado. Se um modelo (10.2) é usado, o erro filtrado  $\varepsilon_F(t, \hat{\theta})$  é dado por (LJUNG, 1999):

$$\begin{aligned}\varepsilon_F(t, \hat{\theta}) &= L(q) \varepsilon(t, \hat{\theta}) = L(q) [y(t) - \hat{y}(t/t-1, \hat{\theta})] = \\ &= L(q) \{y(t) - H^{-1}(q, \hat{\theta}) G(q, \hat{\theta}) u(t) - [1 - H^{-1}(q, \hat{\theta})] y(t)\} = [L^{-1}(q) H(q, \hat{\theta})]^{-1} [y(t) - G(q, \hat{\theta}) u(t)]\end{aligned}$$

O efeito da pré-filtragem é assim idêntico a mudar o modelo do ruído de  $H(q, \hat{\theta})$  para:

$$H_L(q, \hat{\theta}) = L^{-1}(q) \cdot H(q, \hat{\theta})$$

Ao descrever e analisar métodos que empreguem modelos gerais de ruído em sistemas lineares, se usa normalmente  $L(q) \equiv 1$ , visto que a opção da pré-filtragem pode ser contornada pela liberdade na escolha de  $H(q, \hat{\theta})$  (LJUNG, 1999).

### 10.3 MÉTODO DOS MÍNIMOS QUADRADOS

Conforme já foi visto, as estruturas de modelo do tipo regressão linear são muito úteis para descrever sistemas lineares e não lineares. Pode-se escrever:

$$\hat{y}(t) = \varphi_1(t) \cdot \hat{\theta}_1 + \varphi_2(t) \cdot \hat{\theta}_2 + \dots + \varphi_p(t) \cdot \hat{\theta}_p = \varphi^T(t) \hat{\theta}$$

onde  $\hat{y}$  é a variável estimada de saída,  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p$  são parâmetros do modelo a ser determinados e  $\varphi_1, \varphi_2, \dots, \varphi_p$  são variáveis conhecidas. Pode-se definir os seguintes vetores:

$$\varphi^T(t) = [\varphi_1(t) \quad \varphi_2(t) \quad \dots \quad \varphi_p(t)] \quad \text{e} \quad \hat{\theta} = [\hat{\theta}_1 \quad \hat{\theta}_2 \quad \dots \quad \hat{\theta}_p]^T$$

O modelo é indexado pela variável  $t$ , que denota tempo discreto. A regressão linear emprega o seguinte preditor:

$$\hat{y}(t/\hat{\theta}) = \varphi^T(t) \hat{\theta} \quad (10.7)$$

Para a estrutura ARX tem-se que:

$$\begin{aligned}y(t) &= -a_1 \cdot y(t-1) - a_2 \cdot y(t-2) - \dots - a_{n_a} \cdot y(t-n_a) + b_1 \cdot u(t-1) + b_2 \cdot u(t-2) + \dots \\ &\quad + b_{n_b} \cdot u(t-n_b) + \varepsilon(t)\end{aligned}$$

Pode-se reescrever esta equação na forma:

$$\left[1 + a_1 \cdot q^{-1} + a_2 \cdot q^{-2} + \dots + a_{n_a} \cdot q^{-n_a}\right] \cdot y(t) = \left[b_1 \cdot q^{-1} + b_2 \cdot q^{-2} + \dots + b_{n_b} \cdot q^{-n_b}\right] \cdot u(t) + \varepsilon(t)$$

Portanto, para a estrutura ARX, tem-se que:

$$A(q) \cdot y(t) = B(q) \cdot u(t) + \varepsilon(t) \quad (10.8)$$

Usando-se a notação da equação (10.7) para a estrutura ARX, resulta:

$$\varphi^T(t) = [-y(t-1) \quad -y(t-2) \quad \dots \quad -y(t-n_a) \quad u(t-1) \quad u(t-2) \quad \dots \quad u(t-n_b)] \quad (10.9)$$

O vetor de parâmetros é dado por:

$$\theta = [a_1 \quad a_2 \quad \dots \quad a_{n_a} \quad b_1 \quad b_2 \quad \dots \quad b_{n_b}]^T$$

As variáveis  $\varphi(t)$  são chamadas de variáveis de regressão ou regressores e o modelo acima pode ser chamado de modelo de regressão. O problema aqui é encontrar

uma estimativa  $\hat{\theta}$  do vetor de parâmetros  $\theta$  a partir das medições (dados históricos)  $y(1)$ ,  $\varphi(1)$ , ...,  $y(N)$ ,  $\varphi(N)$ , onde  $N$  corresponde ao número coletado de pontos de entrada/saída. Pares de valores de saída e regressores  $\{y(t), \varphi(t), t=1, 2, \dots, N\}$  são coletados, de tal modo que as saídas estimadas do modelo sejam tão próximas quanto possível das variáveis medidas  $y(t)$ , no sentido dos mínimos quadrados. Então, o parâmetro  $\theta$  pode ser escolhido de modo a minimizar a função perda ou função custo dos mínimos quadrados.

Dadas essas medições, um sistema de equações lineares é obtido, a saber (SÖDERSTRÖM; STOICA, 1989):

$$\hat{y}(1) = \varphi^T(1) \hat{\theta}$$

$$\hat{y}(2) = \varphi^T(2) \hat{\theta}$$

$$\vdots$$

$$\hat{y}(N) = \varphi^T(N) \hat{\theta}$$

É conveniente enumerar esses valores usando um argumento  $t$ :

$$y(t), \varphi(t) \quad t = 1, \dots, N$$

Pode-se então formar um sistema de  $N$  equações a partir da sequência de dados medida do processo:

$$y = \Phi \theta + \varepsilon$$

$$\text{onde: } y = \begin{bmatrix} y(n_a + 1) \\ y(n_a + 2) \\ \vdots \\ y(n_a + N) \end{bmatrix} \quad \theta = \begin{bmatrix} a_1 \\ \vdots \\ a_{n_a} \\ b_1 \\ \vdots \\ b_{n_b} \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon(n_a + 1) \\ \varepsilon(n_a + 2) \\ \vdots \\ \varepsilon(n_a + N) \end{bmatrix}$$

$$\Phi = \begin{bmatrix} \varphi(n_a + 1) \\ \varphi(n_a + 2) \\ \vdots \\ \varphi(n_a + N) \end{bmatrix} = \begin{bmatrix} -y(n_a) & -y(n_a - 1) & \dots & -y(1) & | & u(n_b) & u(n_b - 1) & \dots & u(1) \\ -y(n_a + 1) & -y(n_a) & \dots & -y(2) & | & u(n_b + 1) & u(n_b) & \dots & u(2) \\ \vdots & \vdots & \ddots & \vdots & | & \vdots & \vdots & \ddots & \vdots \\ -y(n_a + N - 1) & -y(n_a + N - 2) & \dots & -y(N) & | & u(n_b + N - 1) & u(n_b + N - 2) & \dots & u(N) \end{bmatrix}$$

Deve-se ter  $N \gg n_a + n_b$ .

### 10.3.1 Função perda ou função custo (*loss function*)

Com os dados históricos pode-se substituir a variância  $E[\varepsilon^2(t, \hat{\theta})] = E[y - \hat{y}(t)]^2$  pela seguinte expressão:

$$\frac{1}{2N} \sum_{t=1}^N [y(t) - \hat{y}(t)]^2$$

Segundo Gauss, os parâmetros desconhecidos de um modelo podem ser escolhidos de modo que o somatório dos quadrados das diferenças entre o valor observado e o valor

estimado seja um mínimo, de maneira que as saídas estimadas do modelo  $\hat{y}(t)$  sejam tão próximas quanto possível das variáveis medidas  $y(t)$ . Então, o vetor de parâmetros  $\hat{\theta}$  pode ser escolhido de modo a minimizar a função perda ou função custo dos mínimos quadrados. No caso linear, tem-se:

$$V_N(\hat{\theta}) = \frac{1}{2N} \sum_{t=1}^N \varepsilon^2(t, \hat{\theta}) = \frac{1}{2N} \sum_{t=1}^N [y(t) - \varphi^T(t) \hat{\theta}]^2 \quad (10.10)$$

onde  $V_N(\hat{\theta})$  é chamada de função perda ou função custo. Este é o critério dos mínimos quadrados para a regressão linear (10.7). Pode-se também chegar a (10.10) tomando-se as Equações (10.4) e (10.5), com  $L(q) = 1$  e  $I(\varepsilon) = \varepsilon^2/2$ . Um valor adequado para  $\hat{\theta}$  é aquele que minimize  $V_N(\hat{\theta})$ :

$$\hat{\theta}_N = \arg \min_{\theta} V_N(\hat{\theta})$$

$\hat{\theta}_N$  é a estimativa dos mínimos quadrados. Pode-se então usar  $\varphi^T \hat{\theta}_N$  como uma função de predição. Portanto, o método dos mínimos quadrados determina os parâmetros do modelo, de tal forma que o somatório do quadrado dos erros de predição seja minimizado, sendo, portanto, um caso particular dos chamados métodos do erro de predição (PEM). Repare que este método de selecionar  $\hat{\theta}$  faz sentido caso se tenha ou não imposto um caráter estocástico ao problema. O parâmetro  $\hat{\theta}_N$  é simplesmente o valor que gera o melhor preditor quando aplicado a dados históricos (SÖDERSTRÖM; STOICA, 1989).

Como se supôs que a variável medida  $y(t)$  seja linear nos parâmetros  $\hat{\theta}$  e como a Equação (10.10) é uma função quadrática de  $\varepsilon(t)$ , isto implica que  $V_N(\hat{\theta})$  pode ser minimizada analiticamente:

$$\frac{\partial [V_N(\hat{\theta})]}{\partial \hat{\theta}} = 0$$

Derivando-se (10.10) com relação a  $\hat{\theta}$ , resulta:

$$0 = \frac{1}{N} \sum_{t=1}^N -\varphi(t) \cdot [y(t) - \varphi^T(t) \hat{\theta}]$$

Todo  $\hat{\theta}_N$  que satisfaça:

$$\left[ \sum_{t=1}^N \varphi(t) \varphi^T(t) \right] \hat{\theta}_N = \sum_{t=1}^N \varphi(t) y(t) \quad (10.11)$$

gera o mínimo global de  $V_N(\hat{\theta})$ . Estas equações lineares são conhecidas como **equações normais**. Se a matriz do primeiro termo for inversível, tem-se a estimativa dos mínimos quadrados (LSE - *least squares estimate*) (LJUNG, 1999; SÖDERSTRÖM; STOICA, 1989):

$$\hat{\theta}_N = \arg \min_{\theta} V_N(\hat{\theta}) = \left[ \sum_{t=1}^N \varphi(t) \varphi^T(t) \right]^{-1} \sum_{t=1}^N \varphi(t) y(t) \quad (10.12)$$

Pode-se garantir que a matriz do primeiro termo em (10.11) seja inversível (não singular), se o sinal de entrada  $u(t)$  for persistentemente excitante de ordem  $n_a + n_b$ .

Define-se a matriz quadrada  $P(t)$  como:

$$\mathbf{P}(t) = \left[ \sum_{t=1}^N \varphi(t) \varphi^T(t) \right]^{-1}$$

de modo que se pode escrever:

$$\hat{\theta}_N = \mathbf{P}(t) \left[ \sum_{t=1}^N \varphi(t) y(t) \right]$$

Seja a matriz  $\mathbf{R}(t)$  dada por:

$$\mathbf{R}(t) = \mathbf{P}^{-1}(t) = \sum_{t=1}^N \varphi(t) \varphi^T(t) \quad (10.13)$$

No caso da equação (10.9),  $\varphi(t)$  contém variáveis de entrada e saída com atraso e os elementos de (10.13) são da forma (LJUNG, 1999):

$$[\mathbf{R}(t)]_{ij} = \sum_{t=1}^N y(t-i) \cdot y(t-j) \quad 1 \leq i, j \leq n_a$$

e somas similares de  $u(t-r) \cdot u(t-s)$  ou  $u(t-r) \cdot y(t-s)$  para os outros elementos de  $\mathbf{R}(t)$ . Isto é, eles consistem em estimativas das funções covariância de  $\{y(t)\}$  e  $\{u(t)\}$ . A estimativa dos mínimos quadrados pode assim ser computada usando apenas essas quantidades e é, portanto, relacionada com a análise de correlação.

### 10.3.2 Método dos mínimos quadrados aplicado a sistemas com notação matricial

As expressões da seção anterior podem ser escritas mais compactamente em notação matricial. Assim, organizam-se os regressores  $\varphi(t)$  da expressão (10.9) em uma matriz do tipo:

$$\Phi = \begin{bmatrix} -y(0) & 0 & 0 & \dots & 0 & u(0) & 0 & 0 & \dots & 0 \\ -y(1) & -y(0) & 0 & \dots & 0 & u(1) & u(0) & 0 & \dots & 0 \\ -y(2) & -y(1) & -y(0) & \dots & 0 & u(2) & u(1) & u(0) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & \vdots & \vdots & \vdots & \ddots & 0 \\ -y(N-2) & -y(N-3) & -y(N-4) & \dots & -y(N-1-n_a) & u(N-2) & u(N-3) & u(N-4) & \dots & u(N-1-n_b) \\ -y(N-1) & -y(N-2) & -y(N-3) & \dots & -y(N-n_a) & u(N-1) & u(N-2) & u(N-3) & \dots & u(N-n_b) \end{bmatrix}$$

Reescreve-se esta matriz considerando que em cada linha da mesma existam  $p$  elementos:

$$\Phi = \begin{bmatrix} \varphi_{11} & \dots & \varphi_{1p} \\ \vdots & \ddots & \vdots \\ \varphi_{N1} & \dots & \varphi_{Np} \end{bmatrix} = \begin{bmatrix} \varphi^T(1) \\ \vdots \\ \varphi^T(N) \end{bmatrix} \quad \text{matriz com dimensão } (N \times p)$$

Colocam-se os dados medidos de saída também em um vetor:

$$\mathbf{Y} = [y(1) \ y(2) \ \dots \ y(N)]^T \quad \text{vetor com dimensão } (N \times 1)$$

Deseja-se determinar o vetor de parâmetros:

$$\hat{\theta} = [\hat{\theta}_1 \ \hat{\theta}_2 \ \dots \ \hat{\theta}_p]^T$$



tal que:

$$\Phi \hat{\theta} = Y \quad (10.14)$$

Uma forma de encontrar  $\hat{\theta}$  seria escolher o número de medições  $N$  igual à dimensão de  $\hat{\theta}$ , isto é,  $N = p$ . Então  $\Phi$  se torna uma matriz quadrada. Se a matriz for não singular, o sistema linear de equações (10.14) pode ser resolvido para  $\hat{\theta}$ :

$$\hat{\theta} = \Phi^{-1} Y$$

Na prática, ruído, perturbações e uma adaptação imperfeita do modelo à realidade são boas razões para se usar um número de dados  $N$  maior que  $p$ . Com os dados adicionais é possível obter uma estimativa melhor. Quando  $N > p$ , o sistema linear de equações (10.14) se torna sobre-determinado. Uma solução exata, em geral, não existirá (SÖDERSTRÖM; STOICA, 1989). Neste caso, deseja-se encontrar o vetor de parâmetros  $\hat{\theta}$ , que determine a solução que melhor aproxime os pontos dados. Sejam as equações de erro ou resíduos dados por:

$$E(t, \hat{\theta}) = [\varepsilon(1) \ \varepsilon(2) \ \dots \ \varepsilon(N)]^T = y(t) - \hat{y}(t) = y(t) - \varphi^T(t) \hat{\theta} \quad (10.15)$$

onde os resíduos  $\varepsilon(t)$  são definidos por:

$$\varepsilon(t) = y(t) - \hat{y}(t) = y(t) - \varphi^T(t) \hat{\theta}$$

O vetor de erros de predição  $E(t, \hat{\theta})$  é definido como:

$$E(t, \hat{\theta}) = \begin{pmatrix} \varepsilon(1, \hat{\theta}) \\ \vdots \\ \varepsilon(N, \hat{\theta}) \end{pmatrix} = Y - \hat{Y} = Y - \Phi \hat{\theta}$$

Assim, pelo método dos mínimos quadrados, o objetivo é minimizar a função:

$$E(t, \hat{\theta}) = \|Y - \Phi \hat{\theta}\|^2$$

Na literatura estatística, as equações de erro são frequentemente chamadas **resíduos**. A **estimativa dos mínimos quadrados** de  $\theta$  é definida como o vetor  $\hat{\theta}$  que minimize a função perda. Com a notação aqui apresentada, a função perda pode ser escrita como (SÖDERSTRÖM; STOICA, 1989):

$$V_N(\hat{\theta}) = \frac{1}{2N} \sum_{t=1}^N \varepsilon_N^2(t, \hat{\theta}) = \frac{1}{2N} E_N^2 = \frac{1}{2N} E_N^T E_N = \frac{1}{2N} \|E_N\|^2 \quad (10.16)$$

onde  $\|E_N\|$  denota a norma euclidiana do vetor. De acordo com (10.15), a equação de erro  $E_N(t, \hat{\theta})$  é uma função linear do vetor de parâmetros  $\hat{\theta}_N$ .

Pode-se então escrever (10.16) como:

$$V_N(\hat{\theta}) = \frac{1}{2N} |Y_N - \Phi_N \hat{\theta}_N|^2 = \frac{1}{2N} (Y_N - \Phi_N \hat{\theta}_N)^T (Y_N - \Phi_N \hat{\theta}_N)$$

A solução para o problema dos mínimos quadrados é dada pela seguinte equação normal [vide (10.11)]:

$$[\Phi_N^T \Phi_N] \hat{\theta}_N = \Phi_N^T Y_N$$

Se a matriz  $\Phi_N^T \Phi_N$  for não singular, o mínimo da função perda da equação (10.16) é único, dado pelo vetor  $\hat{\theta}_N$ , tal que [vide (10.12)]:

$$\hat{\theta}_N = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y_N \quad (10.17)$$

A resolução direta desta equação não é simples nem rápida. Mesmo com programas computacionais eficientes (por exemplo, o Matlab), a resolução da inversão de matriz pode gerar diversos erros e é comum que esta operação não possa ser efetuada devido à matriz ser singular para a precisão utilizada pelo software. Neste caso  $P$  é dado por:

$$P = [\Phi_N^T \Phi_N]^{-1} \quad \text{e} \quad R = P^{-1} = \Phi_N^T \Phi_N \quad (10.18)$$

A equação (10.15) fornece a solução para o sistema sobre-determinado ( $N > p$ ) de equações lineares citado em (10.14). Se a matriz  $\Phi_N^T \Phi_N$  for definida positiva, então  $V_N(\hat{\theta})$  tem um único ponto de mínimo dado por (10.17). O valor mínimo correspondente de  $V_N(\hat{\theta})$  é (SÖDERSTRÖM; STOICA, 1989):

$$\min_{\theta} V_N(\theta) = V_N(\hat{\theta}) = \frac{1}{2N} \left[ Y_N^T Y_N - Y_N^T \Phi_N (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y_N \right] \quad (10.19)$$

O critério dos mínimos quadrados pondera todos os erros  $\varepsilon(t)$  igualmente, o que corresponde a afirmar que todas as medidas têm a mesma precisão ou o mesmo peso. Diferentes pesos dos erros podem ser considerados, modificando a função perda para:

$$V_N(\hat{\theta}) = \frac{1}{2N} E_N^T W E_N$$

onde  $W$  é uma matriz diagonal com os “pesos” na diagonal. A estimativa dos mínimos quadrados é então dada por:

$$\hat{\theta}_N = [\Phi_N^T W \Phi_N]^{-1} \Phi_N^T W Y_N$$

### 10.3.3 Exemplos de aplicação do método dos mínimos quadrados

#### 10.3.3.1 Exemplo 1

Assuma o seguinte modelo (SÖDERSTRÖM; STOICA, 1989):

$$y(t) = k$$

Isto significa que uma constante deve ser estimada a partir de um certo número de medições ruidosas. Neste caso:

$$\varphi(t) = 1 \quad \theta = k$$

Portanto, de (10.12):

$$\hat{\theta} = \left[ \sum_{t=1}^N 1 \right]^{-1} \sum_{t=1}^N y(t) = \frac{1}{N} [y(1) + y(2) + \dots + y(N)]$$

Esta expressão corresponde à média aritmética de todas as medições.

#### 10.3.3.2 Exemplo 2

Deseja-se ajustar um modelo quadrático ao seguinte conjunto de dados, com  $N=10$

pontos (EDGAR; HIMMELBLAU, 1988):

<b>x</b>	<b>y</b>
20	73
20	78
30	85
40	90
40	91
50	87
50	86
50	91
60	75
70	65

O modelo desejado é da forma:

$$\hat{y} = \hat{a}_0 + \hat{a}_1 \cdot x + \hat{a}_2 \cdot x^2 \quad \text{ou} \quad \hat{y} = \hat{a}_0 \cdot x_0 + \hat{a}_1 \cdot x_1 + \hat{a}_2 \cdot x_2$$

onde:  $x_0 = 1$        $x_1 = x$       e       $x_2 = x^2$

Primeiro deve-se gerar os seguintes vetores e matrizes:

$$\begin{matrix} x_0 & x & x^2 \\ \begin{bmatrix} 1 & 20 & 400 \\ 1 & 20 & 400 \\ 1 & 30 & 900 \\ 1 & 40 & 1600 \\ 1 & 40 & 1600 \\ 1 & 50 & 2500 \\ 1 & 50 & 2500 \\ 1 & 50 & 2500 \\ 1 & 60 & 3600 \\ 1 & 70 & 4900 \end{bmatrix} & \mathbf{X}^T \mathbf{X} = \begin{bmatrix} N & \sum x & \sum x^2 \\ \sum x & \sum x^2 & \sum x^3 \\ \sum x^2 & \sum x^3 & \sum x^4 \end{bmatrix} = \begin{bmatrix} 10 & 430 & 20.900 \\ 430 & 20.900 & 1.105.000 \\ 20.900 & 1.105.000 & 61.970.000 \end{bmatrix} \end{matrix}$$

$$\mathbf{X}^T \mathbf{Y} = \begin{bmatrix} \sum y \\ \sum x \cdot y \\ \sum x^2 \cdot y \end{bmatrix} = \begin{bmatrix} 821 \\ 35.060 \\ 1.675.000 \end{bmatrix}$$

O vetor:

$$\hat{\mathbf{a}} = [\hat{a}_0 \quad \hat{a}_1 \quad \hat{a}_2]^T$$

é obtido resolvendo-se a seguinte equação:

$$\hat{\mathbf{a}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

Resulta:

$$\hat{a}_0 = 35,66 \quad \hat{a}_1 = 2,63 \quad \hat{a}_2 = -0,032$$

### 10.3.3.3 Exemplo 3

Suponha que se deseje usar o seguinte modelo linear para aproximar  $y$  (EDGAR; HIMMELBLAU, 1988):

$$\hat{y} = \hat{a}_0 + \hat{a}_1 \cdot x$$

A função-objetivo é:

$$V_N = \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N (y_i - \hat{a}_0 - \hat{a}_1 \cdot x_i)^2$$

Os coeficientes desconhecidos são  $\hat{a}_0$  e  $\hat{a}_1$ . São disponíveis  $N$  pares de valores experimentais de  $y_i$  e  $x_i$ . Deseja-se minimizar  $V_N$  com relação a  $\hat{a}_0$  e  $\hat{a}_1$ . Para tanto, tomam-se as primeiras derivadas parciais de  $V_N$  e se iguala a zero:

$$\frac{\partial V_N}{\partial \hat{a}_0} = 0 = -2 \sum_{i=1}^N (y_i - \hat{a}_0 - \hat{a}_1 \cdot x_i) \quad \frac{\partial V_N}{\partial \hat{a}_1} = 0 = 2 \sum_{i=1}^N (y_i - \hat{a}_0 - \hat{a}_1 \cdot x_i) \cdot (-x_i)$$

Rearranjando-se as equações acima, resulta um conjunto de equações lineares nas incógnitas  $\hat{a}_0$  e  $\hat{a}_1$ :

$$\sum_{i=1}^N \hat{a}_0 + \sum_{i=1}^N \hat{a}_1 \cdot x_i = \sum_{i=1}^N y_i \quad \sum_{i=1}^N \hat{a}_0 \cdot x_i + \sum_{i=1}^N \hat{a}_1 \cdot x_i^2 = \sum_{i=1}^N x_i \cdot y_i$$

As expressões acima podem ser reescritas como:

$$\hat{a}_0 \cdot N + \hat{a}_1 \sum_{i=1}^N x_i = \sum_{i=1}^N y_i \quad \hat{a}_0 \sum_{i=1}^N x_i + \hat{a}_1 \sum_{i=1}^N x_i^2 = \sum_{i=1}^N x_i \cdot y_i$$

Suponha então que se disponha dos seguintes dados:

x	Y
0	0
1	2
2	4
3	6
4	8
5	10

Tem-se que:

$$\sum_{i=1}^6 x_i = 15 \quad \sum_{i=1}^6 x_i \cdot y_i = 110 \quad \sum_{i=1}^6 y_i = 30 \quad \sum_{i=1}^6 x_i^2 = 55$$

Então:

$$6 \cdot \hat{a}_0 + 15 \cdot \hat{a}_1 = 30$$

$$15 \cdot \hat{a}_0 + 55 \cdot \hat{a}_1 = 110$$

Resolvendo-se este sistema de equações, resulta:

$$\hat{a}_0 = 0 \quad \hat{a}_1 = 2$$

Portanto:

$$\hat{y} = 2 \cdot x$$

## 10.3.3.4 Exemplo 4

Suponha o seguinte exemplo de aplicação do método dos mínimos quadrados (NORTON, 1986):

A posição  $x$  de um alvo de radar movendo-se em linha reta é observada a intervalos de 0,2 s de 0 a 1 s. Deve-se gerar um modelo para prever a sua posição. Uma forma simples de fazer isso é assumir aceleração constante ao longo da observação e do intervalo de predição, estimar a posição inicial  $\hat{x}_0$ , a velocidade inicial  $\hat{v}_0$  e a aceleração  $\hat{a}$  e prever as futuras posições usando o modelo:

$$\hat{x}(t) = \hat{x}_0 + \hat{v}_0 \cdot t + \frac{\hat{a} \cdot t^2}{2}$$

Pede-se então estimar os parâmetros deste modelo, empregando-se o método dos mínimos quadrados. É dado:

$k$	0	1	2	3	4	5
$t$ (s)	0	0,2	0,4	0,6	0,8	1,0
$x$ (m)	3	59	98	151	218	264

Tem-se que:

$$\hat{x}(t) = \varphi^T(t) \hat{\theta}$$

$$\text{onde: } \varphi^T(t) = \begin{bmatrix} 1 & t & \frac{t^2}{2} \end{bmatrix} \quad \text{e} \quad \hat{\theta} = \begin{bmatrix} \hat{x}_0 \\ \hat{v}_0 \\ \hat{a} \end{bmatrix}$$

Calcula-se então o erro de predição:

$$\varepsilon(t) = x(t) - \hat{x}(t) = x(t) - \varphi^T(t) \hat{\theta}$$

$$\varepsilon^2(t) = x^2(t) + \hat{x}_0^2 + \hat{v}_0^2 \cdot t^2 + \frac{\hat{a}^2 \cdot t^4}{4} - 2 \cdot x \cdot \hat{x}_0 - 2 \cdot x \cdot \hat{v}_0 \cdot t - x \cdot \hat{a} \cdot t^2 + 2 \cdot \hat{x}_0 \cdot \hat{v}_0 \cdot t + \hat{x}_0 \cdot \hat{a} \cdot t^2 + \hat{v}_0 \cdot \hat{a} \cdot t^3$$

Substitui-se o tempo  $t$  pelo seu equivalente discreto:

$$t = k \cdot T \quad \text{onde } T = 0,2 \text{ s}$$

A função custo é dada por:

$$V_N(\hat{\theta}) = \frac{1}{2 \cdot N} \sum_{k=0}^{N-1} \varepsilon^2(k)$$

O valor mínimo da função custo é dado por:

$$\frac{\partial V_N(\hat{\theta})}{\partial \hat{\theta}} = 0$$

Resulta então:

$$\frac{\partial V_N(\hat{\theta})}{\partial \hat{x}_0} = \frac{1}{N} \sum_{k=0}^{N-1} \left[ x(k) - \varphi^T(k) \hat{\theta} \right] \varphi^T(k) \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{N} \sum_{k=0}^{N-1} \left[ x(k) - \hat{x}_0 - \hat{v}_0 \cdot k \cdot T - \frac{\hat{a} \cdot (k \cdot T)^2}{2} \right] = 0 \quad (10.20)$$

$$\frac{\partial V_N(\hat{\theta})}{\partial \hat{v}_0} = \frac{1}{N} \sum_{k=0}^{N-1} \left[ x(k) - \varphi^T(k) \hat{\theta} \right] \varphi^T(k) \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{N} \sum_{k=0}^{N-1} \left[ x(k) - \varphi^T(k) \hat{\theta} \right] \cdot k \cdot T$$

$$\frac{\partial V_N(\hat{\theta})}{\partial \hat{v}_0} = \frac{1}{N} \sum_{k=0}^{N-1} \left[ x(k) \cdot k \cdot T - \hat{x}_0 \cdot k \cdot T - \hat{v}_0 \cdot (k \cdot T)^2 - \frac{\hat{a} \cdot (k \cdot T)^3}{2} \right] = 0 \quad (10.21)$$

$$\frac{\partial V_N(\hat{\theta})}{\partial \hat{a}} = \frac{1}{N} \sum_{k=0}^{N-1} \left[ x(k) - \varphi^T(k) \hat{\theta} \right] \varphi^T(k) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{N} \sum_{k=0}^{N-1} \left[ x(k) - \varphi^T(k) \hat{\theta} \right] \frac{(k \cdot T)^2}{2}$$

$$\frac{\partial V_N(\hat{\theta})}{\partial \hat{a}} = \frac{1}{N} \sum_{k=0}^{N-1} \left[ \frac{x(k) \cdot (k \cdot T)^2}{2} - \frac{\hat{x}_0 \cdot (k \cdot T)^2}{2} - \frac{\hat{v}_0 \cdot (k \cdot T)^3}{2} - \frac{\hat{a} \cdot (k \cdot T)^4}{4} \right] = 0 \quad (10.22)$$

Há  $N=6$  dados, que produzem:

$$\begin{aligned} \sum_{k=1}^N x(k) &= 793 & \sum_{k=1}^N x(k) \cdot k \cdot T &= 580 & \sum_{k=1}^N x(k) \cdot (k \cdot T)^2 &= 475,92 \\ \sum_{k=1}^N k \cdot T &= 3 & \sum_{k=1}^N (k \cdot T)^2 &= 2,2 & \sum_{k=1}^N (k \cdot T)^3 &= 1,8 & \sum_{k=1}^N (k \cdot T)^4 &= 1,57 \end{aligned}$$

Substituindo-se estes resultados nas Equações (10.20), (10.21) e (10.22), resulta:

$$6 \cdot \hat{x}_0 + 3 \cdot \hat{v}_0 + 1,1 \cdot \hat{a} = 793$$

$$3 \cdot \hat{x}_0 + 2,2 \cdot \hat{v}_0 + 0,9 \cdot \hat{a} = 580$$

$$1,1 \cdot \hat{x}_0 + 0,9 \cdot \hat{v}_0 + 0,392 \cdot \hat{a} = 237,96$$

Resolvendo-se este sistema de equações lineares resulta:

$$\begin{bmatrix} \hat{x}_0 \\ \hat{v}_0 \\ \hat{a} \end{bmatrix} = \frac{1}{0,064} \begin{bmatrix} 0,0524 & -0,186 & 0,28 \\ -0,186 & 1,142 & -2,1 \\ 0,28 & -2,1 & 4,2 \end{bmatrix} \begin{bmatrix} 793 \\ 580 \\ 237,96 \end{bmatrix} = \begin{bmatrix} 4,786 \\ 234,5 \\ 55,36 \end{bmatrix}$$

O modelo resultante é:

$$\hat{x}(t) = 4,786 + 234,5 \cdot t + \frac{55,36 \cdot t^2}{2} = 4,786 + 234,5 \cdot t + 27,68 \cdot t^2$$

A Figura 10.2 mostra o valor estimado  $\hat{x}(t)$  e o valor medido  $x(t)$ .

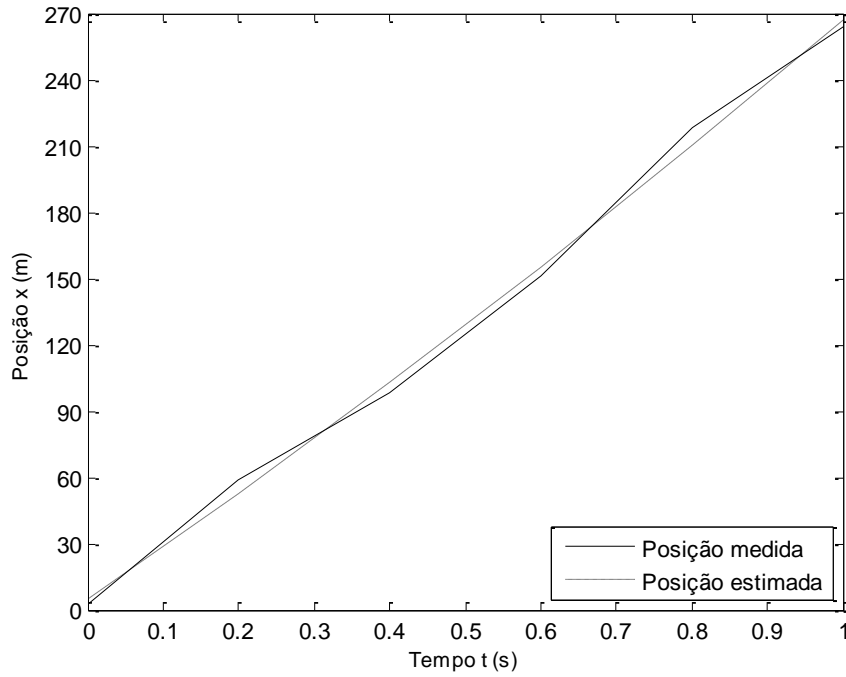


Fig. 10.2 Comparação do valor estimado  $\hat{x}(t)$  com o valor medido  $x(t)$ .

Este problema pode também ser resolvido empregando-se a notação matricial descrita na subseção 10.3.1. O valor de  $\hat{\theta}$  é estimado pela equação (10.17), transcrita a seguir.

$$\hat{\theta}_N = [\Phi_N^T \Phi_N]^{-1} \Phi_N^T Y_N$$

A matriz  $\Phi_N$  e o vetor  $Y_N$  são dados por:

$$\Phi_N = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0,2 & 0,02 \\ 1 & 0,4 & 0,08 \\ 1 & 0,6 & 0,18 \\ 1 & 0,8 & 0,32 \\ 1 & 1 & 0,5 \end{bmatrix} \quad Y_N = \begin{bmatrix} 3 \\ 59 \\ 98 \\ 151 \\ 218 \\ 264 \end{bmatrix}$$

Calculando-se inicialmente o termo  $[\Phi_N^T \Phi_N]^{-1}$ , tem-se que:

$$[\Phi_N^T \Phi_N]^{-1} = \begin{bmatrix} 0,82143 & -2,9464 & 4,4643 \\ -2,9464 & 18,1696 & -33,4821 \\ 4,4643 & -33,4821 & 66,9643 \end{bmatrix}$$

Calculando-se agora o termo  $\Phi_N^T Y_N$ :

$$\Phi_N^T Y_N = \begin{bmatrix} 793 \\ 580 \\ 237,96 \end{bmatrix}$$

Resolvendo-se  $\hat{\theta}_N = [\Phi_N^T \Phi_N]^{-1} \Phi_N^T Y_N$ , resulta:

$$\hat{\theta}_N = \begin{bmatrix} \hat{x}_0 \\ \hat{v}_0 \\ \hat{a} \end{bmatrix} = [\Phi_N^T \Phi_N]^{-1} \Phi_N^T Y_N = \begin{bmatrix} 4,786 \\ 234,5 \\ 55,36 \end{bmatrix}$$

Portanto:

$$\hat{x}(t) = 4,786 + 234,5 \cdot t + 27,68 \cdot t^2$$

Caso se deseje calcular o valor da função perda  $V_N(\hat{\theta})$ , resulta:

$t$ (s)	0	0,2	0,4	0,6	0,8	1
$x$ (m)	3	59	98	151	218	264
$\hat{x}$ (m)	4,79	52,79	103,0	155,5	210,1	267,0
$\varepsilon = x - \hat{x}$ (m)	-1,79	6,21	-5,00	-4,50	7,90	-3,00
$\varepsilon^2$ (m <sup>2</sup> )	3,20	38,56	25,00	20,25	62,41	9,00

$$V_N(\hat{\theta}) = \frac{1}{2 \cdot N} \sum_{k=0}^{N-1} \varepsilon^2(k \cdot T) = 13,20$$

Pode-se calcular este mesmo valor empregando-se a expressão mostrada a seguir.

$$V_N(\hat{\theta}) = \frac{1}{2N} \left[ Y_N^T Y_N - Y_N^T \Phi_N (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y_N \right] \quad (10.23)$$

### 10.3.4 Método dos mínimos quadrados aplicado a modelos não lineares

O critério dos mínimos quadrados:

$$V_N(\hat{\theta}) = \frac{1}{2N} \sum_{t=1}^N [y(t) - g(\varphi(t), \hat{\theta})]^2$$

pode ser usado como uma medida do ajuste de um modelo não linear. A estimativa fica:

$$\hat{\theta}_N = \arg \min_{\theta} V_N(\hat{\theta})$$

idêntica à forma linear. A diferença importante é que pode não ser possível encontrar expressões explícitas para  $\hat{\theta}_N$  como em (10.12) ou (10.17), mas se deve recorrer a técnicas numéricas iterativas.

#### 10.3.4.1 Exemplo de aplicação do método dos mínimos quadrados a sistema não linear

Considere o seguinte sistema não linear:

$$y(t) = b_0 + b_1 \cdot u(t) + b_2 \cdot u^2(t) + e(t)$$

onde  $e(t)$  é um ruído de média zero e desvio padrão 0,1. O sistema é linear nos parâmetros e pode ser escrito na forma:

$$\varphi^T(t) = [1 \quad u(t) \quad u^2(t)] \quad \hat{\theta} = [\hat{b}_0 \quad \hat{b}_1 \quad \hat{b}_2]^T$$

Na prática, a estrutura do modelo é desconhecida e deve-se decidir a escolha do modelo apropriado. Isto é ilustrado, estimando-se os parâmetros dos seguintes modelos:

- Modelo 1:  $\hat{y}(t) = \hat{b}_0$



- Modelo 2:  $\hat{y}(t) = \hat{b}_0 + \hat{b}_1 \cdot u(t)$
- Modelo 3:  $\hat{y}(t) = \hat{b}_0 + \hat{b}_1 \cdot u(t) + \hat{b}_2 \cdot u^2(t)$
- Modelo 4:  $\hat{y}(t) = \hat{b}_0 + \hat{b}_1 \cdot u(t) + \hat{b}_2 \cdot u^2(t) + \hat{b}_3 \cdot u^3(t)$

A Tabela 10.1 mostra a estimativa dos mínimos quadrados para os parâmetros dos diferentes modelos, junto com a função perda resultante.

Tab. 10.1 Estimativa dos parâmetros dos quatro modelos do exemplo do item 10.3.4.1.

Modelo	$\hat{b}_0$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	V
1	3,85				34,46
2	0,57	1,09			1,01
3	1,11	0,45	0,11		0,031
4	1,13	0,37	0,14	-0,003	0,027

Este exemplo mostra que é importante escolher a estrutura correta para conseguir um bom modelo. Com poucos parâmetros não é possível conseguir um bom ajuste para os dados. Se parâmetros demais são usados, o ajuste para os dados medidos é muito bom, mas o ajuste para uma outra série de pontos pode ser ruim.

### REFERÊNCIAS BIBLIOGRÁFICAS

- EDGAR, T. F.; HIMMELBLAU, D. M. **Optimization of chemical processes**. New York, McGraw Hill, 1988.
- LJUNG, L. **System identification: theory for the user**. 2.ed., Englewood Cliffs, Prentice Hall, 1999.
- NORTON, J. P. **An introduction to identification**. London, Academic Press, 1986.
- SMITH, C. L. **Digital computer process control**. Scranton, PA, Intext Educational Publishers, 1972.
- SÖDERSTRÖM, T.; STOICA, P. **System identification**. Hemel Hempstead, U.K., Prentice Hall International, 1989.