Capítulo 2

Fundamentos da Teoria de Estimação

2.1 Introdução

Estimação é a determinação de grandezas físicas não observáveis a partir de grandezas mensuráveis. A teoria de estimação compreende basicamente duas classes de problemas, a saber:

a) Identificação Experimental: A identificação de um Modelo de um sistema através de medidas das suas entradas e saídas, também denominada de modelagem "caixa preta".

b) Filtragem: A determinação dos Estados de um sistema a partir de medidas das suas entradas e saídas.

Nesse capítulo serão apresentados métodos para a estimação de parâmetros de um sistema linear monovariável utilizando dados medidos da sua entrada e saída. A estimação de estados será apresentada em outro capítulo. Assume-se conhecida *a priori* a ordem do modelo. Nesse contexto, considera-se a estimação de parâmetros de um modelo linear de tempo discreto descrito pela equação (2.1). O ítem sobre filtragem será desenvolvido no capítulo 5.

$$y_k = \mathbf{u}_k^T \boldsymbol{\theta} + v_k \tag{2.1}$$

Na equação (2.1) $\boldsymbol{\theta} \in \Re^n$ é o vetor de parâmetros desconhecidos, y_k é a saída medida no instante k, $\mathbf{u}_k^T$ é um vetor com elementos conhecidos (mensuráveis), normalmente a entrada do sistema, e v_k é uma perturbação atuando no processo. O modelo anterior pode ser também escrito como:

$$y_k = \theta_0 u_k + \theta_1 u_{k-1} + \ldots + \theta_{m-1} u_{k-m+1} + v_k \tag{2.2}$$

ou

$$y_k = \sum_{i=0}^{m-1} \theta_i u_{k-i} + v_k$$

ou ainda

$$y_k = \mathbf{u}_k^T \boldsymbol{\theta} + v_k \tag{2.3}$$

com

$$\mathbf{u}_k^T = \begin{bmatrix} u_k & u_{k-1} & \cdots & u_{k-m+1} \end{bmatrix}$$

$$\boldsymbol{\theta}^T = \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_{m-1} \end{bmatrix}$$

Nesta equação deve-se obter um estimador $\hat{\boldsymbol{\theta}}$ dos parâmetros desconhecidos $\boldsymbol{\theta}$ a partir das medidas realizadas. Para tanto, vamos supor que sejam realizadas N medidas, conforme descrito a seguir:

$$\begin{aligned} y_1 &= \mathbf{u}_1^T \boldsymbol{\theta} + v_1 \\ y_2 &= \mathbf{u}_2^T \boldsymbol{\theta} + v_2 \\ &\vdots \\ y_N &= \mathbf{u}_N^T \boldsymbol{\theta} + v_N \end{aligned}$$

$$\mathbf{y} = U\boldsymbol{\theta} + \mathbf{v}$$

$$\mathbf{y}^T = \begin{bmatrix} y_1 & y_2 & \cdots & y_N \end{bmatrix}$$

$$\mathbf{v}^T = \begin{bmatrix} v_1 & v_2 & \cdots & v_N \end{bmatrix}$$

$$U = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_N^T \end{bmatrix}$$

2.2 Método dos Mínimos Quadrados

Este método foi tratado de forma independente por Gauss e Legendre e é um dos métodos mais conhecidos e utilizados nas diversas áreas do conhecimento. Gauss, durante observações astronômicas, afirmou que:

"*O valor mais provável de grandezas desconhecidas é o que minimiza a soma dos quadrados da diferença entre o valor medido e o valor calculado, ponderado pelo grau de precisão da medida*".

Seja $\hat{\boldsymbol{\theta}}$ o valor estimado de $\boldsymbol{\theta}$ e $\hat{\mathbf{y}}$ a saída estimada do sistema, dada por:

$$\hat{\mathbf{y}} = U\hat{\boldsymbol{\theta}} \tag{2.4}$$

Sejam os erros entre os vetores de parâmetros desconhecidos e estimados e entre a saída do sistema e a saída do modelo definidos respectivamente por $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ e $\tilde{\mathbf{y}} = \mathbf{y} - \hat{\mathbf{y}}$. Logo, tem-se que:

$$\tilde{\mathbf{y}} = U\tilde{\boldsymbol{\theta}} + \mathbf{v} \tag{2.5}$$

A equação (2.5) é denominada equação do erro. O estimador é obtido minimizando-se:

$$J(\hat{\boldsymbol{\theta}}) = w(1)\tilde{y}_1^2 + w(2)\tilde{y}_2^2 + \cdots + w(N)\tilde{y}_N^2$$

onde $w(i)$ é a ponderação em cada componente do erro, que é função da precisão da medida. Esta equação pode ser rescrita como segue:

$$J(\hat{\boldsymbol{\theta}}) = \tilde{\mathbf{y}}^T W \tilde{\mathbf{y}}$$

onde $\tilde{\mathbf{y}} = [\tilde{y}_1 \ \tilde{y}_2 \ \cdots \ \tilde{y}_N]^T$ e $W = \mathrm{diag}(w(1) \ w(2) \ \cdots \ w(N))$. Deseja-se:

$$\hat{\boldsymbol{\theta}}_M = \arg\min_{\hat{\boldsymbol{\theta}}} J(\hat{\boldsymbol{\theta}}) = \arg\min_{\hat{\boldsymbol{\theta}}} \|\mathbf{y} - U\hat{\boldsymbol{\theta}}\|_W^2$$

sendo que a função $J(\hat{\boldsymbol{\theta}})$ pode ainda ser rescrita como:

$$J(\hat{\boldsymbol{\theta}}) = \mathbf{y}^T W \mathbf{y} - 2\mathbf{y}^T W U \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\theta}}^T U^T W U \hat{\boldsymbol{\theta}}$$

Sabe-se que $\hat{\boldsymbol{\theta}}_M$ deve ser tal que $dJ/d\hat{\boldsymbol{\theta}} = 0$, o que implica:

$$\frac{dJ}{d\hat{\boldsymbol{\theta}}} = -2[\mathbf{y}^T W U]^T + 2U^T W U \hat{\boldsymbol{\theta}} = 0$$

$$[U^T W U]\hat{\boldsymbol{\theta}} = U^T W \mathbf{y} \tag{2.6}$$

que é denominada Equação Normal e produz o seguinte estimador:

$$\hat{\boldsymbol{\theta}}_M = [U^T W U]^{-1} U^T W \mathbf{y} \tag{2.7}$$

A matriz $[U^T W U]^{-1} U^T W$ é denominada matriz Pseudo-Inversa. O estimador (2.7) é denominado Estimador de Markov, Mínimos Quadrados Ponderados (*Weighted Least Squares* – WLS) ou ainda Mínimos Quadrados Generalizado. Quando $W = \sigma^2 I_N$, onde σ é um escalar e I_N é a matriz de identidade $N \times N$, obtém-se o estimador dos Mínimos Quadrados convencional, descrito por:

$$\hat{\boldsymbol{\theta}}_{MQ} = [U^T U]^{-1} U^T \mathbf{y} \tag{2.8}$$

Nesse caso tem-se que $\hat{\boldsymbol{\theta}}_{MQ}$ é seguramente um ponto de mínimo pois a matriz Hessiana:

$$\frac{d^2 J}{d\hat{\boldsymbol{\theta}}^2} = 2U^T W U$$

é definida positiva posto que $W = \sigma^2 I_N$ é definida positiva. Logo, $\hat{\boldsymbol{\theta}}_{MQ}$ em (2.8) é uma solução ótima. Como é única, é uma solução ótima global. No caso geral do estimador em (2.7) a otimalidade requer apenas que a matriz de ponderação W seja definida positiva.

2.2.1 Interpretação Geométrica do Estimador

A equação $\hat{\mathbf{y}} = U\hat{\boldsymbol{\theta}}$ define o subespaço $\mathcal{R}(U) = \hat{\mathcal{Y}}$. Da equação normal para o estimador dos Mínimos Quadrados tem-se que:

$$(U^T U)\hat{\boldsymbol{\theta}} - U^T \mathbf{y} = 0$$

$$U^T(\mathbf{y} - \hat{\mathbf{y}}) = 0$$

$$U^T \tilde{\mathbf{y}} = 0 \tag{2.9}$$

A equação (2.9) mostra que $\tilde{\mathbf{y}}$ pertence ao espaço nulo de U^T, $\mathcal{N}(U^T)$. Como $\mathcal{N}(U^T) = [\mathcal{R}(U)]^{\perp}$ então $\tilde{\mathbf{y}}$ pertence ao complemento ortogonal ($\perp$) de U, isto é, $\tilde{\mathbf{y}}$ é normal ao subespaço $\hat{\mathcal{Y}}$. A propriedade da ortogonalidade do estimador dos mínimos quadrados pode também ser ilustrada graficamente, conforme mostra a figura 2.1.

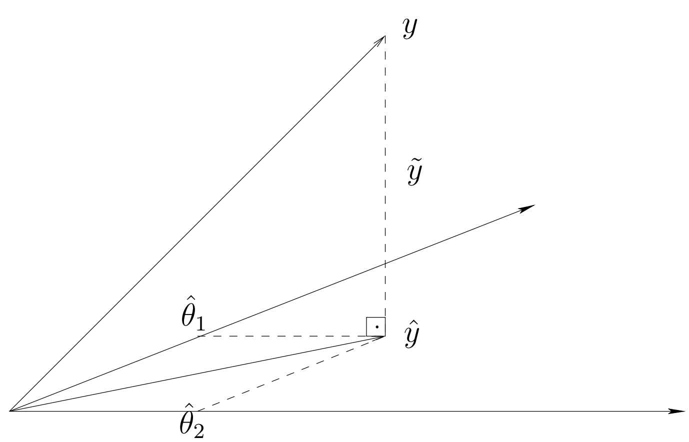

Figura 2.1: Propriedade de Ortogonalidade do Estimador de Mínimos Quadrados.

2.2.2 Exemplos

Exemplo 1

Seja o sistema descrito por:

$$y(t) = \theta_0 + v(t)\ ; \qquad\qquad t = 1, \cdots, N$$

onde $y(t)$ é a saída do sistema e $v(t)$ é uma perturbação com variância $\sigma^2_{v(t)} = 1$. Obter o estimador dos Mínimos Quadrados de θ_0.

Nesse caso tem-se que:

$$U = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \qquad ; \qquad \mathbf{y} = [\, y(1) \cdots y(N)\,]^T$$

$$\hat{\theta}_0 \;=\; [U^T U]^{-1} U^T \mathbf{y} \;=\; \frac{1}{N} \sum_{i=1}^{N} y(i)$$

Portanto, o estimador nesse caso corresponde à média aritmética das medidas. Suponha agora que a variância da perturbação seja dada por $\sigma^2_{v(t)} = \lambda^2_t$. Nesse caso pode-se definir:

$$W = \begin{bmatrix} \lambda_1^2 & 0 & \dots & 0 \\ 0 & \lambda_2^2 & \dots & 0 \\ \vdots & \vdots & \vdots & 0 \\ 0 & \dots & \lambda^2_{N-1} & 0 \\ 0 & 0 & \dots & \lambda^2_N \end{bmatrix}$$

e utilizar o estimador de Markov em (2.7) para obter:

$$\hat{\theta}_0 = \frac{1}{\sum_{j=1}^{N} (1/\lambda_j^2)} \sum_{i=1}^{N} (1/\lambda_i^2) y(i)$$

que é a média ponderada das medidas. Dessa equação pode-se observar que se a medida tem maior precisão (menor variância) então ela possui maior peso no cálculo do estimador.

Exemplo 2

Seja o sistema descrito por:

$$y(t) = \theta_0 u(t) + \theta_1 u(t-1) + v(t)\ ; \qquad t = 1, \cdots, N$$

onde $u(t)$ e $y(t)$ são respectivamente a entrada e a saída do sistema (supostas mensuráveis) e $v(t)$ é uma perturbação. Obter o estimador dos mínimos quadrados de θ_0 e θ_1.

Nesse caso tem-se que:

$$y(t) = \begin{bmatrix} u(t) & u(t-1) \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} + v(t)$$

$$U = \begin{bmatrix} u(1) & u(0) \\ u(2) & u(1) \\ \vdots & \vdots \\ u(N) & u(N-1) \end{bmatrix}$$

e portanto:

$$U^T U = \begin{bmatrix} \sum_{t=1}^{N}(u^2(t)) & \sum_{t=1}^{N}(u(t)u(t-1)) \\ \sum_{t=1}^{N}(u(t-1)u(t)) & \sum_{t=1}^{N}(u^2(t-1)) \end{bmatrix}$$

Observações:

a) A matriz $U^T U$ no caso geral é tal que $U^T U \in \Re^{m\times m}$ onde m é o número de parâmetros desconhecidos.

b) A matriz $U^T U$ é simétrica e seus elementos são somas de "correlações".

c) Para a existência da solução da equação do estimador a matriz $U^T U$ deve ser não singular. Para tanto a entrada $u(t)$ deve apresentar *excitação persistente*, isso é, deve possuir variação suficiente para que o determinante da matriz seja diferente de zero (não nulo).

Considere, por exemplo, uma entrada constante $u(t) = 1$. Nesse caso tem-se:

$$y(t) = (\theta_0 + \theta_1) + v(t)$$

e

$$\frac{1}{N}[U^T U] = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Essa matriz é não singular o que significa que sua inversa não existe. Nesse caso não existe solução única para o problema de estimação dos parâmetros.

Exemplo 3

Considere um sistema estático em que se deseja obter o modelo mais adequado às seguintes medidas:

$$
\begin{array}{lll}
t = 1: & u(1) = 0 & y(1) = 0 \\
t = 2: & u(2) = 1 & y(2) = 0.9 \\
t = 3: & u(3) = 2 & y(3) = 2.1
\end{array}
$$

Considere inicialmente um modelo constante[1], isto é:

$$y(t) = \theta_0$$

Nesse caso tem-se:

$$U = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

e

$$\mathbf{y} = \begin{bmatrix} 0 \\ 0.9 \\ 2.1 \end{bmatrix}$$

O estimador dos Mínimos Quadrados de θ_0 é dado por:

$$\hat{\theta}_0 = [U^T U]^{-1} U^T \mathbf{y}$$

$$\hat{\theta}_0 = 3^{-1}(0 + 0.9 + 2.1) = 1.0$$

que corresponde à média aritmética das medidas. O custo associado à obtenção desse modelo é dado por:

[1] O mais simples!

$$J = \tilde{y}^T \tilde{y} = 1 + 0.1^2 + 1.1^2 = 2.22$$

Vamos a seguir considerar um modelo linear e verificar se ele representa melhor os dados. Seja então:

$$y(t) = \theta_0 + \theta_1 u(t)$$

ou

$$y(t) = \begin{bmatrix} 1 & u(t) \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

Nesse caso a matriz U é dada por:

$$U = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$$

O estimador $\hat{\boldsymbol{\theta}}$ é dado por:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} -0.05 \\ 1.05 \end{bmatrix}$$

O custo associado a esse modelo é $J = 0.015$, que é significativamente menor que aquele associado ao modelo anterior, indicando que esse modelo representa melhor os dados fornecidos.

Considere agora um modelo (não linear) de segunda ordem, isso é:

$$y(t) = \theta_0 + \theta_1 u(t) + \theta_2 u^2(t)$$

que pode ser rescrito como:

$$y(t) = \begin{bmatrix} 1 & u(t) & u^2(t) \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix}$$

$$U = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}$$

Nesse caso $\hat{\boldsymbol{\theta}}$ é dado por:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} 0 \\ 0.75 \\ 0.15 \end{bmatrix}$$

O custo associado a esse modelo é $J = 0.0$ que é o menor valor que pode ser obtido para esses dados. Essa condição ocorre quando o número de medidas é igual ao número de parâmetros a serem estimados e a matriz U é não singular. Na validação do modelo também deve-se incluir uma medida de parcimônia para evitar modelos de ordem elevada (vide seção de validação de modelos).

2.3 Propriedades do Estimador dos Mínimos Quadrados

Definições:

1. O estimador $\hat{\boldsymbol{\theta}} = f(\mathbf{y})$ de $\boldsymbol{\theta}$ diz-se não polarizado se $\hat{\boldsymbol{\theta}}$ possui valor médio igual a $\boldsymbol{\theta}$ qualquer que seja $\boldsymbol{\theta} \in \Re^m$, isto é:

$$\mathcal{E}_{Y/\boldsymbol{\theta}}\{f(\mathbf{y})\} = \boldsymbol{\theta}$$

   onde $\mathcal{E}$ denota a esperança matemática.

2. Um estimador $\hat{\boldsymbol{\theta}} = f(\mathbf{y})$ de $\boldsymbol{\theta}$ diz-se um estimador não polarizado e de variância mínima (*Minimum Variance Unbiased Estimator* – MVUE) se é não polarizado e

$$\mathcal{E}_{Y/\boldsymbol{\theta}}\{(f(\mathbf{y}) - \boldsymbol{\theta})(f(\mathbf{y}) - \boldsymbol{\theta})^T\} \leq \mathcal{E}_{Y/\boldsymbol{\theta}}\{(f^*(\mathbf{y}) - \boldsymbol{\theta})(f^*(\mathbf{y}) - \boldsymbol{\theta})^T\} \quad (2.10)$$

   para qualquer $\boldsymbol{\theta} \in \Re^m$ e qualquer estimador $f^*(\mathbf{y})$ não polarizado.

3. Um estimador $f(\mathbf{y})$ de $\boldsymbol{\theta}$ é o melhor estimador linear não polarizado (*Best Linear Unbiased Estimator* – BLUE) se a equação (2.10) é satisfeita para qualquer $\boldsymbol{\theta} \in \Re^m$ e estimadores f^* da classe de estimadores lineares não polarizados.

Teorema 1 *O estimador de Markov é não polarizado se a perturbação tem média nula ($\mathcal{E}\{\mathbf{v}\} = 0$) e os sinais de entrada e de perturbação são independentes.*

Prova: Da equação do estimador de Markov tem-se que:

$$\hat{\boldsymbol{\theta}} = [U^T W U]^{-1} U^T W [U\boldsymbol{\theta} + \mathbf{v}]$$
$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + [U^T W U]^{-1} U^T W \mathbf{v}$$

$$\mathcal{E}\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta} + \mathcal{E}\{(U^T W U)^{-1} U^T W \mathbf{v}\} \tag{2.11}$$

Se $\mathbf{u}$ e $\mathbf{v}$ são independentes, então a equação (2.11) pode ser rescrita como segue:

$$\mathcal{E}\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta} + \mathcal{E}\{(U^T W U)^{-1} U^T W\}\mathcal{E}\{\mathbf{v}\} = \boldsymbol{\theta}$$

Deve-se notar que a independência entre $\mathbf{u}$ e $\mathbf{v}$ é uma condição suficiente, mas não necessária, pois a partir da equação (2.11) tem-se a seguinte implicação:

$$\mathcal{E}\{(U^T W U)^{-1} U^T W \mathbf{v}\} = 0 \quad \rightarrow \quad \mathcal{E}\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta}$$

Conhecida como condição de ortogonalidade, é mais fraca que a condição de independência, mas mais difícil de ser verificada.

Teorema 2 *Se a perturbação tem média nula ($\mathcal{E}\{\mathbf{v}\} = 0$) e além disso a entrada* $\mathbf{u}$ *é deterministica e a perturbação* $\mathbf{v}$ *é uma variável aleatória com matriz de covariância V então a covariância do erro de estimação do estimador de Markov é dada por:*

$$cov\tilde{\boldsymbol{\theta}} = [U^T W U]^{-1} U^T W V W U [U^T W U]^{-1}$$

onde $V = \mathcal{E}\{\mathbf{v}\,\mathbf{v}^T\}$.

Prova:

$$cov\tilde{\boldsymbol{\theta}} = \mathcal{E}\{[\tilde{\boldsymbol{\theta}} - \mathcal{E}\tilde{\boldsymbol{\theta}}][\tilde{\boldsymbol{\theta}} - \mathcal{E}\tilde{\boldsymbol{\theta}}]^T\}$$

Dado que $\mathcal{E}\tilde{\boldsymbol{\theta}} = \mathcal{E}\{\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\} = \mathcal{E}\{\boldsymbol{\theta}\} - \mathcal{E}\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta} - \boldsymbol{\theta} = 0$, pois $\boldsymbol{\theta}$ é determinístico e o estimador é não polarizado (teorema 1), tem-se:

$$cov\tilde{\boldsymbol{\theta}} = \mathcal{E}\{\tilde{\boldsymbol{\theta}}\,\tilde{\boldsymbol{\theta}}^T\}$$

Do teorema 1 tem-se:

$$\boldsymbol{\theta} - \hat{\boldsymbol{\theta}} = -[U^TWU]^{-1}U^TW\mathbf{v} = \tilde{\boldsymbol{\theta}}$$

e portanto

$$cov\tilde{\boldsymbol{\theta}} = \mathcal{E}\{[U^TWU]^{-1}U^TW\mathbf{v}\mathbf{v}^TWU[U^TWU]^{-1}\}$$

Se a entrada for determinística tem-se que[2]*:*

$$cov\tilde{\boldsymbol{\theta}} = [U^TWU]^{-1}U^TWVWU[U^TWU]^{-1}$$

Quando a perturbação atuando no sistema é dada por $V = \sigma^2 I_N$, isto é, é tal que suas componentes são independentes e identicamente distribuídas (i.i.d.), tem-se que:

$$cov\tilde{\boldsymbol{\theta}} = \sigma^2[U^TU]^{-1}$$

onde σ^2 é um parâmetro da distribuição da medida, que em geral é desconhecido e necessita ser estimado.

Teorema 3 *Um estimador não polarizado de σ^2 é dado por:*

$$\hat{\sigma}^2 = \frac{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}{dim(\mathbf{v}) - dim(\boldsymbol{\theta})} = \frac{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}{N-m} \tag{2.12}$$

Prova:

$$\tilde{\mathbf{y}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - U\hat{\boldsymbol{\theta}}$$

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + [U^TU]^{-1}U^T\mathbf{v} \qquad \textit{(do teorema 1 para } W = I_N\textit{)}$$

$$\tilde{\mathbf{y}} = U\boldsymbol{\theta} + \mathbf{v} - U[\boldsymbol{\theta} + [U^TU]^{-1}U^T\mathbf{v}] = \mathbf{v} - U[U^TU]^{-1}U^T\mathbf{v}$$

$$\tilde{\mathbf{y}} = [I_N - U[U^TU]^{-1}U^T]\mathbf{v}$$

---

[2]Dado que U^TWU é não singular, segue que $([U^TWU]^{-1})^T = ([U^TWU]^T)^{-1} = [U^TWU]^{-1}$.

Definindo $A = I_N - U[U^TU]^{-1}U^T$ *tem-se que* $\tilde{\mathbf{y}} = A\mathbf{v}$*, sendo que a matriz* A *apresenta as propriedades* $A^2 = AA = A$ *e* $A^T = A$*. Portanto:*

$$\mathcal{E}\{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}\} = \mathcal{E}\{\mathbf{v}^T[I_N - U[U^TU]^{-1}U^T][I_N - U[U^TU]^{-1}U^T]\mathbf{v}\} = \mathcal{E}\{\mathbf{v}^TA^2\mathbf{v}\}$$

$$\mathcal{E}\{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}\} = \mathcal{E}\{\mathbf{v}^T[I_N - U(U^TU)^{-1}U^T]\mathbf{v}\} = \mathcal{E}\{\mathbf{v}^TA\mathbf{v}\}$$

Lembrando as seguintes propriedades de matrizes:

$$\mathbf{p}^TP\,\mathbf{p} = tr(P\mathbf{p}\mathbf{p}^T)$$

$$\mathcal{E}\{tr(P)\} = tr(\mathcal{E}\{P\})$$

$$tr(P+Q) = tr(Q+P)$$

$$tr(PQ) = tr(QP)$$

tem-se que:

$$\mathcal{E}\{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}\} = \mathcal{E}\{tr(A\mathbf{v}\mathbf{v}^T)\}$$

$$\mathcal{E}\{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}\} = tr(A\mathcal{E}\{\mathbf{v}\mathbf{v}^T\}) = tr(\sigma^2A)$$

$$\mathcal{E}\{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}\} \;=\; \sigma^2tr(I_N) - \sigma^2tr(U[U^TU]^{-1}U^T) \;=\; \sigma^2(\,N - tr([U^TU]^{-1}U^TU)\,)$$

$$\mathcal{E}\{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}\} = \sigma^2(\,N - tr(I_m)\,) = \sigma^2(N-m)$$

Logo, $\hat{\sigma}^2$ *em (2.12) é um estimador não polarizado de* σ^2 *pois* $\mathcal{E}\{\hat{\sigma}^2\} = \sigma^2$*.*

Teorema 4 *O estimador dos Mínimos Quadrados dado por* $\hat{\boldsymbol{\theta}} = [U^TU]^{-1}U^T\mathbf{y}$ *possui as seguintes propriedades para* $V = \mathcal{E}\{\mathbf{v}\mathbf{v}^T\} = \sigma^2I_N$ *e* $\mathcal{E}\{\mathbf{v}\} = 0$*:*

a) *É uma função linear das medidas;*

b) *É um estimador não polarizado;*

c) *É um estimador BLUE, isto é:*

$$\mathcal{E}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\} \leq \mathcal{E}\{(\boldsymbol{\theta}^\star - \boldsymbol{\theta})(\boldsymbol{\theta}^\star - \boldsymbol{\theta})^T\}$$

onde $\boldsymbol{\theta}^\star$ *é qualquer estimador linear das medidas.*

Prova:

a) $\hat{\boldsymbol{\theta}} = [U^T U]^{-1} U^T \mathbf{y} = C\mathbf{y}$.

b) Já foi demonstrado (teorema 1).

c) Seja um estimador $\boldsymbol{\theta}^\star = C\mathbf{y}$ *tal que* $\mathcal{E}\boldsymbol{\theta}^\star = \boldsymbol{\theta}$ *(não polarizado). Logo, é simples demonstrar que* $\mathcal{E}\{C\mathbf{y}\} = \boldsymbol{\theta}$ *implica* $CU\boldsymbol{\theta} = \boldsymbol{\theta}$, *dado que* $\mathcal{E}\{\mathbf{v}\} = 0$. *A covariância de* $\boldsymbol{\theta}^\star$ *é dada por:*

$$\mathcal{E}\{(\boldsymbol{\theta}^\star - \boldsymbol{\theta})(\boldsymbol{\theta}^\star - \boldsymbol{\theta})^T\} = \mathcal{E}\{(C\mathbf{y} - \boldsymbol{\theta})(C\mathbf{y} - \boldsymbol{\theta})^T\} =$$

$$= \mathcal{E}\{(C\mathbf{y} - CU\boldsymbol{\theta} + CU\boldsymbol{\theta} - \boldsymbol{\theta})(C\mathbf{y} - CU\boldsymbol{\theta} + CU\boldsymbol{\theta} - \boldsymbol{\theta})^T\} =$$

$$= \mathcal{E}\{(C\mathbf{y} - CU\boldsymbol{\theta})(C\mathbf{y} - CU\boldsymbol{\theta})^T\} = C\mathcal{E}\{(\mathbf{y} - U\boldsymbol{\theta})(\mathbf{y} - U\boldsymbol{\theta})^T\}C^T =$$

$$C\mathcal{E}\{\mathbf{v}\mathbf{v}^T\}C^T = CC^T\sigma^2$$

Seja $D = C - (U^T U)^{-1} U^T$. *Então, dado que* $CU\boldsymbol{\theta} = \boldsymbol{\theta}$ *implica* $CU = I_m$, *tem-se que:*

$$DD^T = [C - (U^T U)^{-1} U^T][C - (U^T U)^{-1} U^T]^T$$

$$DD^T = CC^T - CU(U^T U)^{-1} - (U^T U)^{-1} U^T C^T + (U^T U)^{-1}$$

$$DD^T = CC^T - (U^T U)^{-1}$$

Portanto, como $DD^T \geq 0$ *e* $(U^T U)^{-1} > 0$ *(assumindo que a inversa existe), tem-se* $cov\hat{\boldsymbol{\theta}} \leq cov\boldsymbol{\theta}^\star$, *pois é simples demonstrar que* $cov(\hat{\boldsymbol{\theta}}) = cov(\tilde{\boldsymbol{\theta}}) = \sigma^2 [U^T U]^{-1}$ *(teorema 2).*

Teorema 5 *Seja $\mathcal{E}\{\mathbf{v}\} = 0$ e $cov\mathbf{v} = V$. Então o estimador BLUE de $\boldsymbol{\theta}$ é dado por:*

$$\hat{\boldsymbol{\theta}} = [U^T V^{-1} U]^T U^T V^{-1} \mathbf{y}$$

Prova: Como V é uma matriz simétrica definida positiva, tem-se que V pode ser escrita como $V = PP^T$ com P triangular e não singular. Definindo $\mathbf{y}^\star = P^{-1}\mathbf{y}$ tem-se que:

$$\mathbf{y} = U\boldsymbol{\theta} + \mathbf{v}$$

$$P^{-1}\mathbf{y} = P^{-1}U\boldsymbol{\theta} + P^{-1}\mathbf{v}$$

$$\mathbf{y}^\star = U^\star\boldsymbol{\theta} + \mathbf{v}^\star$$

A partir dessa transformação tem-se:

$$cov\mathbf{v}^\star = \mathcal{E}\{[\mathbf{y}^\star - U^\star\boldsymbol{\theta}][\mathbf{y}^\star - U^\star\boldsymbol{\theta}]^T\} = \mathcal{E}\{P^{-1}(\mathbf{y} - U\boldsymbol{\theta})(\mathbf{y} - U\boldsymbol{\theta})^T P^{-T}\}$$

$$cov\mathbf{v}^\star = P^{-1} V P^{-T} = P^{-1} P P^T P^{-T} = I_N$$

sendo $P^{-T} \stackrel{\triangle}{=} (P^{-1})^T = (P^T)^{-1}$, uma vez que P é não singular. Com essa transformação a perturbação passa a satisfazer as condições do teorema 4. Aplicando então o teorema no sistema transformado resulta:

$$\hat{\boldsymbol{\theta}} = [U^{\star T} U^\star]^{-1} U^{\star T} \mathbf{y}^\star$$

$$\hat{\boldsymbol{\theta}} = [U P^{-T} P^{-1} U]^{-1} U^T P^{-T} P^{-1} \mathbf{y} = [U^T V^{-1} U]^{-1} U^T V^{-1} \mathbf{y}$$

O termo P^{-1} é denominado filtro embranquecedor.

Teorema 6 *(Gauss-Markov) Seja $\mathbf{v}$ um vetor aleatório com $\mathcal{E}\{\mathbf{v}\} = 0$ e $\mathcal{E}\{\mathbf{v}\mathbf{v}^T\} = V$. O estimador de variância mímima é obtido minimizando o critério:*

$$J = (\mathbf{y} - U\boldsymbol{\theta})^T W (\mathbf{y} - U\boldsymbol{\theta}) \quad W = V^{-1}$$

2.4 Estimador de Bayes

2.4.1 Definições

O método de estimação apresentado na seção anterior assume que os valores desconhecidos a serem estimados são deterministícos. Nos métodos de estimação de Bayes os parâmetros desconhecidos $\boldsymbol{\theta}$ são assumidos como sendo variáveis aleatórias e dispõe-se de alguma informação sobre a sua função de distribuição de probabilidade, denominada informação *à-priori*. Como será visto os métodos de Bayes obtém o estimador dos parametros desconhecidos realizando uma combinação desta informação inicial com a informação contida nas medidas realizadas. Para tanto assume-se que o vetor de medidas $\boldsymbol{y}_k$ é descrito pela equação (2.13):

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{v}_k) \tag{2.13}$$

onde $\boldsymbol{h}$ é uma função não linear relacionando as variáveis desconhecidas e a perturbação com as medidas. Em geral, $\boldsymbol{h}$ é uma função desconhecidaa.$\boldsymbol{v}_k$ é um vetor de variáveis aleatórias, com função densidade de probabilidade suposta conhecida, representando as perturbações na medida e no sistema. No estimador de Bayes, dada a informação inicial sobre a função densidade de probabilidade(f.d.p) do vetor de variáveis aleatórias $\boldsymbol{\theta}$, isto é, $p(\boldsymbol{\theta})$ deseja-se calcular a f.d.p da variável aleaória $\boldsymbol{\theta}$ quando dispõe-se de um conjunto de medidas da saída do sistema descrito por (2.13). Isto é, deve-se calcular $p(\boldsymbol{\theta}|\boldsymbol{Y})$ onde: $\boldsymbol{Y}^T = \begin{bmatrix} \boldsymbol{y}_1^T & \boldsymbol{y}_2^T & \cdots & \boldsymbol{y}_N^T \end{bmatrix}$. A nova informação sobre o vetor de parâmetros $\boldsymbol{\theta}$ é calculada utilizando o regra de Bayes conforme segue:

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{Y})} \tag{2.14}$$

A solução da equação(2.14) é o estimador de Bayes, onde $p(\boldsymbol{\theta}|\boldsymbol{Y})$ é a densidade *a-posteriori* de $\boldsymbol{\theta}$. Contudo, em geral, calcular a função densidade *a-posteriori* utilizando (2.14) é um problema complexo. Assim procura-se obter informação sobre alguma estatistica desta densidade, e esta informação é denominada estimador ($\hat{\boldsymbol{\theta}}$) de $\boldsymbol{\theta}$. Os diferentes estimadores de Bayes são obtidos otimzando algum criterio função do erro de estimação dado por

$\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$. Para tanto define-se o funcional J conforme segue:

$$J(\hat{\boldsymbol{\theta}}) = \mathcal{E}(R(\tilde{\boldsymbol{\theta}})) \tag{2.15}$$

$$J(\hat{\boldsymbol{\theta}}) = \int R(\tilde{\boldsymbol{\theta}}) p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) d(\boldsymbol{\theta})$$

$$J(\hat{\boldsymbol{\theta}}) = \int \int R(\tilde{\boldsymbol{\theta}}) p_{\boldsymbol{\theta}\boldsymbol{Y}}(\boldsymbol{\theta}, \boldsymbol{Y}) d(\boldsymbol{\theta}) d(\boldsymbol{Y})$$

onde $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ é a densidade de probabilidade marginal de $\boldsymbol{\theta}$, $p_{\boldsymbol{\theta}\boldsymbol{Y}}(\boldsymbol{\theta}, \boldsymbol{Y})$ é a densidade conjunta das variáveis aleatórias $\boldsymbol{\theta}$ e $\boldsymbol{Y}$. O estimador ótimo de Bayes é obtido minimizando-se o critério $J(\hat{\boldsymbol{\theta}})$. Variando-se a função de custo $R(\tilde{\boldsymbol{\theta}})$ obtém-se os diferentes estimadores de Bayes. A seguir apresenta-se os dois estimadores mais comuns citados na literatura.

2.4.2 Estimador de Bayes de Mínimos Quadrados

Seja a função de custo $R(\tilde{\boldsymbol{\theta}}) = \tilde{\boldsymbol{\theta}}^T Q \tilde{\boldsymbol{\theta}}$ onde Q é uma matriz simétrica semi-definida positiva. Neste caso o critério a ser minimizado pode ser escrito como:

$$J(\hat{\boldsymbol{\theta}}) = \int \int \tilde{\boldsymbol{\theta}}^T Q \tilde{\boldsymbol{\theta}} p_{\boldsymbol{\theta}\boldsymbol{Y}}(\boldsymbol{\theta}, \boldsymbol{Y}) d(\boldsymbol{\theta}) d(\boldsymbol{Y})$$

Utilizando a definição de função de densidade de probabilidade condicional pode-se reescrever a equação acima como:

$$J(\hat{\boldsymbol{\theta}}) = \int \{ \int \tilde{\boldsymbol{\theta}}^T Q \tilde{\boldsymbol{\theta}} p_{\boldsymbol{\theta}|\boldsymbol{Y}}(\boldsymbol{\theta}|\boldsymbol{Y}) d(\boldsymbol{\theta}) \} p_{\boldsymbol{Y}}(\boldsymbol{Y}) d(\boldsymbol{Y})$$

Como $p_{\boldsymbol{Y}}(\boldsymbol{Y})$ é sempre positiva, minimizar J é equivalente a minimizar a integral interna da equação anterior, isto é, deve-se minimizar

$$J_{BMQ}(\hat{\boldsymbol{\theta}}) = \int \tilde{\boldsymbol{\theta}}^T Q \tilde{\boldsymbol{\theta}} p_{\boldsymbol{\theta}|\boldsymbol{Y}}(\boldsymbol{\theta}|\boldsymbol{Y}) d(\boldsymbol{\theta})$$

A condição necessária para a minimização de J_{BMQ} é dada por:

$$\left.\frac{\partial J_{BMQ}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right|_{\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{BMQ}} = -2Q \int_{-\infty}^{\infty} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) p_{\boldsymbol{\theta}|\boldsymbol{Y}}(\boldsymbol{\theta}|\boldsymbol{Y}) d(\boldsymbol{\theta}) \Big|_{\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{BMQ}} = 0$$

$$\hat{\boldsymbol{\theta}}_{BMQ} \int_{-\infty}^{\infty} p_{\boldsymbol{\theta}|\boldsymbol{Y}}(\boldsymbol{\theta}|\boldsymbol{Y}) d(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \boldsymbol{\theta} p_{\boldsymbol{\theta}|\boldsymbol{Y}}(\boldsymbol{\theta}|\boldsymbol{Y}) d(\boldsymbol{\theta})$$

$$\hat{\boldsymbol{\theta}}_{BMQ} = \int_{-\infty}^{\infty} \boldsymbol{\theta} p_{\boldsymbol{\theta}|\boldsymbol{Y}}(\boldsymbol{\theta}|\boldsymbol{Y}) d(\boldsymbol{\theta})$$

$$\hat{\boldsymbol{\theta}}_{BMQ} = \mathcal{E}(\boldsymbol{\theta}|\boldsymbol{Y})$$

Esta solução é um ponto de mínimo pois tem-se que:

$$\frac{\partial^2 J_{BMQ}(\hat{\boldsymbol{\theta}})}{\partial^2 \hat{\boldsymbol{\theta}}} = 2Q > 0$$

2.4.3 Propriedades do Estimador de Bayes de Mínimos Quadrados

a) É um operador linear

Prova: Seja o estimador de Bayes de Mínimos Quadrados dado por $\hat{\boldsymbol{\theta}}_{BMQ}$, a matriz A e o vetor $\boldsymbol{b}$ com dimensões apropriadas, então:

$$\mathcal{E}((A\hat{\boldsymbol{\theta}}_{BMQ} + \boldsymbol{b})|\boldsymbol{Y}) = A\mathcal{E}(\boldsymbol{\theta}|\boldsymbol{Y}) + \boldsymbol{b} = A\hat{\boldsymbol{\theta}}_{BMQ} + \boldsymbol{b}$$

Sejam os estimadores de Bayes de variância mínima $\hat{\boldsymbol{\theta}}_{BMQ}$ e $\hat{\boldsymbol{\alpha}}_{BMQ}$ com a mesma dimensão,então:

$$\mathcal{E}((\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\alpha}})|\boldsymbol{Y}) = \mathcal{E}(\hat{\boldsymbol{\theta}}|\boldsymbol{Y}) + \mathcal{E}(\hat{\boldsymbol{\alpha}}|\boldsymbol{Y}) = \hat{\boldsymbol{\theta}}_{BMQ} + \hat{\boldsymbol{\alpha}}_{BMQ}$$

b) É um estimador não polarizado

Prova:

$$\mathcal{E}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{BMQ}) == \mathcal{E}(\boldsymbol{\theta}) - \mathcal{E}(\mathcal{E}(\boldsymbol{\theta}|\boldsymbol{Y})) = \mathcal{E}(\boldsymbol{\theta}) - \mathcal{E}(\boldsymbol{\theta}) = 0$$

c) O erro de estimação $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{BMQ}$ é ortogonal a qualquer função $\boldsymbol{g}(\boldsymbol{Y})$.

Prova: Deve-se provar que:

$$\mathcal{E}\{\boldsymbol{g}(\boldsymbol{Y})\tilde{\boldsymbol{\theta}}^T\} = 0$$

Utilizando a propriedade de esperança matemática deve-se provar que:

$$\mathcal{E}\{\boldsymbol{g}(\boldsymbol{Y})\tilde{\boldsymbol{\theta}}^T|\boldsymbol{Y}\} = 0$$

Para cada valor $\boldsymbol{y}$ de $\boldsymbol{Y}$ tem-se que:

$$\mathcal{E}\{\boldsymbol{g}(\boldsymbol{Y})\tilde{\boldsymbol{\theta}}^T|\boldsymbol{Y} = \boldsymbol{y}\} = \mathcal{E}\{\boldsymbol{g}(\boldsymbol{Y})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{BMQ})^T|\boldsymbol{Y} = \boldsymbol{y}\} =$$

$$= \boldsymbol{g}(\boldsymbol{y})\mathcal{E}\{(\boldsymbol{\theta}^T - \hat{\boldsymbol{\theta}}^T_{BMQ})|\boldsymbol{Y} = \boldsymbol{y}\}$$

$$= \boldsymbol{g}(\boldsymbol{y})\{\hat{\boldsymbol{\theta}}^T_{BMQ} - \hat{\boldsymbol{\theta}}^T_{BMQ}\} = 0$$

2.4.4 Estimador de máximo à-posteriori(MAP)

Neste caso a função de custo é selecionada de forma a ponderar igualmente os erros de estimação maiores que um valor infinetesimal. Assumindo, sem perda de generalidade, que $\boldsymbol{\theta} \in \boldsymbol{R}$, tem-se o seguinte critério de custo:

$$R(\tilde{\boldsymbol{\theta}}) = \begin{cases} 0 & \text{se } \tilde{\boldsymbol{\theta}} < \varepsilon \\ 1 & \text{se } \tilde{\boldsymbol{\theta}} \geq \varepsilon \end{cases}$$

com $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$e $\varepsilon \to 0$. Assim tem-se que:

$$J(\hat{\boldsymbol{\theta}}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(\tilde{\boldsymbol{\theta}}) p_{\boldsymbol{\theta}|\boldsymbol{Y}}(\boldsymbol{\theta}|\boldsymbol{Y}) d(\boldsymbol{\theta}) p_{\boldsymbol{Y}}(\boldsymbol{Y}) d(\boldsymbol{Y})$$

$$J(\hat{\boldsymbol{\theta}}) = \int_{-\infty}^{\infty} \{1 - \int_{\hat{\boldsymbol{\theta}}-\varepsilon}^{\hat{\boldsymbol{\theta}}+\varepsilon} p_{\boldsymbol{\theta}|\boldsymbol{Y}}(\boldsymbol{\theta}|\boldsymbol{Y}) d(\boldsymbol{\theta})\} p_{\boldsymbol{Y}}(\boldsymbol{Y}) d(\boldsymbol{Y})$$

Como $p_{\boldsymbol{Y}}(\boldsymbol{Y})$ é positiva, para minimizar $J(\hat{\boldsymbol{\theta}})$ em relação a $\hat{\boldsymbol{\theta}}$ deve-se maximizar

$$\int_{\hat{\boldsymbol{\theta}}-\varepsilon}^{\hat{\boldsymbol{\theta}}+\varepsilon} p_{\boldsymbol{\theta}|\boldsymbol{Y}}(\boldsymbol{\theta}|\boldsymbol{Y}) d(\boldsymbol{\theta})$$

Quando $\varepsilon \to 0$ esta função é maximizada para o valor de $\hat{\boldsymbol{\theta}}$ que maximiza $p_{\boldsymbol{\theta}|\boldsymbol{Y}}(\boldsymbol{\theta}|\boldsymbol{Y}) d(\boldsymbol{\theta})$. Logo

$$\hat{\boldsymbol{\theta}}_{MAP} = \max_{\boldsymbol{\theta}} p_{\boldsymbol{\theta}|\boldsymbol{Y}}(\boldsymbol{\theta}|\boldsymbol{Y})$$

2.5 Estimador de Máxima Verossimilhança

Neste caso assume-se que a saída do sistema($\boldsymbol{y}$) pode ser descrita por um elemento de uma família de funções densidades de probabilidade $p(.|\boldsymbol{\theta})$ onde $\boldsymbol{\theta}$ é o vetor de parâmetros desconhecidos. Assim para uma medida $\boldsymbol{y} = \boldsymbol{y_{med}}$ a função $p(\boldsymbol{y} = \boldsymbol{y_{med}}|\boldsymbol{\theta})$ é uma função de $\boldsymbol{\theta}$ denominada função de verossimilhança.

Definição: O estimador de máxima verossimilhança é uma função $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ tal que:

$$\hat{\boldsymbol{\theta}}(\boldsymbol{y}) = \max_{\boldsymbol{\theta}} p(\boldsymbol{y}|\boldsymbol{\theta}) \tag{2.16}$$

Desigualdade de Cramer-Rao

Teorema 7 *Desigualdade de Cramer-Rao*
Seja $p(\boldsymbol{y}, \boldsymbol{\theta})$ e a matriz $M_{\boldsymbol{\theta}} = \{m_{ij}\}$ dada por:

$$m_{ij} = \mathcal{E}_{\boldsymbol{\theta}}\{\frac{\partial \log p(\boldsymbol{y}, \boldsymbol{\theta})}{\partial \theta_i}\frac{\partial \log p(\boldsymbol{y}, \boldsymbol{\theta})}{\partial \theta_j}\} \tag{2.17}$$

com $\boldsymbol{\theta}^T = [\theta_1\theta_2 \cdots \theta_p]$. Se $M_{\boldsymbol{\theta}}$ é uma matriz não singular e seja $\boldsymbol{g}(\boldsymbol{y})$ um estimador não polarizado de $\boldsymbol{\theta}$,então:

$$cov_{\boldsymbol{\theta}}\boldsymbol{g}(\boldsymbol{y}) \geq M_{\boldsymbol{\theta}}^{-1}$$

$M_{\boldsymbol{\theta}}$ é denominada matriz de informação de Fisher.

Definições

a) Estimador Eficaz

Um estimador é eficaz quando é válido o sinal de igualdade na desigualdade de Cramer-Rao.

Exemplo: Eficiência do Estimador dos Mínimos Quadrados.Seja um sistema descrito por:

$$y = U\theta + v \quad p(v) = N(0, \sigma^2 I) \;\; y \in \Re^N$$

Nesse caso a função distribuição de probabilidade condicional $p(y/\theta)$ é dada por:

$$p(y/\theta) =$$
$$= \frac{\exp(-0.5(y-U\theta)^T(\sigma^2 I_N)^{-1}(y-U\theta))}{(2\pi)^{N/2} det[\sigma^2 I_N]^{0.5}}$$

Como $det([\sigma^2 I_N]^{0.5}) = (\sigma^2)^{N/2}$

$$\ln p(y/\theta) = ln(2\pi)^{-N/2} +$$

$$+ln(\sigma^2)^{-N/2} - \frac{(y-U\theta)^T((y-U\theta)}{2\sigma^2}$$

$$\ln p(y/\theta) = -\frac{N}{2} ln(2\pi) - \frac{N}{2} ln(\sigma^2) - \frac{(y-U\theta)^T((y-U\theta)}{2\sigma^2}$$

e

$$\frac{\partial \ln p(y/\theta)}{\partial\theta} = \frac{\partial[y^Ty - \theta^T U^T \underline{y} - y^T U\theta + \theta^T U^T U\theta]}{2\sigma^2\partial\theta}$$

$$= \frac{-U^Ty - U^Ty + (U^TU + U^TU)\theta}{2\sigma^2}$$

$$\frac{\partial \ln p(y/\theta)}{\partial\theta} = \frac{[-2U^T(y-U\theta)]}{2\sigma^2}$$

e

$$I = \mathcal{E}\{\frac{\partial \ln p(y/\theta)}{\partial\theta}\}\{\frac{\partial \ln p(y/\theta)}{\partial\theta}\}^T =$$

$$= \frac{1}{\sigma^4}\mathcal{E}\{U^T(y-U\theta)(y-U\theta)^TU\}$$

$$= \frac{1}{\sigma^4}U^T\mathcal{E}vv^TU$$

$$= \frac{U^TU}{\sigma^2}$$

e

$$cov\hat{\theta} \geq \mathcal{E}[(U^TU)^{-1}]\sigma^2 = cov\hat{\theta}_{MQ}$$

que é a covariância do estimador dos mínimos quadrados. Portanto o estimador dos mínimos quadrados é eficiente. Este resultado mostra que o estimador dos Mínimos quadrados apresenta a menor variância entre todos os estimadores lineares e não lineares não polarizados, quando a perturbação for Gaussiana e o sistema puder ser descrito pelo modelo $y(t) = u^T\theta + v(t)$.

Teorema 8 *A condição necessária e suficiente para que um estimador não polarizado seja eficaz é que:*

$$M(\theta)[\hat{\theta}_{ef} - \theta] = [\frac{\partial \ln p(y:\theta)}{\partial \theta}]^T$$

onde $M(\theta)$, independente de y, é a matriz de informação de Fisher.(Obs: assume-se, sem perda de generalidade, que $\theta \in R$.

b) Estimador Consistente

Um Estimador é consistente se:

$$\lim_{n \to \infty} P\{\|\hat{\theta}_n - \theta\|^2 = 0\} \to 1$$

Teorema 9 *Se existir um estimador $f(y)$ não polarizado de θ que satisfaz o limite de Cramer-Rao, isto é, ele é eficaz, então este estimador também é o estimador de máxima verossimilhança.*

Teorema 10 *Seja*

$$y = U\theta + v \quad p(v) = N(0, V) \;\; y \in \Re^m$$

então o estimador de máxima verossimilhança de θé dado por:

$$\hat{\theta}_{ML} = [U^T V^{-1} U]^{-1} U^T V^{-1} y = \hat{\theta}_{Markov}$$

Teorema 11 *O estimador de máxima verossimilhança é consistente e normal com densidade dada por $N(\theta, M^{-1})$*

2.5.1 Aplicação em Sistemas Lineares com Ruído Gaussiano

Seja

$$y = Hx + v$$

$$p_0(x) = N(x_0, X_0) \text{ e } p_v(v) = N(0, V)$$

$$\hat{x} = \mathcal{E}(x/y) = x_0 + C_{xy} Y^{-1} (y - \bar{y})$$

$$C_{x/y} = X - C_{xy}Y^{-1}C_{xy}$$
$$y \sim N(Hx_0, HX_0H^T + V)$$
$$\mathcal{E}(y-\bar{y})(y-\bar{y})^T = \mathcal{E}((Hx+v-Hx_0)(Hx+v-Hx_0)^T) =$$
$$H\mathcal{E}((x-x_0)(x-x_0)^T)H^T + \mathcal{E}vv^T + H\mathcal{E}(x-x_0)\mathcal{E}v^T +$$
$$+\mathcal{E}v\mathcal{E}(x-x_0)^TH^T =$$
$$= HX_0H^T + V$$
$$C_{yx} = \mathcal{E}(y-\bar{y})(x-x_0)^T = HX_0$$
$$C_{xy} = X_0H^T$$
$$\hat{x} = x_0 + X_0H^T(V + HX_0H^T)^{-1}(y - Hx_0)$$
$$C_{x/y} = X_0 - X_0H^T(HX_0H^T + V)^{-1}HX_0$$

2.6 Estimador Linear Seqüencial

Seja o sistema descrito por:

$$y = \Theta_1 u_1 + \Theta_2 u_2 + \ldots + \Theta_m u_m + v$$
$$y = u^T\Theta + v$$
$$\Theta^T = [\Theta_1\Theta_2\ldots\Theta_m]$$
$$u^T = [u_1u_2\ldots u_m] \;\; v \sim N(0, \sigma_k^2)$$

$$\begin{aligned} y_1 &= u_1^T\Theta + v_1 \\ y_2 &= u_2^T\Theta + v_2 \\ &\vdots \\ y_k &= u_k^T\Theta + v_k \end{aligned}$$

$$y_k = U_k\Theta + v$$
$$U_k = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_k^T \end{bmatrix} \quad V_k = \begin{bmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \sigma_k^2 \end{bmatrix}$$

$$\hat{\Theta}_k = [U_k^T V_k^{-1} U_k]^{-1} U_k^T V_k^{-1} y_k$$

onde o índice k representa que o estimador é calculado com as medidas disponíveis até o instante k. Agora deseja-se obter o estimador quando realiza-se uma nova medida, isto é, deseja-se:

$$\hat{\Theta}_{k+1} = f(\hat{\Theta}_k, y_{k+1})$$
$$\hat{\Theta}_{k+1} = [U_{k+1}^T V_{k+1}^{-1} U_{k+1}]^{-1} U_{k+1}^T V_{k+1}^{-1} y_{k+1}$$

Tem-se que:

$$[U_{k+1}^T V_{k+1}^{-1} U_{k+1}]^{-1} = \left[[U_k^T \vdots u_{k+1}] \left[\begin{array}{c|c} V_k & 0 \\ \hline 0 & \sigma_{k+1}^2 \end{array} \right]^{-1} \left[\begin{array}{c} U_k \\ \cdots \\ u_{k+1}^T \end{array} \right] \right]^{-1}$$
$$= [\ U_k^T \quad V_k^{-1} \quad U_k \ + (1/\sigma_{k+1}) u_{k+1} u_{k+1}^T]^{-1}$$

Seja:

$$P_{k+1} = [\ U_{k+1}^T \quad V_{k+1}^{-1} \quad U_{k+1}\ ]^{-1}$$
$$P_{k+1} = [P_k^{-1} + (1/\sigma_{k+1}^2) u_{k+1} u_{k+1}^T]^{-1}$$

Propriedade:

$$[X^{-1} + H^T V^{-1} H]^{-1} = X - XH^T [HXH^T + V]^{-1} HX$$

Seja $X = P_k^{-1}$; $H^T = u_{k+1}$ e $V^{-1} = 1/\sigma_{k+1}^2$ tem-se que:

$$P_{k+1} = P_k - P_k u_{k+1} [u_{k+1}^T P_k u_{k+1} + \sigma_{k+1}^2]^{-1} u_{k+1}^T P_k \tag{2.18}$$

$$\hat{\Theta}_{k+1} = P_{k+1} [U_k^T u_{k+1}] \left[\begin{array}{cc} V_k & 0 \\ 0 & \sigma_{k+1}^2 \end{array} \right]^{-1} \left[\begin{array}{c} \underline{y}_k \\ y_{k+1} \end{array} \right]$$
$$\hat{\Theta}_{k+1} = P_{k+1} [\ U_k^T V_k^{-1} \underline{y}_k \ + u_{k+1} y_{k+1} / \sigma_{k+1}^2]$$

Que resulta em:

$$\hat{\Theta}_{k+1} = \hat{\Theta}_k + K_k \varepsilon_{k+1}$$
$$K_k = P_k u_{k+1} / (\sigma_{k+1}^2 + u_{k+1}^T P_k u_{k+1})$$
$$P_{k+1} = P_k - P_k u_{k+1} u_{k+1}^T P_k / (\sigma_{k+1}^2 + u_{k+1}^T P_k u_{k+1})$$
$$\varepsilon_{k+1} = y_{k+1} - u_{k+1}^T \hat{\Theta}_k$$

O par $(\hat{\Theta}_k P_k)$ representa o menor número de variáveis caracterizando os dados de entrada-saída que devem ser atualizados a cada nova medida.