

# Capítulo 4

## Estimação de Parâmetros

Antes de começarmos a discutir o problema e o procedimento de estimação de parâmetros, é conveniente retornar à Seção 1.1 e à Figura 1.1, para relembrar o processo clássico de construção do conhecimento científico. As observações experimentais, quando inseridas no contexto científico, têm como objetivo fundamental permitir que se compreendam os vínculos que existem (e se de fato existem) entre as diversas variáveis que compõem um problema. O puro exercício de observação e armazenamento de dados experimentais não constitui um processo científico. Para que o processo científico se ponha em marcha, é necessário estabelecer relações, propor explicações, construir e testar teorias. Como já discutido na Seção 1.1, é muito comum ainda hoje ouvir discussões sobre a natureza do trabalho científico, classificando os pesquisadores ora como experimentalistas, ora como teóricos. Essa discussão é vazia e fútil! Não aceite ser classificado dessa forma! Todo bom pesquisador exercita o lado experimental (que fornece a matéria-prima básica para interpretação da realidade e validação de teorias e modelos) e o lado teórico (que permite estabelecer as relações entre as diversas variáveis do problema, tornando assim possível a compreensão dos eventos passados e a previsão dos eventos futuros). É óbvio que certas pessoas têm mais afinidade com o computador que com a chave de fendas, e vice-versa. É por isso que o trabalho conjunto e em equipe é tão importante para o bom desenvolvimento da pesquisa científica!

Nesse contexto, as técnicas de estimação de parâmetros constituem as ferramentas básicas para estabelecimento e interpretação dos vínculos existentes entre as diversas variáveis de um problema. Como discutido nos capítulos seguintes, os procedimentos de estimação de parâmetros são as ferramentas que tornam possível a interpretação qualitativa e quantitativa dos dados experimentais, tornando também possível a discriminação das variáveis relevantes de um problema, a construção de modelos preditivos, a simulação e projeto de processos, etc. Dessa forma, os procedimentos de estimação de parâmetros constituem a ponte que conecta as observações experimentais à interpretação teórica e quantitativa do problema. Será surpreendente para muitos descobrir ao longo dos próximos capítulos que toda a atividade experimental científica desemboca em procedimentos de estimação de parâmetros, mesmo quando essa atividade não está sendo aparentemente realizada. Mais surpreendentemente ainda será descobrir que toda metodologia de planejamento experimental encontra-se intimamente ligada aos procedimentos de estimação de parâmetros. Afinal, é para isso mesmo que os experimentos são feitos! Por isso, preste muita atenção nas discussões que se seguem!

## 4.1. Modelos, Modelagem e Simulação

Mas, afinal, o que é um modelo? Podemos definir um **modelo** como uma estrutura que tenta descrever de forma aproximada a realidade, baseado em um conjunto de observações experimentais. Um modelo é, portanto, uma estrutura que estabelece vínculo entre variáveis distintas e permite explicar os eventos passados e prever de alguma forma o comportamento do sistema em condições experimentais ainda não realizadas. Se, dadas certas condições experimentais, um observador é capaz de fazer algum tipo de previsão sobre o resultado futuro a ser ainda observado, então o observador dispõe de um modelo.

Como já discutido na Seção 1.1, é importante lembrar que o modelo não deve ser jamais confundido com a realidade. **O modelo é apenas uma tentativa de explicar a realidade**, baseado nas observações disponíveis e em um conjunto de hipóteses admitidas pelo pesquisador. Nenhum modelo é capaz de descrever a realidade completamente porque todo dado experimental é corrompido por erros de medida e porque nem todas as variáveis podem ser controladas e/ou medidas precisamente durante os testes experimentais, como já discutido nos capítulos anteriores. Por isso, o bom pesquisador não tem apego a qualquer modelo e está pronto a modificá-lo sempre que uma nova observação experimental confiável (ou seja, que pode ser reproduzida, respeitados os limites impostos pela incerteza experimental) não pode ser explicada pela estrutura original. Darwin e Einstein são gênios da humanidade exatamente porque tiveram a coragem de reinterpretar a realidade sugerida por um novo conjunto de dados experimentais. E o modelo evolucionário e da relatividade são ainda considerados bons modelos porque as novas observações experimentais puderam ser previstas e explicadas de forma adequada com essas estruturas teóricas. Mas quem sabe o que nos reserva o futuro? É claro que não é fácil propor mudanças em um modelo muito testado e amplamente aceito pela comunidade científica. Por isso, há que se ter cuidado com a interpretação dos dados experimentais. Mas há também que se ter a coragem de propor novas interpretações da realidade. Não pode haver oportunidade maior para um pesquisador que um dado experimental que não pode ser explicado pelo modelo!

O modelo pode ser apresentado de várias formas, a depender dos propósitos do pesquisador, das limitações das observações experimentais e da complexidade do fenômeno investigado. Algumas dessas formas são apresentadas a seguir.

### 4.1.1. O Modelo Matemático

O modelo matemático é aquele que propõe que as relações entre as diversas variáveis de um problema podem ser descritas de forma matemática precisa. Por exemplo:

$$y = 5x + 4z^2 + 1$$

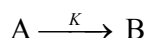
$$\frac{d^2y}{dx^2} + \frac{1}{Pe} \frac{dy}{dx} = Da y^2, y(0) = 0, \left. \frac{dy}{dx} \right|_{x=L} = 0$$

No primeiro caso, a relação é explícita: dados os valores das variáveis  $x$  e  $z$ , é possível calcular diretamente o valor da variável  $y$ . No segundo caso, a relação é implícita: dados

os valores de  $Pe$ ,  $Da$ ,  $L$  e um certo valor de  $x$ , é necessário primeiramente resolver a equação para achar o valor de  $y$ . (Muitas vezes, a resolução do modelo só pode ser feita com o auxílio de métodos numéricos sofisticados.) De qualquer forma, em ambos os casos as relações matemáticas são bem estabelecidas.

A grande vantagem dos modelos matemáticos é que eles podem ser usados para fazer previsões quantitativas sobre o comportamento futuro do sistema estudado. Modelos matemáticos são, portanto, ferramentas fundamentais das ciências exatas, já que são eles que permitem o projeto de novos processos e equipamentos.

**Exemplo 4.1** - Suponha que um vaso de reação tem que ser projetado para conduzir a reação



onde  $K$  é a constante de velocidade da reação. Suponha que o reagente  $A$  é fornecido em concentração conhecida  $C_{A0}$  por uma empresa já estabelecida no mercado e será transformado no produto  $B$  dentro do vaso de volume desconhecido  $V$ . O tempo total da batelada é  $t_B$ , sendo que é necessário, além de promover a reação, também descarregar, limpar e carregar o vaso entre bateladas sucessivas. Esses tempos são representados respectivamente por  $t_R$ ,  $t_D$ ,  $t_L$  e  $t_C$ . Assim, para fazer o projeto da unidade, é necessário descrever primeiramente o lucro da empreitada. A função lucro  $L$  pode ser escrita como

$$L(V, t_R) = N_B V (C_{A0} - C_{Af}(t_R)) \$_B - N_B V C_{A0} \$_A - N_B (V C_{Af}(t_R))^m \$_O - V^n \$_I$$

onde  $N_B$  é o número total de bateladas realizadas ao longo da vida útil do equipamento  $\Delta t$ , dado por

$$N_B = \frac{\Delta t}{(t_R + t_D + t_L + t_C)}$$

e  $\$_B$ ,  $\$_A$ ,  $\$_O$  e  $\$_I$  são respectivamente os preços de mercado para o produto, o reagente, a operação (que aumenta com o teor de reagente no final da batelada, por causa da necessária purificação do produto) e o investimento (que aumenta com o aumento das dimensões do equipamento).  $C_{Af}(t_R)$  é o teor residual de  $A$  no produto final.

O primeiro termo da equação do lucro representa os ganhos obtidos com a venda do produto  $B$ ; o segundo termo, os custos devidos à compra do reagente  $A$ ; o terceiro, os custos operacionais do processo; e o quarto, os custos do investimento. Quanto maior o tempo de reação, menor o teor residual final de  $A$ , maior a quantidade de produto  $B$  e menores os custos operacionais. No entanto, quanto maior  $t_R$ , menor o número de bateladas produzidas ao longo da vida útil do equipamento. (Admite-se que  $t_D$ ,  $t_L$  e  $t_C$  são constantes.) Quanto maior o volume  $V$  do reator, maiores são as quantidades produzidas do produto  $B$ , mas também maiores são os custos operacionais e de investimento. Por isso, deve haver um ponto ótimo, ou de máximo lucro. O projeto consiste em achar esse ponto de máximo lucro.

O ponto de máximo pode ser encontrado fazendo-se

$$\frac{\partial L}{\partial V} = f_1(V, t_R) = 0$$

$$N_B (C_{A0} - C_{Af}(t_R)) \$_B - N_B C_{A0} \$_A - \frac{m N_B}{V} (V C_{Af}(t_R))^m \$_O - n V^{n-1} \$_I = 0$$

e

$$\frac{\partial L}{\partial t_R} = f_2(V, t_R) = 0$$

$$\begin{aligned} \frac{\partial N_B}{\partial t_R} \left[ V (C_{A0} - C_{Af}(t_R)) \$_B - V C_{A0} \$_A - (V C_{Af}(t_R))^m \$_O \right] - \\ - N_B V \frac{\partial C_{Af}}{\partial t_R} \$_B - \frac{m N_B}{C_{Af}(t_R)} (V C_{Af}(t_R))^m \$_O = 0 \end{aligned}$$

Repare que as duas equações acima dependem de duas variáveis ( $V$  e  $t_R$ ) e, portanto, podem ser resolvidas usando-se a técnica matemática adequada para resolução de equações algébricas, como a técnica de Newton-Raphson. Contudo, é preciso nesse caso definir quem é a função  $C_{Af}(t_R)$ ; ou em outras palavras, como a concentração de A no reator depende do tempo de reação. Ou ainda, é necessário definir o modelo matemático que descreve de forma apropriada a evolução da reação no tempo. Admitindo-se que a reação é de primeira ordem, então:

$$\frac{dC_A}{dt} = -K C_A, \quad C_A(0) = C_{A0}$$

de forma que

$$C_{Af}(t_R) = C_{A0} \exp(-K t_R)$$

Como saber se a reação é de primeira ordem ou que a relação acima é válida? A resposta correta é: fazendo experimentos e construindo o modelo. Como saber o valor de  $K$ ? A resposta correta é: estimando parâmetros. O engenheiro que usa a informação acima para fazer o projeto do reator nem sempre percebe que é usuário dos procedimentos de modelagem e estimação de parâmetros já executados por um outro observador. Sem o modelo e sem o parâmetro, o projeto bem embasado é impossível.

#### 4.1.2. O Modelo Conceitual

O modelo conceitual é aquele que estabelece vínculos qualitativos entre as várias variáveis de um problema, sem que se estabeleçam necessariamente vínculos matemáticos quantitativos precisos. Esses modelos são muito utilizados em campos da ciência onde a mensuração das variáveis é complexa ou impossível, como na área de Ciências Humanas, e como protótipo de modelos matemáticos precisos, a serem desenvolvidos *a posteriori*.

Um exemplo clássico de modelo conceitual é o bem conhecido “Complexo de Édipo”, usado para explicar as relações que se estabelecem entre pais e filhos na Psicologia. Admite-se que os filhos homens disputam a atenção da mãe com o pai e que a maneira com que se desenrola essa disputa pode resultar na formação de pessoas adultas saudáveis ou no desenvolvimento de sérios desvios de comportamento. Repare que a mensuração das grandezas envolvidas é virtualmente impossível nesse caso, impedindo a construção de um modelo matemático preciso do fenômeno. Isso não diminui, no entanto, a importância do modelo nem impede o seu uso para a solução de problemas reais da Psicologia.

Um modelo conceitual extremamente importante na Engenharia é a idéia de que a massa, a energia e a quantidade de movimento do universo se conservam. (O primeiro modelo é chamado de Princípio de Lavoisier; o segundo modelo é chamado de Primeira Lei da Termodinâmica; o terceiro modelo é constituído pelas Leis Fundamentais de Newton para o Movimento.) Esse é o ponto de partida para a quase totalidade dos modelos matemáticos fenomenológicos que se estabelecem nos vários ramos da Engenharia. Assim, partindo-se do pressuposto de que massa, energia e quantidade de movimento se conservam, é possível estabelecer vínculos matemáticos quantitativos precisos entre várias variáveis envolvidas em um fenômeno físico particular.

**Exemplo 4.2** - Suponha que uma massa  $M_0$  de água é colocada em uma caixa d'água, com comprimento  $L$ , altura  $H$  e largura  $W$ . Suponha que um medidor de nível (bóia) fornece a altura  $h$  de água no interior da caixa d'água em qualquer instante de tempo. É possível com essas informações acompanhar a quantidade de água consumida?

Sabendo-se que a densidade da água é  $\rho = 1000 \text{ kg/m}^3$ , a massa de água contida no interior da caixa d'água pode ser obtida em qualquer instante de tempo como

$$M = \rho L W h$$

Admitindo-se que a massa total do universo (nesse caso, a caixa d'água) permanece constante, é possível escrever

$$M_0 = M + M_C$$

onde  $M_C$  é a massa total de água consumida, já que a água não poderia desaparecer como num golpe de magia. Portanto

$$M_C = M_0 - \rho L W h$$

Repare que a equação acima só pode ser escrita depois de se admitir que a massa total do universo tem que permanecer constante. O modelo conceitual precede, portanto, o modelo matemático preciso.

Observe que já se sabe há muitos anos que o Princípio de Lavoisier não é válido quando os objetos se movem com grande conteúdo de energia. Nesse caso, é necessário utilizar a Teoria da Relatividade para acompanhar com maior precisão as variações de

massa do universo. A Teoria da Relatividade admite, entre outras coisas, que existe uma relação direta entre massa e energia, que assim seriam manifestações distintas de uma mesma grandeza mais fundamental. O Princípio de Lavoisier, no entanto, descreve bastante bem o comportamento de sistemas de baixa energia, constituindo por isso a base de toda a Engenharia Química. Isso mostra que um modelo não precisa ser completo nem descrever todos os detalhes da realidade para que seja útil e possa ser usado na prática. Na realidade, um modelo é útil quando fornece as respostas desejadas, com precisão compatível com a precisão experimental, com o maior grau de simplicidade possível. Utilidade e complexidade não são sinônimos!

#### 4.1.3. O Modelo Físico - A Planta Piloto

Em muitos problemas é inconveniente ou impossível realizar experimentos diretamente no sistema estudado. Por exemplo, em uma fábrica não há muito espaço para experimentação, uma vez que experimentos mal sucedidos podem resultar em acidentes ou em desvios das metas de produção. Em ambientes naturais, experimentos mal sucedidos podem comprometer a saúde do ecossistema e resultar em catástrofes ambientais. Nesses casos, é comum construir modelos físicos reais do sistema estudado, as chamadas plantas pilotos. As plantas pilotos são modelos físicos reais, quase sempre de escala muito inferior à escala dos sistemas realmente estudados, utilizados para fazer experimentações e estudos que podem ser vinculados ao comportamento dos sistemas reais de interesse. Por exemplo, uma planta piloto que produz 10 kg de plástico por hora pode ser usada como modelo de uma fábrica que produz 30 toneladas de plástico por hora. Assim, testes de produção só são efetuados na planta industrial depois de terem sido aprovados na planta piloto, onde os custos e riscos são muito menores. Também nesse caso o modelo não deve ser confundido com a realidade, já que a planta industrial é provavelmente muito mais complexa que a planta piloto usada para representá-la, em função do maior volume de peças e equipamentos.

Pode-se de certa forma dizer que um modelo resulta sempre de um trabalho de investigação em qualquer área do conhecimento, já que o objetivo central da ciência é correlacionar dados e fatos. Para o engenheiro, pela própria natureza prática e exata da Engenharia, a tarefa de construir modelos para um sistema atinge o seu clímax quando resulta num conjunto consistente de relações matemáticas que permita a descrição quantitativa do sistema. Essa atividade é designada genericamente de **modelagem**. Para os fins desse livro, define-se especificamente como modelagem àquelas atividades relacionadas ao desenvolvimento de relações matemáticas precisas entre as várias variáveis de um problema.

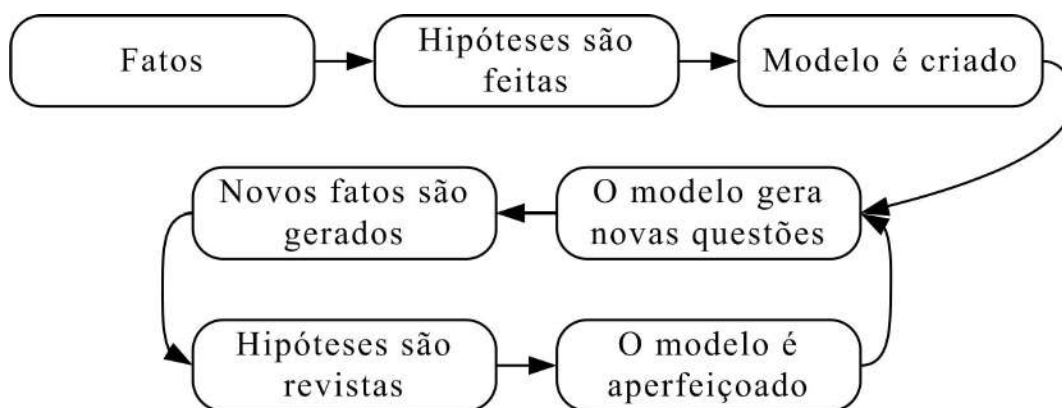


Figura 4.1 - O Conceito de Modelagem.

O conceito de modelagem pode ser representado como na Figura 4.1. Portanto, a representação é similar à utilizada para representar o ciclo clássico do método científico na Figura 1.1, pois o modelo resulta naturalmente da compreensão do problema analisado. É importante observar na Figura 4.1 que o modelo está necessariamente vinculado ao conjunto de hipóteses efetuadas pelo observador e aos dados experimentais obtidos. Logo, para que descreva adequadamente dados reais, a modelagem não dispensa os dados experimentais. **MUITO PELO CONTRÁRIO!** Apenas os dados experimentais podem permitir a validação e/ou negação do modelo criado. Não é possível, portanto, fazer modelagem sem experimentação. Modelagem sem experimentação é especulação!

As atividades de modelagem são muito importantes nas diversas áreas da ciência porque são muitas as utilidades de um bom modelo matemático. Primeiramente, o modelo matemático permite armazenar uma quantidade enorme de informação experimental. Por exemplo, a Teoria Clássica da Gravitação diz que a força de atração exercida mutuamente por dois corpos de massas  $M_1$  e  $M_2$ , separados por uma distância  $r$  é igual a

$$F = G \frac{M_1 M_2}{r^2}$$

Repare que a equação acima substitui os infinitos registros experimentais possíveis para as forças de atração entre corpos de massas distintas, separados por distâncias distintas. 10 símbolos (contando os sinais e índices) substituindo infinitos registros experimentais. Que bela concisão! Somemos a essa equação a Segunda Lei de Newton

$$M \frac{dv}{dt} = F$$

e já somos capazes de descrever o comportamento de uma infinidade de sistemas físicos reais e de interesse tecnológico. Que enxuto! Que beleza!

Em segundo lugar, como já discutido, os modelos podem ser usados para prever o comportamento de sistemas de interesse. É a atividade usualmente denominada de

**simulação.** Modelos são utilizados para simular o comportamento de sistemas físicos reais. Pode-se definir como simulação ao uso do modelo que representa o sistema real para estudar o comportamento do sistema físico real. Podem ser consideradas atividades de simulação as atividades de projeto, de otimização, de análise de estratégias, de treinamento, etc., sempre que o modelo for usado como representação adequada do sistema real. A atividade de simulação é extremamente importante do ponto de vista prático, já que resolver equações é em geral mais barato e mais rápido, além de muito mais seguro, que promover testes experimentais no sistema físico real. Assim, um teste experimental que a simulação indique que resultará em resposta inadequada não deve ser realizado. Por outro lado, a realização de um experimento que a simulação indique que resultará em uma resposta mais adequada do processo deve ser incentivada. Se a melhoria não for de fato obtida, não se deve desprezar o modelo ou renegar a atividade de modelagem. Ao invés disso, deve-se modificar e aperfeiçoar o modelo. Se a resposta obtida for satisfatória, o modelo é validado e a confiança no modelo e na qualidade da interpretação do fenômeno físico aumenta.

## 4.2. Classificação de Modelos

Muitas vezes é conveniente classificar os modelos em grupos distintos, em função dos diferentes conjuntos de ferramentas e técnicas matemáticas disponíveis para análise.

### 4.2.1. Modelos Teóricos e Empíricos

Diz-se que um modelo é teórico quando as equações que relacionam as diversas variáveis do problema são derivadas a partir de pressupostos teóricos fundamentais, como as leis de conservação de massa, energia e quantidade de movimento. Pode-se dizer que os modelos teóricos derivam de modelos conceituais que procuram interpretar o fenômeno físico estudado. Diz-se que o modelo é empírico quando as equações utilizadas para descrever as relações observadas entre as diversas variáveis do problema são postuladas, não havendo qualquer pressuposto teórico que justifique a princípio a relação utilizada. Muito freqüentemente, os modelos teóricos podem ser propostos *a priori*, antes mesmo da observação experimental (o que não significa que modelos teóricos prescindam das observações experimentais, uma vez que alguns pressupostos teóricos podem não ser verdadeiros para o problema particular analisado). Por sua vez, o desenvolvimento de modelos empíricos depende completamente da obtenção de dados experimentais confiáveis e da criatividade do analista.

**Exemplo 4.3** - Considere o problema de mistura num vaso de processo, como apresentado abaixo:



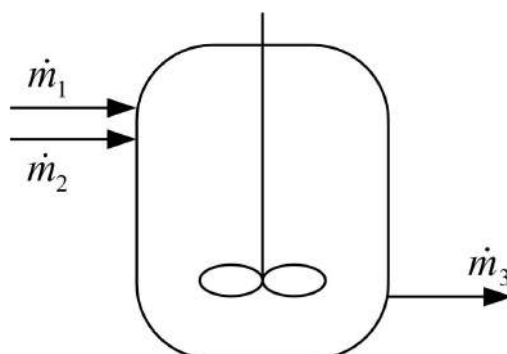


Figura 4.2 - Mistura contínua em um vaso de processo.

Admitindo-se que a massa se conserva, é possível escrever a seguinte relação matemática

$$\dot{m}_3 = \dot{m}_1 + \dot{m}_2$$

que é um modelo teórico simples, que relaciona as variáveis do problema.

Admita que um conjunto de medidas experimentais é obtido em campo, na forma

$\dot{m}_1$ (kg/h)	1.0	2.0	2.0
$\dot{m}_2$ (kg/h)	2.0	2.0	1.0
$\dot{m}_3$ (kg/h)	3.1	3.9	2.9

Observe que a relação teórica não é obedecida exatamente. No primeiro caso parece sair mais massa do que entra, enquanto no segundo e terceiro casos parece acontecer o contrário. Isso significa que o modelo teórico está errado? Não necessariamente. Lembre que os instrumentos estão sujeitos a pequenos erros de medida e que não há precisão absoluta em nenhum processo de medição. Há que se avaliar de forma mais precisa como os desvios observados se comparam aos erros de predição. Além disto, quem garante que não pode haver um pequeno vazamento nas linhas? Nesse caso, o problema não seria do pressuposto teórico, mas do inadequado controle experimental. (Aliás, alguém saberia dizer por que a massa se conserva obrigatoriamente?)

Suponha agora que um modelo empírico será construído a partir dos dados disponíveis. Nesse caso, propõe-se *a priori* que

$$\dot{m}_3 = \alpha \dot{m}_1 + \beta \dot{m}_2$$

onde  $\alpha$  e  $\beta$  são constantes a serem determinadas a partir das medidas experimentais. Para isso, utiliza-se aqui um procedimento chamado de mínimos quadrados, que consiste em reduzir ao mínimo as diferenças do quadrado dos desvios observados entre as medidas experimentais e calculadas. Nesse caso,

$$F = \sum_{i=1}^3 (\dot{m}_{3i}^e - \dot{m}_{3i}^m)^2 = \sum_{i=1}^3 (\dot{m}_{3i}^e - \alpha \dot{m}_{1i}^e - \beta \dot{m}_{2i}^e)^2$$

deve ser mínimo. O superescrito  $e$  indica o dado experimental, enquanto o superescrito  $m$  designa o dado obtido a partir do modelo. Para obter o valor mínimo,

$$\frac{\partial F}{\partial \alpha} = \sum_{i=1}^3 2(\dot{m}_{3i}^e - \alpha \dot{m}_{1i}^e - \beta \dot{m}_{2i}^e)(\dot{m}_{1i}^e) = 0$$

$$\frac{\partial F}{\partial \beta} = \sum_{i=1}^3 2(\dot{m}_{3i}^e - \alpha \dot{m}_{1i}^e - \beta \dot{m}_{2i}^e)(\dot{m}_{2i}^e) = 0$$

resultando em

$$\begin{cases} 16.7 = 9\alpha + 8\beta \\ 16.9 = 8\alpha + 9\beta \end{cases} \Rightarrow \begin{cases} \alpha = 0.8882 \\ \beta = 1.0882 \end{cases}$$

e no modelo empírico

$$\dot{m}_3 = 0.8882\dot{m}_1 + 1.0882\dot{m}_2$$

Repare que nenhum argumento teórico sustenta a relação apresentada acima, mas somente o fato de descrever de forma adequada os dados experimentais obtidos. Abaixo são mostrados os resultados experimentais e as previsões obtidas com os modelos teórico e empírico. Entre parênteses são mostrados os desvios observados entre a previsão do modelo e o dado experimental.

$\dot{m}_3$ (kg/h)	3.1	3.9	2.9	(dado experimental)
$\dot{m}_3$ (kg/h)	3.0 (-0.1)	4.0 (+0.1)	3.0 (+0.1)	(modelo teórico)
$\dot{m}_3$ (kg/h)	3.06 (-0.04)	3.95 (+0.05)	2.86 (-0.04)	(modelo empírico)

É importante enfatizar que nesse texto não se faz qualquer diferenciação ou discriminação dos modelos somente pela forma com que foram gerados. Em outras palavras, não se considera aqui que modelos teóricos são necessariamente melhores que os modelos empíricos somente porque estão baseados em pressupostos teóricos. Deixa-se essa questão para o analista e seu problema particular. De qualquer forma, é importante dizer que modelos teóricos usualmente permitem **extrapolações** muito mais confiáveis que aquelas obtidas com modelos empíricos. Isso ocorre porque em geral é muito mais razoável admitir que os pressupostos teóricos se mantenham válidos em ampla faixa de experimentação (Figura 4.3) do que admitir que a estrutura matemática proposta se mantenha constante ao longo de todas as possíveis condições experimentais (Figura 4.4). Por isso, pode-se dizer sem muito rigor que modelos empíricos revelam a estrutura local das relações existentes entre as diversas variáveis, enquanto modelos teóricos permitem desvendar uma estrutura muito mais geral a partir de umas poucas observações experimentais. Por outro lado, modelos empíricos são em geral mais simples e fáceis de derivar, permitindo construção mais rápida e barata, quando

comparada à construção de modelos teóricos mais detalhados do processo. Portanto, não parece surpreendente o fato de modelos empíricos serem preferidos para realizar **interpolações** e desenvolver aplicações em linha, como em algoritmos de controle de processos. Por essa razão, esse texto dá igual importância aos modelos teóricos e empíricos, entendendo que cada grupo particular de modelos encontra também seu nicho particular de aplicações.

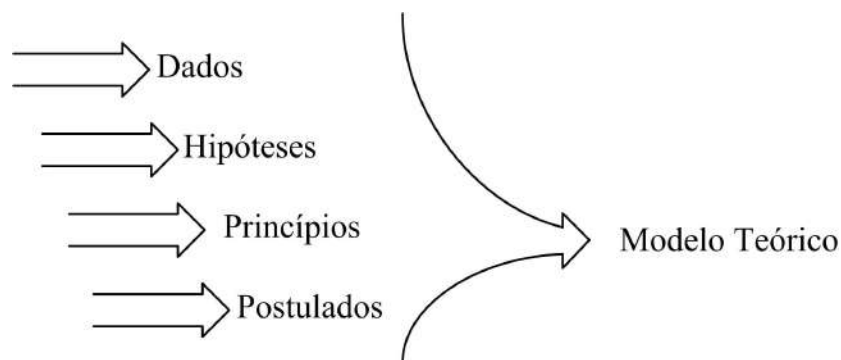


Figura 4.3 - O Desenvolvimento de um Modelo Teórico.

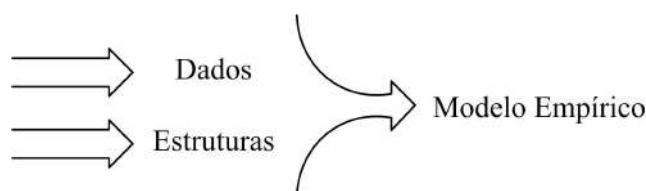


Figura 4.4 - O Desenvolvimento de um Modelo Empírico.

#### 4.2.2. Modelos Lineares e Não Lineares

Diz-se que o modelo é linear quando ele satisfaz uma das seguintes propriedades:

**Propriedade 4.1** - Sejam  $\mathbf{y}^T = [y_1 \ y_2 \ \dots \ y_{NY}]$  um conjunto de variáveis, chamadas de **variáveis de saída** ou de **variáveis dependentes**, e  $\mathbf{x}^T = [x_1 \ x_2 \ \dots \ x_{NX}]$  um segundo conjunto de variáveis, chamadas de **variáveis de entrada** ou de **variáveis independentes**. Seja ainda um **modelo matemático explícito** na forma

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{NY} \end{bmatrix} = \begin{bmatrix} f_1(x_1 \ x_2 \ \dots \ x_{NX}) \\ f_2(x_1 \ x_2 \ \dots \ x_{NX}) \\ \dots \\ f_{NY}(x_1 \ x_2 \ \dots \ x_{NX}) \end{bmatrix} = \mathbf{f}(\mathbf{x}) \quad (4.1)$$

O modelo matemático explícito da Equação (4.1) é linear se

$$\mathbf{y} = \mathbf{f}(\alpha \mathbf{x} + \beta \mathbf{w}) = \alpha \mathbf{f}(\mathbf{x}) + \beta \mathbf{f}(\mathbf{w}) \quad (4.2)$$

onde  $\alpha$  e  $\beta$  são escalares quaisquer.

O modelo descrito pela Equação (4.1) é dito explícito porque permite a obtenção direta dos valores das variáveis dependentes a partir dos valores das variáveis independentes. A definição de modelo linear é extremamente importante porque os modelos lineares via de regra permitem a obtenção de soluções analíticas para os problemas de simulação, otimização e estimação de parâmetros, como será visto adiante.

**Propriedade 4.2** - Sejam  $\mathbf{y}^T = [y_1 \ y_2 \ \dots \ y_{NY}]$  um conjunto de variáveis, chamadas de **variáveis de saída** ou de **variáveis dependentes**, e  $\mathbf{x}^T = [x_1 \ x_2 \ \dots \ x_{NX}]$  um segundo conjunto de variáveis, chamadas de **variáveis de entrada** ou de **variáveis independentes**. Seja ainda um **modelo matemático implícito** na forma

$$\begin{bmatrix} g_1(x_1 \ x_2 \ \dots \ x_{NX}; \ y_1 \ y_2 \ \dots \ y_{NY}) \\ g_2(x_1 \ x_2 \ \dots \ x_{NX}; \ y_1 \ y_2 \ \dots \ y_{NY}) \\ \dots \\ g_{NY}(x_1 \ x_2 \ \dots \ x_{NX}; \ y_1 \ y_2 \ \dots \ y_{NY}) \end{bmatrix} = \mathbf{g}(\mathbf{z}) \quad (4.3)$$

onde  $\mathbf{z}^T = [x_1 \ x_2 \ \dots \ x_{NX} \ ; \ y_1 \ y_2 \ \dots \ y_{NY}]$ . O modelo matemático implícito da Equação (4.3) é linear se

$$\mathbf{g}(\alpha \mathbf{z} + \beta \mathbf{w}) = \alpha \mathbf{g}(\mathbf{z}) + \beta \mathbf{g}(\mathbf{w}) \quad (4.4)$$

onde  $\alpha$  e  $\beta$  são escalares quaisquer.

O modelo descrito pela Equação (4.3) é dito implícito porque não permite a obtenção direta dos valores das variáveis dependentes a partir dos valores das variáveis independentes. Nesse caso, diz-se que o modelo tem que ser resolvido. Uma vez resolvido, a Equação (4.3) ganha a forma da Equação (4.1). Repare que o significado de "resolvido" aqui é muito tênue. Por exemplo, a aplicação de uma técnica numérica permite resolver o sistema sem que seja necessário encontrar uma solução analítica fechada para o problema. Por isso, na grande maioria das vezes será admitido nesse livro que o modelo matemático tem a forma da Equação (4.1), mesmo que uma solução analítica fechada não seja disponível para o problema. Nesse caso, estará sendo admitido implicitamente que uma técnica numérica pode ser usada para resolver o sistema.

**Exemplo 4.4** - Seja o modelo matemático implícito abaixo, que relaciona o conjunto de variáveis dependentes  $\mathbf{y}$  com o conjunto de variáveis independentes  $\mathbf{x}$

$$\mathbf{g}(\mathbf{z}) = \mathbf{C} \mathbf{z} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \mathbf{A} \mathbf{x} + \mathbf{B} \mathbf{y}$$

onde  $\mathbf{C}$  é uma matriz de dimensão  $NY \times (NX + NY)$ ,  $\mathbf{A}$  é uma matriz de dimensão  $NY \times NX$  e  $\mathbf{B}$  é uma matriz de dimensão  $NY \times NY$ . Nesse caso,

$$\mathbf{g}(\alpha \mathbf{w} + \beta \mathbf{u}) = \mathbf{C}(\alpha \mathbf{w} + \beta \mathbf{u}) = \alpha \mathbf{C} \mathbf{w} + \beta \mathbf{C} \mathbf{u} = \alpha \mathbf{g}(\mathbf{w}) + \beta \mathbf{g}(\mathbf{u})$$

Logo, o modelo matricial proposto é linear. Repare que a equação implícita

$$\mathbf{g}(\mathbf{z}) = \mathbf{C} \mathbf{z} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \mathbf{A} \mathbf{x} + \mathbf{B} \mathbf{y} = \mathbf{0}$$

pode ser resolvida como

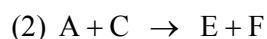
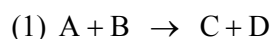
$$\mathbf{y} = -\mathbf{B}^{-1} \mathbf{A} \mathbf{x} = \mathbf{f}(\mathbf{x})$$

passando a ter a forma explícita, desde que a matriz  $\mathbf{B}$  possa ser invertida. Nesse caso,

$$\mathbf{f}(\alpha \mathbf{w} + \beta \mathbf{u}) = -\mathbf{B}^{-1} \mathbf{A} (\alpha \mathbf{w} + \beta \mathbf{u}) = -\alpha \mathbf{B}^{-1} \mathbf{A} \mathbf{w} - \beta \mathbf{B}^{-1} \mathbf{A} \mathbf{u} = \alpha \mathbf{f}(\mathbf{w}) + \beta \mathbf{f}(\mathbf{u})$$

confirmando a linearidade.

Modelos lineares matriciais como aqui descritos aparecem naturalmente durante a formulação de balanços de massa, com ou sem reação. Por exemplo, sejam as seguintes reações químicas:



Então, as seguintes equações de balanço podem ser escritas para um tanque fechado onde ocorrem as reações:

$$M_{A0} - M_A - \xi_1 - \xi_2 = 0$$

$$M_{B0} - M_B - \xi_1 = 0$$

$$M_{C0} - M_C + \xi_1 - \xi_2 = 0$$

$$M_{D0} - M_D + \xi_1 = 0$$

$$M_{E0} - M_E + \xi_2 = 0$$

$$M_{F0} - M_F + \xi_2 = 0$$

onde  $\xi_1$  e  $\xi_2$  são os graus de avanço das reações 1 e 2 respectivamente. As equações acima podem ser também escritas como:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} M_{A0} - M_A \\ M_{B0} - M_B \\ M_{C0} - M_C \\ M_{D0} - M_D \\ M_{E0} - M_E \\ M_{F0} - M_F \end{bmatrix} + \begin{bmatrix} -1 & -1 \\ -1 & 0 \\ +1 & -1 \\ +1 & 0 \\ 0 & +1 \\ 0 & +1 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = 0$$

**Exemplo 4.5** - Seja o modelo parabólico apresentado a seguir.

$$y = x^2$$

então

$$(\alpha w + \beta z)^2 = \alpha^2 w^2 + 2 \alpha \beta w z + \beta^2 z^2 \neq \alpha w^2 + \beta z^2$$

Logo, o modelo é não linear.

**Exemplo 4.6** - O conceito de linearidade não é absoluto para uma equação e depende das variáveis consideradas no problema. Por exemplo, seja o modelo na forma

$$y = \alpha_1 x_1^2 + \alpha_2 x_2^2$$

O modelo é linear nas variáveis  $\mathbf{\alpha}^T = [\alpha_1 \quad \alpha_2]$  e não linear nas variáveis  $\mathbf{x}^T = [x_1 \quad x_2]$ . Portanto, é necessário definir as variáveis consideradas para que o conceito de linearidade faça sentido.

Da mesma forma, seja o modelo matemático implícito abaixo, que relaciona a variável dependente  $y$  com a variável independente  $x$

$$g(y, x) = \frac{dy}{dx} + 4y, \quad y(0) = y_0$$

Fazendo-se

$$\mathbf{z} = \begin{bmatrix} y \\ x \end{bmatrix} = \alpha \mathbf{w} + \beta \mathbf{u} = \alpha \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \beta \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} \alpha w_1 + \beta u_1 \\ \alpha w_2 + \beta u_2 \end{bmatrix}$$

então

$$y = \alpha w_1 + \beta u_1 \Rightarrow \frac{dy}{dx} = \alpha \frac{dw_1}{dx} + \beta \frac{du_1}{dx}$$

$$x = \alpha w_2 + \beta u_2 \Rightarrow \frac{d}{dw_2} = \alpha \frac{d}{dx} \Rightarrow \frac{d}{du_2} = \beta \frac{d}{dx}$$

$$y = \alpha w_1 + \beta u_1 \Rightarrow y_0 = \alpha w_{10} + \beta u_{10}$$

Combinando as duas equações acima nos termos da equação original

$$\frac{dy}{dx} = \alpha \frac{dw_1}{dx} + \beta \frac{du_1}{dx} = \frac{dw_1}{dw_2} + \frac{du_1}{du_2}$$

$$4y = 4 \alpha w_1 + 4 \beta u_1$$

e

$$\left( \frac{dw_1}{dw_2} + 4 \alpha w_1 \right) + \left( \frac{du_1}{du_2} + 4 \beta u_1 \right) \neq \alpha g(w_1, w_2) + \beta g(u_1, u_2)$$

de maneira que o modelo é não linear. No entanto, admitindo-se que  $x$  não é uma variável relevante do problema e que não pode ser manipulada, então

$$g(y) = \frac{dy}{dx} + 4y, \quad y(0) = y_0$$

Fazendo-se

$$y = \alpha w_1 + \beta u_1 \Rightarrow \frac{dy}{dx} = \alpha \frac{dw_1}{dx} + \beta \frac{du_1}{dx}$$

que combinada com a equação original

$$\alpha \left( \frac{dw_1}{dx} + 4w_1 \right) + \beta \left( \frac{du_1}{dx} + 4u_1 \right) = \alpha g(w_1, x) + \beta g(u_1, x)$$

resultando em um modelo linear. Portanto, sempre que a hipótese de linearidade for levantada, é necessário definir o conjunto de variáveis que estão sendo consideradas no problema.

### 4.2.3. Modelos Determinísticos e Estocásticos

Como já discutido na Seção 1.3, modelos determinísticos são aqueles que associam a cada experimento um resultado experimental bem definido, enquanto modelos estocásticos associam a cada condição experimental um conjunto de possíveis resultados, cada qual com uma certa probabilidade de ocorrer. De maneira pragmática, um modelo determinístico associa a cada pergunta sempre uma mesma resposta, enquanto modelos estocásticos associam a cada pergunta um conjunto de respostas possíveis, com diferentes probabilidades. Um modelo estocástico admite, portanto, que um mesmo experimento pode resultar em diferentes respostas, algumas das quais são muito prováveis, enquanto outras são pouco prováveis.

**Exemplo 4.7** - Seja o modelo dado na forma

$$\frac{dy}{dt} = -y, \quad y(0) = y_0 \Rightarrow y(t) = y_0 e^{-t}$$

Dada uma certa condição inicial, a trajetória dinâmica obtida é sempre a mesma. O modelo é, portanto, determinístico.

**Exemplo 4.8** - Para simular a difusão do composto A num segundo composto B, monta-se uma rede (ou *grid*) e trocam-se as posições de A com um de seus vizinhos B, de forma aleatória, até que A atinja o outro lado da rede. O tempo (ou número de iterações) que A demora para atingir o outro lado da rede caracteriza a velocidade de difusão de A no meio.

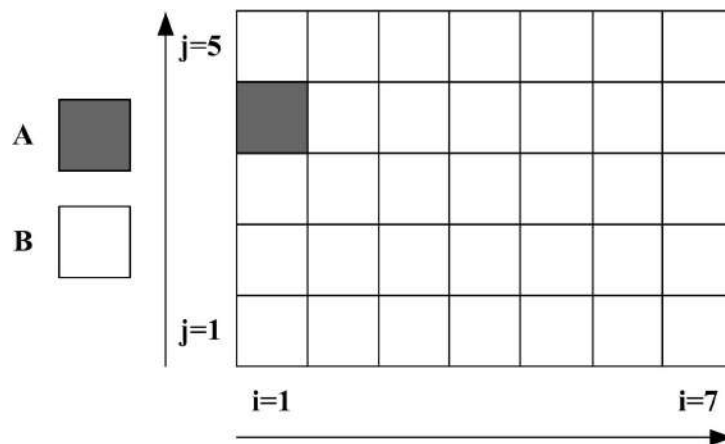


Figura 4.5 - Rede para Simulação da Difusão.

A simulação é conduzida com a geração de números pseudo-aleatórios com a equação abaixo

$$X_{k+1} = 11X_k - \text{Trunc}(11X_k)$$

usando como semente inicial o número  $X_0=0.35312356$ . O processo difusivo é simulado com as seguintes regras:

- Se  $0.00 < X_{k+1} < 0.25$ ;  $i_{k+1} = i_k - 1$ ;  $j_{k+1} = j_k$ ; ou seja, a molécula anda para trás;
- Se  $0.25 < X_{k+1} < 0.50$ ;  $i_{k+1} = i_k$ ;  $j_{k+1} = j_k - 1$ ; ou seja, a molécula anda para baixo;
- Se  $0.50 < X_{k+1} < 0.75$ ;  $i_{k+1} = i_k + 1$ ;  $j_{k+1} = j_k$ ; ou seja, a molécula anda para frente;
- Se  $0.75 < X_{k+1} < 1.00$ ;  $i_{k+1} = i_k$ ;  $j_{k+1} = j_k + 1$ ; ou seja, a molécula anda para cima.

A simulação é interrompida quando  $i_{k+1} = 7$ ; ou seja, quando a molécula atinge a outra extremidade da Figura 4.5. A simulação é realizada 1000 vezes, usando como semente para o algoritmo de geração de números pseudo-aleatórios o último número gerado na etapa anterior. A Figura 4.6 mostra o caminho trilhado pela molécula na rede de difusão durante a primeira simulação. A Figura 4.7 mostra o número de iterações obtidas ao longo das 1000 simulações e a qualidade do ajuste exponencial. Pode-se dizer que o tempo característico de difusão segue a distribuição exponencial. (É curioso observar que os balanços determinísticos resultam na mesma relação exponencial com o tempo, mostrando que nem sempre é possível distinguir com exatidão um modelo determinístico de um modelo estocástico.) O número médio de iterações necessárias para atingir o lado oposto da rede é igual a 78, com variância igual a 4442.



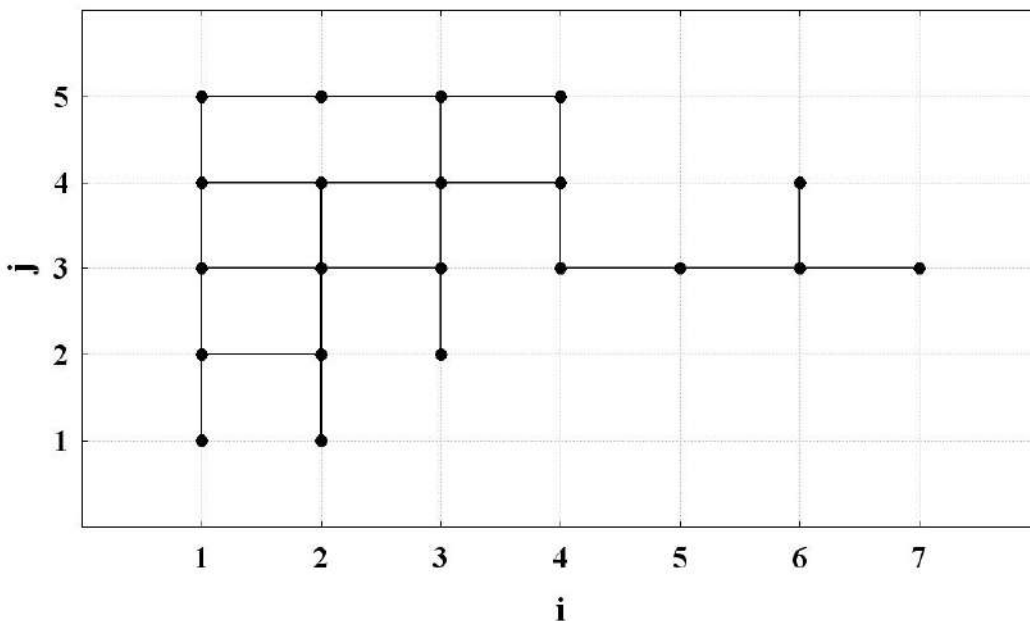


Figura 4.6- Caminho difusivo percorrido pela molécula A durante a primeira simulação.

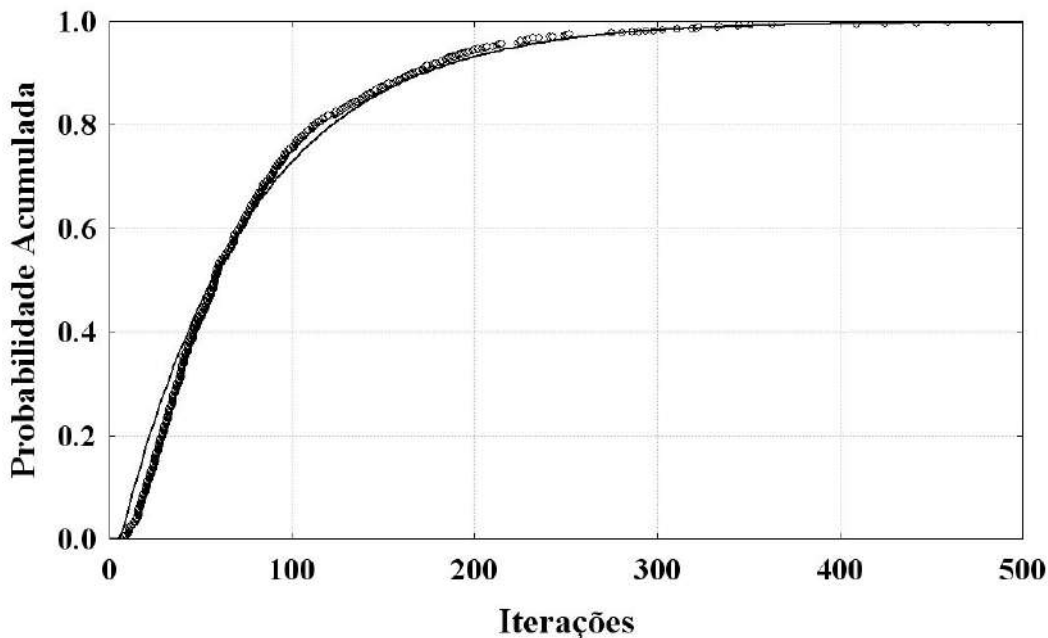
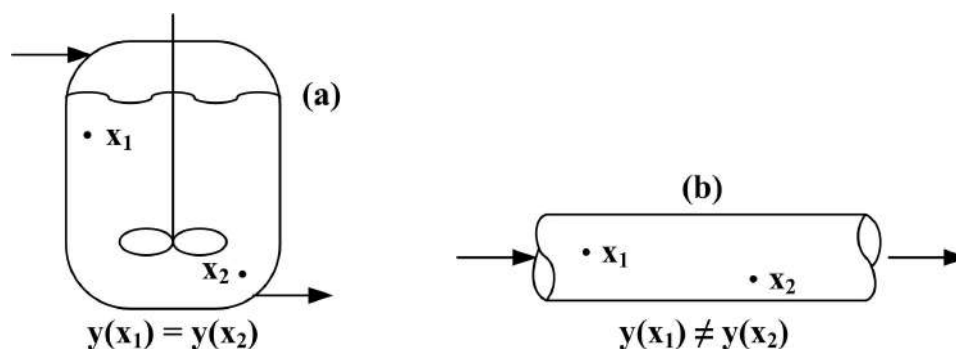


Figura 4.7- Distribuição do tempo necessário para percorrer a rede de difusão e comparação com o ajuste exponencial ( $f(i)=\text{Expon}(i-6; 59)$ ).

Modelos e simulações deste tipo são usualmente chamados de Modelos e Simulações de Monte Carlo.

#### 4.2.4. Modelos a Parâmetros Concentrados e a Parâmetros Distribuídos

Na engenharia é muito adequado também classificar os modelos quanto à forma com que são descritas as variações espaciais das variáveis de interesse. Diz-se que o modelo é a parâmetros concentrados quando as variações espaciais são desprezíveis e as propriedades não mudam com a posição. O exemplo clássico é o modelo do tanque de mistura. Nesse caso, admitindo-se a validade de hipótese de mistura perfeita, as propriedades são as mesmas em qualquer ponto do espaço. Por outro lado, diz-se que o modelo é a parâmetros distribuídos quando as variações espaciais são importantes e não podem ser desprezadas. Nesse caso há, portanto, heterogeneidade espacial. O exemplo clássico é o modelo do reator tubular. Veja a Figura 4.8 apresentada a seguir.



**Figura 4.8** - Exemplo de sistemas a parâmetros concentrados (a) e a parâmetros distribuídos (b).

Essa classificação é útil porque os modelos matemáticos que descrevem sistemas a parâmetros distribuídos ganham em geral a forma de equações diferenciais parciais, cuja resolução requer o uso de procedimentos numéricos bastante específicos. Os modelos a parâmetros concentrados, por sua vez, quase sempre são constituídos por equações algébricas ou equações diferenciais ordinárias de primeira ordem, cuja solução numérica é muito mais simples.

### 2.2.5. Modelos Estacionários e Dinâmicos

Finalmente, é bastante útil classificar os modelos quanto à dependência temporal das propriedades e/ou variáveis que ele descreve. O modelo é dito dinâmico quando uma ou mais variáveis do modelo mudam no tempo. Aplicações em controle de processos, por exemplo, requerem estruturas dinâmicas para análise, uma vez que se procura detectar e corrigir problemas que possam ocorrer com o processo ao longo do tempo. O modelo é chamado de estacionário quando as variáveis não mudam no tempo. O desenvolvimento de projetos de máquinas e equipamentos em geral parte do pressuposto do comportamento estacionário, para que seja possível determinar as dimensões ótimas do equipamento e a condição ótima de operação.

Essa classificação é útil porque os modelos matemáticos que descrevem sistemas dinâmicos quase sempre requerem a implementação de rotinas numéricas de integração, como os algoritmos clássicos de Euler e Runge-Kutta. Os modelos estacionários, por sua vez, quer sejam a parâmetros concentrados ou a parâmetros distribuídos (depois de implementados os procedimentos de discretização), quase sempre resultam em sistemas de equações algébricas, a serem resolvidos com técnicas numéricas clássicas, como de Newton-Raphson, desenvolvidas para resolução de sistemas de equações algébricas.

**Exemplo 4.9** - Seja o modelo abaixo

$$\begin{aligned} \frac{\partial C}{\partial t} &= D \frac{\partial^2 C}{\partial x^2} - v \frac{\partial C}{\partial x} - K C \\ C(0, x) &= 0 \\ C(t, 0) &= C_0 \\ \left. \frac{\partial C}{\partial x} \right|_{x=L} &= 0 \end{aligned}$$

que descreve as variações de concentração ( $C$ ) de um certo reagente A ao longo da posição axial ( $x$ ) de um reator tubular de comprimento  $L$ , ao longo do tempo ( $t$ ).  $D$  é o chamado coeficiente de dispersão do reagente A no tubo;  $v$  é a velocidade média do escoamento ao longo do tubo; e  $K$  é a constante de velocidade da reação. A primeira condição de contorno diz que o reator está inicialmente vazio do composto A (só contém solvente, por exemplo); a segunda condição de contorno diz que a concentração da corrente de alimentação é constante e igual a  $C_0$ ; e a terceira condição de contorno diz que nada muda a partir da saída do reator.

O modelo acima é um modelo dinâmico a parâmetros distribuídos. A versão estacionária desse modelo, usada frequentemente para o projeto de reatores químicos e obtida quando as variações temporais desaparecem, tem a forma

$$\begin{aligned} 0 &= D \frac{d^2 C}{dx^2} - v \frac{dC}{dx} - K C \\ C(0) &= C_0 \\ \left. \frac{dC}{dx} \right|_{x=L} &= 0 \end{aligned}$$

Repare que a complexidade do modelo é bastante menor.

Para resolver o modelo, é bastante comum adotar esquemas de discretização. Embora isso não seja necessário nos dois modelos acima, pois soluções analíticas fechadas podem ser desenvolvidas para ambos os casos, admita que o reator tubular é dividido em  $N$  "fatias" de comprimento  $\Delta L = L/N$  ao longo do comprimento  $L$ , como representado na Figura 4.9. Nesse caso, em um ponto  $i$  qualquer no interior do reator, é possível escrever

$$\begin{aligned} \left. \frac{dC}{dx} \right|_{x=x_i} &\approx \frac{C_{i+1} - C_{i-1}}{2 \Delta L} \\ \left. \frac{d^2 C}{dx^2} \right|_{x=x_i} &\approx \frac{\left. \frac{dC}{dx} \right|_{x=x_i + \Delta L/2} - \left. \frac{dC}{dx} \right|_{x=x_i - \Delta L/2}}{\Delta L} \approx \frac{\frac{C_{i+1} - C_i}{\Delta L} - \frac{C_i - C_{i-1}}{\Delta L}}{\Delta L} = \frac{C_{i+1} - 2C_i + C_{i-1}}{\Delta L^2} \end{aligned}$$

de maneira que o modelo estacionário fica na forma de um conjunto de  $N$  equações algébricas

$$0 = \frac{D}{\Delta L^2} (C_{i+1} - 2C_i + C_{i-1}) - \frac{v}{2 \Delta L} (C_{i+1} - C_{i-1}) - K C_i, \quad i = 1 \dots N$$

com

$C_0$  conhecido (primeira condição de contorno)

$C_{N+1} = C_N$  (segunda condição de contorno)

que pode ser resolvido com a precisão desejada e imposta pela discretização (número de "fatias"  $N$ ).

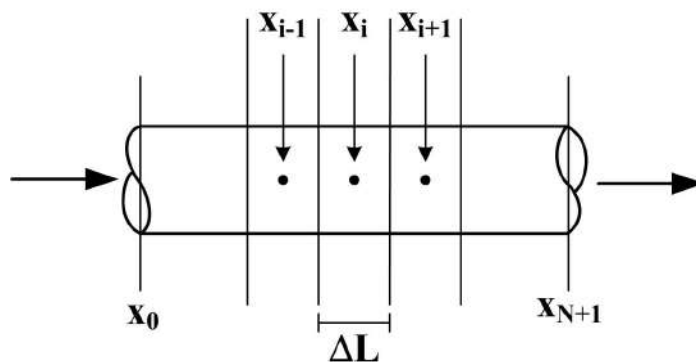


Figura 4.9 - Esquema de discretização de diferenças finitas.

Para o caso muito especial em que  $N$  é igual a 1 (a discretização mais simples possível)

$$0 = \frac{D}{\Delta L^2} (C_2 - 2C_1 + C_0) - \frac{v}{2 \Delta L} (C_2 - C_0) - K C_1$$

com

$C_0$  conhecido (primeira condição de contorno)

$C_2 = C_1$  (segunda condição de contorno)

e

$$0 = \frac{D}{\Delta L^2} (C_1 - 2 C_1 + C_0) - \frac{v}{2 \Delta L} (C_1 - C_0) - K C_1$$

$$C_1 = C_2 = \frac{\left( \frac{D}{\Delta L^2} + \frac{v}{2 \Delta L} \right)}{\left( \frac{D}{\Delta L^2} + \frac{v}{2 \Delta L} + K \right)} C_0$$

De maneira similar, o modelo dinâmico fica na forma

$$\frac{dC_i}{dt} = \frac{D}{\Delta L^2} (C_{i+1} - 2C_i + C_{i-1}) - \frac{v}{2\Delta L} (C_{i+1} - C_{i-1}) - K C_i, \quad i = 1 \dots N$$

com

$$\begin{aligned} C_i(0) &= 0 \text{ (primeira condição de contorno)} \\ C_0(t) &\text{ conhecido (segunda condição de contorno)} \\ C_{N+1}(t) &= C_N(t) \text{ (terceira condição de contorno)} \end{aligned}$$

Assim, para  $N$  igual a 1

$$\frac{dC_1}{dt} + \left( \frac{D}{\Delta L^2} + \frac{v}{2\Delta L} + K \right) C_1 = \left( \frac{D}{\Delta L^2} + \frac{v}{2\Delta L} \right) C_0$$

com

$$C_1(0) = 0$$

Repare como o esquema de discretização reduz a complexidade matemática do modelo, ao mesmo tempo em que aumenta o número de equações a serem resolvidas. Repare ainda que o procedimento de discretização reduziu o modelo estacionário diferencial original a um conjunto de equações algébricas, reduzindo o modelo dinâmico diferencial original a um conjunto de equações diferenciais ordinárias.

É importante salientar que, uma vez desenvolvido um modelo matemático, é fundamental que possamos RESOLVÊ-LO. Um modelo matemático que não pode ser resolvido não tem qualquer utilidade. Além disso, um modelo matemático mal resolvido é ineficiente. Por isso, de maneira pouco precisa, pode-se dizer que fazer simulações é resolver o modelo muitas vezes, para diferentes condições. Pode-se inclusive dizer que um modelo é o pacote constituído pelas equações que representam o sistema e as técnicas numéricas que permitem resolver as equações. Portanto, as técnicas numéricas utilizadas constituem uma parte importante do modelo utilizado para descrever o processo. Normalmente não atentamos para esse fato porque nos acostumamos a pensar em modelos explícitos, em que dado  $x$  é possível obter  $y$  diretamente, como na Equação (4.1). Isso nem sempre é verdade, como mostram os Exemplos 4.8 e 4.9. Às vezes é necessário um pouco mais de trabalho e criatividade. Daí a enorme importância da Matemática e da Computação, em particular das ferramentas numéricas, na área de Modelagem e Simulação de Processos, e mais especificamente para a disciplina de Estimação de Parâmetros. É importante salientar, no entanto, que para os fins desse livro admite-se que o modelo sempre pode ser resolvido de forma eficiente pelo analista. É problema do analista, portanto, resolver o modelo por ele desenvolvido.

### 4.3. Definição do Problema de Estimação de Parâmetros

Como já discutido exhaustivamente ao longo desse capítulo, um problema fundamental em qualquer trabalho científico é o de correlacionar dados (construir modelos). No entanto, a construção do modelo envolve a definição de ao menos duas entidades básicas distintas:

a) A estrutura do modelo

$$y = \alpha x^2 \quad (\text{relação quadrática entre } x \text{ e } y)$$

$$y = \alpha e^{\beta x} \quad (\text{relação exponencial entre } x \text{ e } y)$$

b) Os parâmetros do modelo

$\alpha$  e  $\beta$  nas relações acima

A estrutura do modelo é a forma funcional através da qual as diversas variáveis do problema estão relacionadas. Os parâmetros do modelo são os números que tornam possível a previsão quantitativa das relações existentes entre as várias variáveis do problema, através da estrutura do modelo. Por exemplo, quando se diz que duas variáveis  $x$  e  $y$  estão relacionadas linearmente, apenas se estabelece o tipo de relação funcional que existe entre as duas variáveis analisadas como

$$y = \alpha x + \beta$$

No entanto, para que o modelo seja útil e possa ser utilizado para fazer previsões ou simulações, é necessário definir adicionalmente quem são os coeficientes angular ( $\alpha$ ) e o coeficiente linear ( $\beta$ ) da reta. Caso contrário, de pouco serve o modelo. Repare que a estrutura do modelo pode ser gerada de diversas maneiras, de forma empírica ou fundamentada em preceitos teóricos. De qualquer forma, sem os parâmetros a estrutura pura do modelo quase nunca faz sentido.

**Exemplo 4.10** - A Lei de Fourier da transferência de calor, gerada a partir da observação experimental, diz que a taxa de transferência de calor que se estabelece entre os dois planos opostos de uma parede é proporcional à diferença de temperaturas existente entre os dois planos, é proporcional à área de contato entre os dois planos, e é inversamente proporcional à distância entre os dois planos (espessura da parede). Levada ao limite infinitesimal de espessura da parede, a Lei de Fourier ganha a forma

$$\dot{q} = -k A \frac{dT}{dx}$$

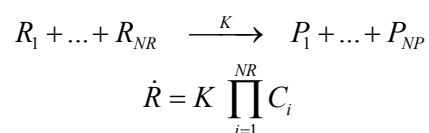
onde  $\dot{q}$  é a taxa de transferência de calor (energia / tempo),  $A$  é a área de contato entre os planos,  $T$  é a temperatura e  $x$  é o comprimento medido ao longo da espessura da parede.  $k$  é a chamada condutividade térmica do material ((energia comprimento)/(tempo temperatura)). O sinal de menos indica que o calor flui sempre do lado mais quente para o lado mais frio; ou seja, flui na direção contrária do gradiente de temperaturas. Se a condutividade térmica do material é constante (ou se a parede é suficientemente fina), então

$$\dot{q} = -k A \frac{(T_2 - T_1)}{L}$$

A equação acima é uma equação fundamental para o projeto de isolamentos. Dadas as características do material isolante ( $k$ ), do sistema avaliado ( $A$ ,  $T_2$ ,  $T_1$ ) e a máxima perda de calor admissível ( $\dot{q}$ ), obtém-se a quantidade necessária de isolante ( $L$ ). No entanto, para que a equação seja de fato útil, é necessário conhecer o parâmetro

$k$ . A medição e estudo da condutividade térmica de materiais é um problema fundamental da área de sistemas térmicos.

De outra forma, a aplicação de princípios teóricos rigorosos (e elegantes) permitem afirmar que as taxas de reação química observadas em sistemas gasosos diluídos, formado por moléculas esféricas rígidas, são proporcionais às concentrações dos reagentes. Essa dependência funcional é conhecida como Lei de Ação das Massas e pode ser escrita na forma



onde  $R_i$  designa o reagente  $i$ ,  $P_i$  designa o produto  $i$ ,  $\dot{R}$  designa a velocidade da reação (moles / (tempo volume) ),  $C_i$  é a concentração da espécie  $i$  (moles / volume) e  $K$  é a constante de velocidade da reação, que depende da temperatura de acordo com a equação

$$K = K_0 \exp \left[ -\frac{\Delta E}{RT} \right]$$

que é a conhecida Lei de Arrhenius.  $K_0$  é o fator de frequência da reação,  $\Delta E$  é a energia de ativação da reação (energia / mol ),  $R$  é a constante universal dos gases (1.9876 cal / (mol K) ) e  $T$  é a temperatura absoluta. A Lei de Ação das Massas e a Lei de Arrhenius são fundamentais para o projeto de reatores químicos e são objetos de estudo da disciplina de Cinética das Reações Químicas. No entanto, são de pouco valor se os valores do fator de frequência ( $K_0$ ) e da energia de ativação ( $\Delta E$ ), característicos da reação química investigada, não são conhecidos.

Um problema fundamental para todos aqueles envolvidos com atividades científicas, em particular aqueles envolvidos com a compreensão quantitativa de como uma variável do problema influencia as demais, é determinar os parâmetros do modelo. Por exemplo, como determinar  $\dot{q}$ ,  $k$ ,  $A$ ,  $L$ ,  $T$ ,  $\dot{R}$ ,  $C$ ,  $K_0$  e  $\Delta E$  nos problemas do Exemplo 4.10 A resposta pode ser obtida usualmente das seguintes formas:

- Fixando alguns valores típicos de projeto para algumas variáveis independentes. Por exemplo, no caso da troca de calor, os valores do  $\dot{q}$  máximo admissível e de ao menos uma das temperaturas são normalmente estabelecidos *a priori* pelas características do projeto;
- Calculando as variáveis dependentes através do modelo. Por exemplo, no problema de troca de calor, a espessura de isolante  $L$  é obtida como função das demais variáveis do problema;
- Medindo variáveis e parâmetros com instrumentação adequada. Por exemplo, no problema de troca de calor, a área de contato pode ser obtida medindo-se as dimensões características do meio que está sendo isolado;

- d) Consultando a literatura especializada. Por exemplo, no caso da troca de calor, um manual pode ser consultado para se observar se a condutividade térmica do isolante considerado já foi avaliada e reportada por outros pesquisadores.

Embora o parágrafo anterior induza o leitor a pensar que o problema de avaliação dos parâmetros é simples, isso não é verdade absolutamente. Vejamos alguns pontos curiosos.

- a) A literatura especializada não fornece todos os dados necessários para a realização de qualquer projeto. Muito pelo contrário! Quanto mais importante e relevante o problema do ponto de vista tecnológico e econômico, menos provável é que se encontrem informações relevantes do problema na literatura pública. Todo engenheiro já passou pela experiência de procurar dados na literatura sem sucesso. Além disso, essa estratégia apenas transfere de mãos o problema fundamental da avaliação dos parâmetros, não o solucionando. Aliás, grande parte das correlações propostas na literatura resulta de extenso trabalho de modelagem e estimação de parâmetros realizados por terceiros;
- b) A medição de certos parâmetros é virtualmente impossível, seja porque não existe técnica experimental disponível para esse fim (por exemplo, para medir os valores de  $K_0$  e  $\Delta E$  que caracterizam uma reação química), seja porque o parâmetro não tem qualquer significado físico real (como os coeficientes  $\alpha$  e  $\beta$  da correlação empírica linear ou quaisquer outros coeficientes de natureza empírica), seja porque o parâmetro é na realidade definido pela relação estabelecida entre as variáveis fundamentais do problema (como a condutividade térmica na Lei de Fourier), ou seja, porque nem mesmo a relação existente entre as variáveis é de fato conhecida.

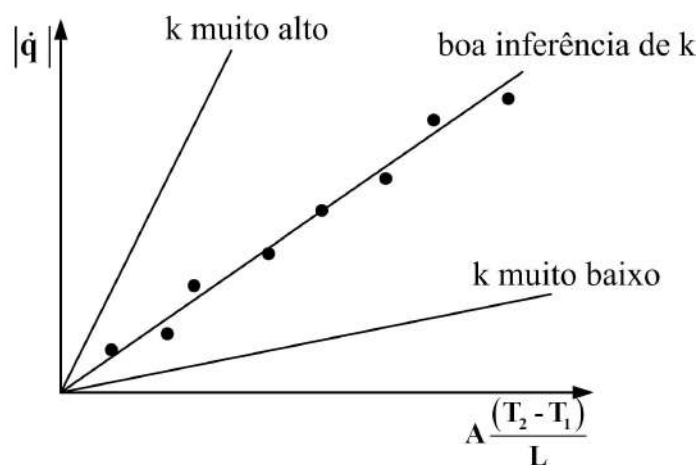
Por todas as razões apresentadas acima, independentemente da estrutura ou origem do modelo matemático utilizado, quase sempre há certos números ou parâmetros que não podem ser medidos nem avaliados *a priori* pelo analista, mas sem os quais não é possível nem usar o modelo nem estabelecer vínculos entre as variáveis. Como proceder então nesses casos? A resposta para esse problema constitui o conjunto de ferramentas conhecidas como técnicas de **estimação de parâmetros**. Estimar parâmetros consiste fundamentalmente em **inferir** os valores dos parâmetros que não podem ser medidos nem avaliados *a priori*, a partir de uma comparação estabelecida entre dados experimentais e um modelo disponível para o processo, cujo desempenho é afetado pelo parâmetro de interesse. Estimar parâmetros consiste, portanto, em obter dados experimentais e comparar esses dados com estruturas que pretendem explicá-los. Estimar parâmetros consiste em exercitar, portanto, as componentes experimental e teórica de uma investigação científica.

**Exemplo 4.11** - Para o problema da transferência de calor do Exemplo 4.10, não é possível de fato medir o valor de  $k$ . Não existe um instrumento (“condutivômetro térmico”? ☺) que possa ser conectado ao material para fornecer diretamente o valor de  $k$ . Mas a Lei de Fourier diz que

$$\dot{q} = -k A \frac{(T_2 - T_1)}{L}$$



em que todas as demais grandezas físicas, com exceção de  $k$ , podem ser medidas durante um experimento de troca de calor. Suponha que um pedaço do isolante considerado é prensado entre duas paredes de dimensões bem definidas (a espessura  $L$  e a área  $A$  medidos), mantidas a temperaturas constantes ( $T_1$  e  $T_2$  medidos) através da manipulação da quantidade de calor dissipada por uma resistência elétrica ( $\dot{q}$  medido). Nesse caso, suponha que várias medidas (experimentos) são feitas. Segundo a Lei de Fourier, a condutividade térmica  $k$  é o fator de proporcionalidade (coeficiente angular) existente entre a medida  $\dot{q}$  e o grupo de medidas  $A \frac{(T_2 - T_1)}{L}$ . Se essas medidas são lançadas em um gráfico, na forma abaixo, é possível inferir de alguma maneira o valor de  $k$ . Portanto, não é exagero dizer que o problema de estimação de parâmetros é equivalente à construção de um **sensor virtual** (*softsensor*) para medição das variáveis que não podem ser medidas diretamente com instrumentos físicos. Portanto, o “condutivômetro térmico” é o procedimento de estimação de parâmetros.



**Figura 4.10** - Inferência da condutividade térmica do isolante a partir de outras medidas experimentais.

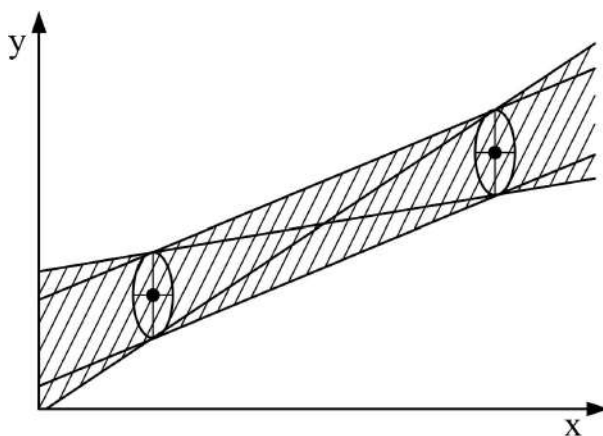
#### 4.4. Características Fundamentais do Problema de Estimação de Parâmetros

Embora o procedimento de estimação de parâmetros esteja filosoficamente ilustrado na Figura 4.12, é preciso definir o problema de forma mais rigorosa e precisa, para que sejam eliminadas quaisquer ambigüidades de condução do processo e a possibilidade da intervenção deletéria e desavisada do analista. Por isso, é conveniente tentar colocar primeiramente em palavras em que consiste o procedimento de estimação de parâmetros. Uma possível definição do problema pode ser apresentada na forma: *"Estimar parâmetros é usar um modelo como referência e variar os parâmetros até que as predições do modelo passem o mais próximo possível dos dados experimentais, respeitadas as incertezas características da medição."*

A definição introduzida acima é bastante interessante porque é composta por um conjunto de palavras-chaves que antecipam a formulação do problema matemático que

caracteriza o procedimento de estimação de parâmetros. Vejamos como alguns desses elementos fundamentais caracterizam o problema de estimação ou inferência:

- a) Há um **modelo de referência**. Essa é uma característica fundamental do processo de estimação de parâmetros. O modelo de referência serve de molde, em torno do qual os dados experimentais devem ser encaixados. Portanto, o procedimento de estimação de parâmetros pressupõe uma tentativa de compreender a realidade experimental e não pode ser conduzido se o analista não quer ou não tem coragem de propor uma explicação, seja ela empírica ou teórica, para as suas observações experimentais. Logo, para o problema de estimação de parâmetros o modelo é um dado conhecido, cabendo ao analista prover esse dado.
- b) Os parâmetros são variados. Dessa forma, **durante o procedimento de estimação de parâmetros os parâmetros do modelo são as verdadeiras variáveis consideradas**. Genericamente, durante uma simulação com o modelo  $y = f(x; \alpha)$ , admite-se que os parâmetros  $\alpha$  são conhecidos e são feitos estudos sobre como as variáveis independentes  $x$  influenciam as variáveis dependentes  $y$ . O problema de estimação de parâmetros consiste, ao contrário, em observar quão longe ou perto dos dados experimentais  $x$  e  $y$  conhecidos o modelo passa, quando os parâmetros  $\alpha$  são modificados. (Por isso, não surpreende que em muitas áreas da engenharia o problema de estimação de parâmetros seja chamado de **problema inverso**.) Durante a estimação dos parâmetros podem ser feitas, portanto, muitas simulações das condições experimentais, para diferentes conjuntos de valores de parâmetros.
- c) O modelo deve passar o mais próximo possível dos dados experimentais. Logo, o procedimento de estimação de parâmetros pressupõe a existência de uma **métrica**; ou seja, de uma função que mede a distância existente entre os dados experimentais e os dados previstos pelo modelo. Todo o procedimento de estimação de parâmetros depende da definição dessa métrica, que em última análise diz se as previsões feitas com o modelo são boas ou ruins. Mais ainda, o procedimento de estimação de parâmetros pressupõe a implementação de algum tipo de rotina de **otimização**, dado que as previsões feitas com o modelo não devem estar apenas próximas dos dados experimentais, mas sim o mais próximo possível. Portanto, a similaridade entre os dois conjuntos, experimentos e previsões, deve ser máxima.
- d) Devem ser respeitadas as incertezas características da medição. Logo, não deve ser esquecido que as medidas experimentais contêm erros e que **os erros influenciam o processo de inferência dos parâmetros**. Essa questão é ilustrada na Figura 4.11 apresentada abaixo. Suponha que dois pares de dados experimentais estão disponíveis para que se estimem os parâmetros característicos da reta. Como os dados experimentais contêm erros, os valores *verdadeiros* dos dados não são conhecidos, mas é possível avaliar uma região de confiança, com grau de precisão escolhido pelo analista, onde se espera encontrar os valores *verdadeiros*. Observe, no entanto, que a incerteza no dado experimental provoca como consequência uma incerteza na reta que une os dois pares de dados experimentais. Portanto, é fundamental reconhecer que o procedimento de estimação de parâmetros, por estar baseado na análise de dados experimentais que contêm um certo grau de incerteza, resulta em valores que também contêm um certo grau de incerteza. Portanto, o procedimento de estimação de parâmetros deve ser interpretado à luz dos conhecimentos básicos da Estatística.



**Figura 4.11** - Os pontos ilustram as medidas experimentais. As regiões ovais indicam as incertezas experimentais. A região hachurada indica onde estão as possíveis retas que descrevem os dados experimentais.

Baseado na discussão anterior e por conveniência de apresentação, o problema de estimação de parâmetros é segmentado aqui em três subproblemas. O primeiro subproblema consiste em definir uma métrica (também chamada de **função objetivo**) adequada para o problema. A definição de uma métrica apropriada é de fundamental importância, pois é baseado nessa métrica que se desenvolve o processo de inferência paramétrica. O segundo subproblema consiste em achar o ponto de ótimo da métrica formulada (mínimo ou máximo, a depender da lógica subjacente à métrica utilizada). Quase sempre, a definição do ponto de ótimo só é possível com o auxílio de técnicas numéricas. Algumas dessas técnicas numéricas são apresentadas e discutidas ao longo do Capítulo 5. No entanto, para alguns poucos problemas específicos descritos por modelos lineares nos parâmetros, é possível encontrar o ponto de ótimo analiticamente. O terceiro subproblema consiste finalmente em formular uma interpretação estatística precisa dos parâmetros obtidos e da qualidade da previsão efetuada com o auxílio do modelo. Cada um desses três subproblemas fundamentais é discutido nas seções que seguem.

**Exemplo 4.12** - Em conformidade com o Exemplo 4.11, suponha que o seguinte conjunto de dados experimentais está disponível

$\dot{q}$ (cal / h)	1050	2000	2950	4000
$A \frac{(T_2 - T_1)}{L}$ (m K)	10	20	30	40

a partir dos quais pretende-se estimar o valor da condutividade térmica  $k$ . Para simplificar a apresentação e ajudar a fixação da nomenclatura proposta, o problema é descrito na forma

$$y = \alpha x$$

onde  $x$  é a variável independente  $\left(x = A \frac{(T_2 - T_1)}{L}\right)$ ,  $y$  é a variável dependente ( $\dot{q}$ ) e  $\alpha$  é o parâmetro estimado ( $k$ ). Suponha ainda que a seguinte métrica é proposta para descrever a distância entre os dados experimentais e os dados previstos pelo modelo:

$$F = \sum_{i=1}^{NE} (y_i^e - y_i^m)^2 = \sum_{i=1}^{NE} (y_i^e - \alpha x_i^e)^2$$

onde  $NE$  representa o número total de experimentos analisados (4, nesse caso) e os superescritos  $e$  e  $c$  designam respectivamente os dados medidos experimentalmente e os dados previstos com o modelo. Repare que, de fato, à medida que a diferença entre os dados experimentais e os dados do modelo aumenta,  $F$  também aumenta. Desta forma, quanto maior o valor de  $F$ , mais distante as previsões feitas com o modelo estão dos dados experimentais disponíveis. Para minimizar a distância do modelo aos dados experimentais, deve-se manipular os parâmetros de forma adequada. Nesse caso, o ponto de mínimo é encontrado quando

$$\frac{\partial F}{\partial \alpha} = \sum_{i=1}^{NE} 2(y_i^e - \alpha x_i^e)(-x_i^e) = 0$$

Resolvendo a equação acima em termos de  $\alpha$ , é possível obter

$$\alpha = \frac{\sum_{i=1}^{NE} (y_i^e x_i^e)}{\sum_{i=1}^{NE} (x_i^e)^2}$$

Para o caso particular analisado,

$$\alpha = \frac{(10 \cdot 1050 + 20 \cdot 2000 + 30 \cdot 2950 + 40 \cdot 4000)}{(10^2 + 20^2 + 30^2 + 40^2)} = \frac{299000}{3000} = 99.67$$

Admitindo que os dados experimentais  $x_i$  têm erros desprezíveis, estando os erros de medição concentrados em  $y_i$ , e que as medidas experimentais são independentes, então é possível utilizar as Equações (1.37) e (1.44) para escrever

$$\text{Var}\{\alpha\} = \frac{\sum_{i=1}^{NE} (\text{Var}\{y_i^e\} [x_i^e]^2)}{\left[\sum_{i=1}^{NE} (x_i^e)^2\right]^2} \Rightarrow \sigma_\alpha^2 = \frac{\sum_{i=1}^{NE} (x_i^e)^2 \sigma_{y_i}^2}{\left[\sum_{i=1}^{NE} (x_i^e)^2\right]^2}$$

que relaciona os erros de medida experimental com os erros paramétricos. Se os erros de medição são iguais em todas as condições de experimentação

$$\sigma_{\alpha}^2 = \frac{\sum_{i=1}^{NE} (x_i^e)^2}{\left[ \sum_{i=1}^{NE} (x_i^e)^2 \right]^2} \sigma_y^2 = \frac{\sigma_y^2}{\sum_{i=1}^{NE} (x_i^e)^2} = \frac{\sigma_y^2}{3000}$$

Observe que a equação acima indica claramente que as incertezas experimentais viram incertezas paramétricas durante o processo de estimação de parâmetros. Observe ainda que a natureza das incertezas paramétricas está intimamente relacionada à natureza dos erros experimentais (se as hipóteses feitas em relação aos erros experimentais fossem diferentes, a fórmula acima não poderia ser escrita - há que se caracterizar de forma apropriada os erros experimentais!), à natureza do modelo (a equação acima só pode ser escrita para o modelo proposto - cada modelo dá origem a um problema novo!), à natureza da métrica (a equação acima só pode ser escrita para a função objetivo utilizada - cada métrica dá origem a um problema novo!) e aos valores experimentais medidos (se as medidas  $x_i$  mudarem, mudam os erros paramétricos; assim, é possível interferir no desempenho do modelo escolhendo bem as condições experimentais!).

Para o caso particular analisado, se  $\sigma_y^2 = 3000$  (cal/h)<sup>2</sup>, então

$$\sigma_{\alpha}^2 = \frac{3000}{3000} = 1 \Rightarrow \alpha = 99.67 \pm 2 \sigma_{\alpha} = 99.67 \pm 2$$

onde foi admitido comportamento normal com aproximadamente 95% de confiança. (O bom procedimento científico talvez nos obrigasse a escrever  $\alpha = 100 \pm 2$ , dado que não se deve usar mais casas decimais que as permitidas pela precisão experimental. Esse rigor será muitas vezes ignorado ao longo desse texto.) A hipótese de normalidade será discutida bastante ao longo dessa e das próximas seções.

O desempenho do modelo pode ser comparado às medidas experimentais na forma

$\dot{q}^e$ (cal / h)	1050	2000	2950	4000
$\dot{q}^m$ (cal / h)	996.7	1993.4	2990.1	3986.8
$\varepsilon = \dot{q}^e - \dot{q}^m$	53.3	6.6	-40.1	13.2

de maneira que a variância em torno do zero pode ser calculada como

$$\sigma_{\varepsilon}^2 = \frac{53.3^2 + 6.6^2 + 40.1^2 + 13.2^2}{3} = 1555.6$$

com três graus de liberdade. (O número de graus de liberdade perdidos durante o procedimento de estimação de parâmetros é igual ao número de parâmetros estimados. Essa questão será discutida com detalhes ao longo dessa seção.) Comparando-se a variância experimental da medida com a variância dos desvios observados, utilizando-se para isso o teste F com três graus de liberdade no numerador (desvios de predição observados), infinitos graus de liberdade no denominador (admite-se que a variância

experimental da medida é o valor verdadeiro, disponibilizado por estudo anterior), e 95% de confiança, verifica-se que

$$\frac{1}{F(\infty, 3; 0.975)} = \frac{1}{13.902} < \frac{\sigma_\varepsilon^2}{\sigma_y^2} = \frac{1555.6}{3000} = 0.518 < F(3, \infty; 0.975) = 3.1161$$

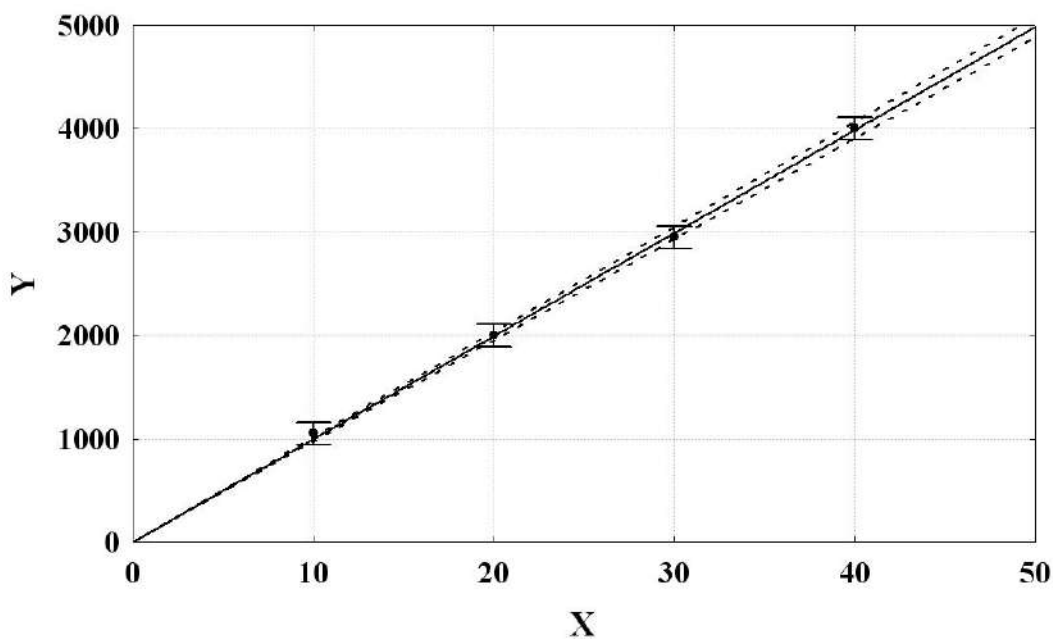
é satisfeita. Logo, a variâncias experimental característica das medidas de  $y$  e a variância dos desvios de predição não podem ser consideradas diferentes. Portanto, o modelo pode ser considerado bom, com incertezas de predição comparáveis às incertezas experimentais.

Mais ainda, admitindo-se a normalidade dos erros de medição, com aproximadamente 95% de confiança os erros de medida são da ordem de  $\pm 2 \sigma_y = \pm 2 \sqrt{3000} \approx \pm 110$ . Como nenhum dos desvios observados é maior que isso, não há pontos suspeitos ou "outliers". Assim, tudo indica que o procedimento de estimação foi bem executado.

Finalmente, como o modelo pode ser usado para fazer previsões de  $y$  *a posteriori*, é conveniente calcular os erros de previsão com o modelo. Nesse caso, usando novamente as Equações (1.37) e (1.44), chega-se a

$$\text{Var}\{\hat{y}\} = \text{Var}\{\alpha x\} \Rightarrow \hat{\sigma}_y^2 = x^2 \sigma_\alpha^2 = x^2$$

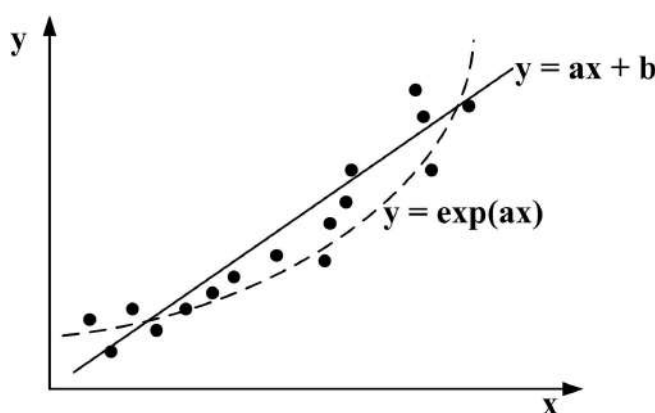
que é a variância de predição inerente ao modelo, uma vez que foram desprezados os possíveis erros experimentais de  $x$  e  $y$ . Toda essa informação está contida na Figura 4.12 abaixo. Observe como os erros de predição mudam com  $x$  e nesse caso particular crescem, à medida que nos afastamos do zero.



**Figura 4.12** - Resultados do procedimento de estimação do Exemplo 4.12. Barras verticais denotam os erros experimentais, a linha cheia é o modelo e as linhas tracejadas indicam o intervalo de confiança das previsões feitas com o modelo.

### 4.5. A Definição da Função Objetivo

Para que seja possível introduzir a noção de proximidade ou distância, é necessário primeiramente definir uma métrica. Para que se note a importância de introduzir uma definição precisa de uma métrica, a Figura 4.13 ilustra como pode ser difícil decidir que função está mais próxima dos dados experimentais quando uma transformação matemática precisa não está disponível.



**Figura 4.13** - Ilustração sobre a necessidade de definir uma métrica. Qual a função mais próxima dos dados experimentais?

Do ponto de vista estritamente matemático, define-se como uma função distância entre dois elementos quaisquer  $x$  e  $y$  de um conjunto, representada usualmente por  $d(x,y)$ , uma função que satisfaz os seguintes axiomas:

- a)  $d(x,y)$  é um número real estritamente positivo; ou seja,

$$d(x,y) \in \mathbb{R} \ ; \ d(x,y) \geq 0 \tag{4.5}$$

- b)  $d(x,y)$  é igual a zero se e somente se  $x=y$ ; ou seja,

$$d(x,y) = 0 \Leftrightarrow x = y \tag{4.6}$$

- c)  $d(x,y)$  é uma transformação comutativa; ou seja,

$$d(x,y) = d(y,x) \tag{4.7}$$

- d)  $d(x,y)$  satisfaz a desigualdade do triângulo; ou seja,

$$d(x,y) \leq d(x,z) + d(z,y) \tag{4.8}$$

É importante enfatizar que a necessidade de satisfazer os axiomas apresentados acima guarda estreito vínculo com a nossa idéia de distância física existente no mundo real. O primeiro axioma exige que a distância seja um número real positivo mensurável, como usual na nossa escala de valores. Por exemplo, alguém consegue imaginar o significado de se associar um número complexo ou um número negativo à distância entre duas cidades em um mapa geográfico? O segundo axioma generaliza a idéia de que se dois pontos são diferentes, então a distância entre eles não pode ser igual a zero. Da mesma forma, não parece fazer sentido associar um valor diferente de zero para a distância de um ponto a ele mesmo. O terceiro axioma generaliza a idéia de que a distância entre dois pontos quaisquer deve ser independente da escolha de qual deles é considerado como origem ou referência para a trajetória. Finalmente, o quarto axioma generaliza a idéia de que a trajetória mais curta possível entre dois pontos é aquela que liga diretamente esses dois pontos. Portanto, as Equações (4.5-8) introduzem formalmente conceitos com que já estamos bem acostumados de lidar no mundo real.

**Exemplo 4.13** - Seja o conjunto dos números reais. Sejam  $x$  e  $y$  dois números reais quaisquer. Então

$$d(x, y) = |x - y|$$

define uma métrica em  $\mathfrak{R}$ . Veja que o primeiro axioma é satisfeito naturalmente pela função módulo. O segundo axioma também é naturalmente satisfeito, porque o único número real que tem módulo igual a zero é o próprio zero. O terceiro axioma também é naturalmente satisfeito, uma vez que os módulos de números opostos (ou seja, de sinais distintos) são iguais. Finalmente, para mostrar a desigualdade do triângulo, suponha sem perda de generalidade que  $x < y$ . Então,

$$d(x, y) = |x - y| = |x - z + z - y| \begin{cases} < |x - z| + |z - y|, & z < x \\ = |x - z| + |z - y|, & x < z < y \\ < |x - z| + |z - y|, & z > y \end{cases}$$

Por exemplo, se  $x = 1$  e  $y = 3$

$$d(1, 3) = |1 - 3| = 2 \begin{cases} < |1 - 0| + |0 - 3| = 4 \\ = |1 - 2| + |2 - 3| = 2 \\ < |1 - 4| + |4 - 3| = 4 \end{cases}$$

Portanto, o valor absoluto da diferença entre dois números reais é uma medida da distância entre esses dois números.

Para o problema de estimação de parâmetros, os pontos  $\mathbf{x}$  e  $\mathbf{y}$  são na verdade os conjuntos de valores (vetores) que contêm os dados experimentais e os dados calculados com o modelo. Suponha que  $NE$  dados experimentais estão disponíveis e organizados em um vetor de dados experimentais ( $\mathbf{y}^e$ ) e estão sendo comparados a igual número de dados calculados com um modelo e organizados para representar condições



experimentais semelhantes ( $\mathbf{y}^m$ ). Nesse caso, pode-se considerar que  $\mathbf{y}^e$  e  $\mathbf{y}^m$  são elementos do  $\mathfrak{R}^{NE}$ ; ou seja, vetores de números reais com dimensão  $NE$ . Nesse caso, qualquer métrica utilizada para descrever uma distância em  $\mathfrak{R}^{NE}$  pode ser também usada para descrever a distância entre os dados experimentais e os dados calculados pelo modelo.

**Exemplo 4.14** - Sejam os conjuntos de dados experimentais e calculados, representados por  $\mathbf{y}^e$  e  $\mathbf{y}^m$ , constituídos por números reais. Então

$$d(\mathbf{y}^e, \mathbf{y}^m) = \left[ \sum_{i=1}^{NE} (y_i^e - y_i^m)^2 \right]^{1/2}$$

define uma métrica para o problema de estimação de parâmetros. Veja que o primeiro axioma é satisfeito naturalmente pela função quadrática, que resulta sempre num número real positivo. O segundo axioma também é naturalmente satisfeito, uma vez que o único número real cujo quadrado é igual a zero é o próprio zero. O terceiro axioma também é naturalmente satisfeito, uma vez que os quadrados de números opostos (ou seja, de sinais distintos) são iguais. Finalmente, para mostrar a desigualdade do triângulo, é conveniente lembrar apenas que

$$(y_i^e - y_i^m)^2 = |y_i^e - y_i^m|^2 \Rightarrow \sqrt{(y_i^e - y_i^m)^2} = |y_i^e - y_i^m|$$

de maneira que o resultado do Exemplo 4.13 também pode ser usado aqui para garantir a desigualdade do triângulo.

Portanto, a soma dos quadrados das diferenças entre as componentes de dois vetores reais é uma medida da distância entre esses dois vetores.

Os axiomas definidos pelas Equações (4.5-8) permitem definir um conjunto virtualmente infinito de métricas para o  $\mathfrak{R}^{NE}$ . Por exemplo, pode-se mostrar que as seguintes métricas satisfazem os axiomas propostos:

$$d_1(\mathbf{y}^e, \mathbf{y}^m) = \left[ \sum_{i=1}^{NE} (y_i^e - y_i^m)^2 \right]^{1/2} \quad (4.9a)$$

$$d_2(\mathbf{y}^e, \mathbf{y}^m) = \left[ \sum_{i=1}^{NE} (y_i^e - y_i^m)^N \right]^{1/N}, \quad N \text{ par} \quad (4.9b)$$

$$d_3(\mathbf{y}^e, \mathbf{y}^m) = \left[ \sum_{i=1}^{NE} w_i (y_i^e - y_i^m)^N \right]^{1/N}, \quad w_i \text{ positivo, } N \text{ par} \quad (4.9c)$$

$$d_4(\mathbf{y}^e, \mathbf{y}^m) = \left[ \sum_{i=1}^{NE} |y_i^e - y_i^m| \right] \quad (4.9d)$$

$$d_s(\mathbf{y}^e, \mathbf{y}^m) = \exp \left[ \sum_{i=1}^{NE} (y_i^e - y_i^m)^2 \right] - 1 \quad (4.9e)$$

Surge, portanto, intuitivamente a necessidade de perguntar qual deve ser a melhor métrica para descrever o problema de estimação de parâmetros. Do ponto de vista estritamente matemático, essa questão não faz qualquer sentido e todas as expressões acima (e infinitas outras) podem ser igualmente utilizadas para descrever a distância entre os pontos experimentais e os pontos obtidos com o auxílio do modelo. Contudo, um axioma adicional é imposto ao problema de estimação de parâmetros na forma:

e)  $d(\mathbf{x}, \mathbf{y})$  deve conter significação estatística.

Por exemplo, um exercício de derivação de uma métrica com significação estatística é apresentado a seguir. Suponha para isso que a estrutura do modelo está correta; ou seja, que o modelo é perfeito, embora os parâmetros do modelo sejam eventualmente desconhecidos. Essa hipótese, chamada de **Hipótese do Modelo Perfeito** será usada muitas vezes ao longo do texto e é apresentada a seguir.

#### Hipótese Fundamental 4.1 - A Hipótese do Modelo Perfeito

Admita que um modelo  $\mathbf{y} = \mathbf{f}(\mathbf{x}; \boldsymbol{\alpha})$  é usado para descrever um problema físico. O modelo é perfeito se é capaz de descrever exatamente as relações existentes entre as variáveis do problema. Nesse caso, as medidas experimentais não obedecem exatamente as relações impostas pelo modelo única e exclusivamente por causa dos inevitáveis desvios experimentais. Nesse caso,  $\mathbf{y}^e = \mathbf{f}(\mathbf{x}^e + \boldsymbol{\varepsilon}_x; \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_\alpha) + \boldsymbol{\varepsilon}_y$ .

A Hipótese do Modelo Perfeito é obviamente uma idealização sobre a compreensão do problema físico e impossível de ser atendida completamente, pois já foi extensamente discutido que nenhum modelo descreve todos os detalhes da realidade. Dessa forma, nenhum modelo pode ser de fato perfeito. Apesar disso, admitimos que a estrutura do modelo matemático utilizado para representar os dados experimentais é muito boa. Portanto, qualquer desvio eventualmente observado entre o dado experimental e o dado calculado com o modelo é devido única e exclusivamente às incertezas experimentais. Admita, portanto, que

$$y^e = y^m + \varepsilon$$

onde  $y^e$  é o valor observado experimentalmente,  $y^m$  é o valor calculado pelo modelo e  $\varepsilon$  é o desvio entre estes dois valores devido ao erro experimental. Na realidade,  $\varepsilon$  deve conter também os erros de modelagem, desprezados quando se utiliza a hipótese do modelo perfeito. No entanto, é muito difícil usar uma outra hipótese para o trabalho de estimação de parâmetros *a priori*, pois se os erros de modelagem fossem conhecidos, o modelo poderia ser melhorado e não haveria razão a princípio para se utilizar o modelo errado.

Se os erros são aleatórios e simétricos, espelhando um bom procedimento experimental, espera-se que eles tenham média igual a zero. Essa é uma das hipóteses fundamentais associadas à natureza dos dados experimentais, chamada de **Hipótese do Experimento Bem Feito**. Admite-se que, se o experimento é bem feito e o modelo é perfeito, o erro experimental não deve apresentar qualquer tipo de tendência ou **polarização**, flutuando em torno do valor zero. Dessa maneira, o valor médio esperado para o experimento é o próprio valor calculado com o modelo, pois

$$E\{y^e\} = E\{y^m + \varepsilon\} = E\{y^m\} + E\{\varepsilon\} = E\{y^m\} = y^m \quad (4.10)$$

Se o experimento é bem feito, não há por que acreditar que o erro experimental é maior que o erro mínimo inevitável, devido às incertezas experimentais. Em função da hipótese do experimento bem feito, uma avaliação da variância dos erros experimentais pode ser feita na forma

$$\sigma_y^2 = \frac{\sum_{i=1}^{NE} (y_i^e - y_i^m)^2}{\nu} \quad (4.11)$$

onde  $\nu$  é o número de graus de liberdade. A Equação (4.11), segundo o Exemplo 4.14, constitui uma métrica para o problema de estimação de parâmetros. Portanto, deseja-se minimizar a função

$$F_{Obj} = \sum_{i=1}^{NE} (y_i^e - y_i^m)^2 \quad (4.12)$$

que é uma métrica com interpretação estatística precisa.

#### **Hipótese Fundamental 4.2 - A Hipótese do Experimento Bem Feito**

Diz-se que o experimento é bem feito se os erros de medição cometidos durante a condução dos procedimentos experimentais são tão pequenos que é possível admitir que a probabilidade de encontrar os dados experimentais é máxima. Alternativamente, diz-se que o experimento é bem feito se os erros de medição cometidos durante a condução dos erros experimentais são equivalentes ao conteúdo mínimo de erro admissível para o processo de medição.

A Hipótese do Experimento Bem Feito permeia toda a análise estatística e numérica do problema de estimação de parâmetros. Do ponto de vista prático, é difícil não considerar essa hipótese durante a formulação do problema, já que a negação dessa hipótese condena de certa forma o conjunto experimental que está sendo analisado. Se os experimentos não são bem feitos, ou se os erros experimentais são muito grandes, parece razoável sugerir ao analista que os dados experimentais sejam medidos novamente; ou seja, recomenda-se a repetição dos experimentos.

A função objetivo definida na Equação (4.12) é uma medida do erro experimental, se as hipóteses do experimento bem feito e do modelo perfeito são

adequadas. Portanto, minimizar essa função objetivo é o mesmo que dizer que o erro experimental não deve ser superior ao menor valor possível, em consonância com a hipótese do experimento bem feito. A Equação (4.12) define a função objetivo de **mínimos quadrados**. Deve ficar claro que apenas a interpretação estatística torna essa métrica melhor que as demais métricas definidas pelas Equações (4.9a-e).

**Exemplo 4.15** - Um exemplo de aplicação da técnica de mínimos quadrados é o caso clássico da reta. Suponha que

$$y^m = \alpha x + \beta$$

A função de mínimos quadrados fica na forma

$$F_{Obj} = \sum_{i=1}^{NE} (y_i^e - \alpha x_i^e - \beta)^2$$

Os valores de  $\alpha$  e  $\beta$  são então obtidos minimizando-se o valor da função objetivo

$$\frac{\partial F_{Obj}}{\partial \alpha} = \sum_{i=1}^{NE} 2(y_i^e - \alpha x_i^e - \beta)(-x_i^e) = 0$$

$$\frac{\partial F_{Obj}}{\partial \beta} = \sum_{i=1}^{NE} 2(y_i^e - \alpha x_i^e - \beta)(-1) = 0$$

Há, portanto, duas equações a resolver e duas incógnitas a determinar, uma vez que os dados experimentais são conhecidos. Portanto:

$$\begin{cases} -2 \sum_{i=1}^{NE} (y_i^e x_i^e - \alpha (x_i^e)^2 - \beta x_i^e) = 0 \\ -2 \sum_{i=1}^{NE} (y_i^e - \alpha x_i^e - \beta) = 0 \end{cases}$$

$$\begin{cases} \alpha \sum_{i=1}^{NE} (x_i^e)^2 + \beta \sum_{i=1}^{NE} (x_i^e) = \sum_{i=1}^{NE} (y_i^e x_i^e) \\ \alpha \sum_{i=1}^{NE} (x_i^e) + \beta NE = \sum_{i=1}^{NE} (y_i^e) \end{cases}$$

resultando em

$$\alpha = \frac{NE \left[ \sum_{i=1}^{NE} (y_i^e x_i^e) \right] - \left[ \sum_{i=1}^{NE} (y_i^e) \right] \left[ \sum_{i=1}^{NE} (x_i^e) \right]}{NE \left[ \sum_{i=1}^{NE} (x_i^e)^2 \right] - \left[ \sum_{i=1}^{NE} (x_i^e) \right]^2}$$

$$\beta = \frac{\sum_{i=1}^{NE} (y_i^e)}{NE} - \alpha \frac{\sum_{i=1}^{NE} (x_i^e)}{NE}$$

**Exemplo 4.16** - Um outro exemplo de aplicação da técnica de mínimos quadrados é o caso clássico da parábola. Suponha que

$$y^m = \alpha x^2 + \beta x + \gamma$$

A função de mínimos quadrados fica na forma

$$F_{Obj} = \sum_{i=1}^{NE} \left( y_i^e - \alpha (x_i^e)^2 - \beta x_i^e - \gamma \right)^2$$

Os valores de  $\alpha$ ,  $\beta$  e  $\gamma$  são obtidos minimizando-se o valor da função objetivo

$$\frac{\partial F_{Obj}}{\partial \alpha} = \sum_{i=1}^{NE} 2 \left( y_i^e - \alpha (x_i^e)^2 - \beta x_i^e - \gamma \right) \left( -(x_i^e)^2 \right) = 0$$

$$\frac{\partial F_{Obj}}{\partial \beta} = \sum_{i=1}^{NE} 2 \left( y_i^e - \alpha (x_i^e)^2 - \beta x_i^e - \gamma \right) \left( -x_i^e \right) = 0$$

$$\frac{\partial F_{Obj}}{\partial \gamma} = \sum_{i=1}^{NE} 2 \left( y_i^e - \alpha (x_i^e)^2 - \beta x_i^e - \gamma \right) \left( -1 \right) = 0$$

Há, portanto, três equações a resolver e três incógnitas a determinar, uma vez que os dados experimentais são conhecidos. A solução analítica para esse problema pode ser facilmente derivada. Isto ocorre sempre que o modelo é linear nos parâmetros.

**Exemplo 4.17** - Um outro exemplo de aplicação da técnica de mínimos quadrados é o caso clássico da função exponencial. Suponha que

$$y = \alpha e^{\beta x}$$

A função de mínimos quadrados fica na forma

$$F_{Obj} = \sum_{i=1}^{NE} \left( y_i^e - \alpha e^{\beta x_i^e} \right)^2$$

Os valores de  $\alpha$  e  $\beta$  são obtidos minimizando-se o valor da função objetivo

$$\frac{\partial F_{Obj}}{\partial \alpha} = \sum_{i=1}^{NE} 2 \left( y_i^e - \alpha e^{\beta x_i^e} \right) \left( -e^{\beta x_i^e} \right) = 0$$

$$\frac{\partial F_{Obj}}{\partial \beta} = \sum_{i=1}^{NE} 2 \left( y_i^e - \alpha e^{\beta x_i^e} \right) \left( -\alpha x_i^e e^{\beta x_i^e} \right) = 0$$

Há, portanto, duas equações a resolver e duas incógnitas a determinar, uma vez que os dados experimentais são conhecidos. Uma solução analítica para esse problema não pode ser derivada. Isto ocorre por causa da natureza não linear do modelo. A solução do problema requer, portanto, o uso de técnicas numéricas como as discutidas no Capítulo 5. A necessidade de usar técnicas numéricas para resolver um problema supostamente tão simples mostra que não é possível, de maneira geral, conduzir estudos de estimação de parâmetros longe do computador. O usuário de procedimentos de estimação de parâmetros deve estar, portanto, habilitado a utilizar procedimentos numéricos de estimação.

Uma maneira comum de propor uma solução analítica para o problema é escrever o modelo na forma

$$z^m = \ln(y^m) = \ln(\alpha) + \beta x = \hat{\alpha} + \beta x$$

e a função objetivo na forma

$$F_{Obj} = \sum_{i=1}^{NE} \left( z_i^e - \hat{\alpha} - \beta x_i^e \right)^2$$

de maneira a poder usar a solução apresentada no Exemplo 4.15. Deve-se prestar a atenção para o fato, no entanto, de que toda a significação estatística da função objetivo pode ter sido jogada fora nesse caso, já que a variável medida efetivamente foi  $y$  e não  $z = \ln(y)$ . (Essa questão será discutida com um pouco mais de detalhes na Seção 5.8 do Capítulo 5.) Além disso, minimizar a função

$$F_{Obj} = \sum_{i=1}^{NE} \left( y_i^e - y_i^m \right)^2$$

não é equivalente a minimizar a função

$$F_{Obj} = \sum_{i=1}^{NE} \left( \ln(y_i^e) - \ln(y_i^m) \right)^2$$

e valores diferentes dos parâmetros estimados provavelmente serão encontrados. Por isso, o usuário deve resistir o máximo possível à tentativa de introduzir transformações dos dados experimentais. Voltaremos a esse ponto adiante.

Os Exemplos 4.15 e 4.16 mostram que, para modelos lineares nos parâmetros, a aplicação da técnica de mínimos quadrados admite solução analítica. Já para o modelo exponencial do Exemplo 4.17 é necessário algum método numérico para que a solução

seja encontrada. Assim, de forma generalizada pode-se definir um modelo linear nos parâmetros como:

$$y^m(\mathbf{x}, \boldsymbol{\alpha}) = \sum_{j=1}^{NP} \alpha_j f_j(\mathbf{x}) \quad (4.13)$$

O modelo proposto tem  $NP$  parâmetros,  $\boldsymbol{\alpha}^T = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_{NP}]$ , associados a  $NP$  funções  $f_j(\mathbf{x})$ , que transformam as  $NX$  variáveis independentes,  $\mathbf{x}^T = [x_1 \ x_2 \ \dots \ x_{NX}]$ , na variável dependente  $y$ . Cada uma das variáveis é medida em cada uma das  $NE$  condições experimentais. A função de mínimos quadrados fica então na forma

$$F_{Obj} = \sum_{i=1}^{NE} \left( y_i^e - \sum_{j=1}^{NP} \alpha_j f_j(\mathbf{x}_i^e) \right)^2 \quad (4.14)$$

Os valores de  $\boldsymbol{\alpha}$  são obtidos minimizando-se o valor da função objetivo em relação a cada um de seus componentes, na forma

$$\frac{\partial F_{Obj}}{\partial \alpha_k} = \sum_{i=1}^{NE} 2 \left( y_i^e - \sum_{j=1}^{NP} \alpha_j f_j(\mathbf{x}_i^e) \right) (-f_k(\mathbf{x}_i^e)) = 0, \quad k = 1..NP \quad (4.15)$$

que resulta no seguinte sistema de equações

$$\sum_{j=1}^{NP} \alpha_j \left( \sum_{i=1}^{NE} f_j(\mathbf{x}_i^e) f_k(\mathbf{x}_i^e) \right) = \sum_{i=1}^{NE} y_i^e f_k(\mathbf{x}_i^e), \quad k = 1..NP \quad (4.16)$$

Há, portanto,  $NP$  equações a resolver e  $NP$  incógnitas a determinar, uma vez que os dados experimentais são conhecidos. A solução analítica para esse problema pode ser facilmente derivada se a notação matricial é utilizada. Sejam

$$\mathbf{M} = \begin{bmatrix} \sum_{i=1}^{NE} f_1(\mathbf{x}_i^e) f_1(\mathbf{x}_i^e) & \sum_{i=1}^{NE} f_1(\mathbf{x}_i^e) f_2(\mathbf{x}_i^e) & \dots & \sum_{i=1}^{NE} f_1(\mathbf{x}_i^e) f_{NP}(\mathbf{x}_i^e) \\ \sum_{i=1}^{NE} f_2(\mathbf{x}_i^e) f_1(\mathbf{x}_i^e) & \sum_{i=1}^{NE} f_2(\mathbf{x}_i^e) f_2(\mathbf{x}_i^e) & \dots & \sum_{i=1}^{NE} f_2(\mathbf{x}_i^e) f_{NP}(\mathbf{x}_i^e) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{NE} f_{NP}(\mathbf{x}_i^e) f_1(\mathbf{x}_i^e) & \sum_{i=1}^{NE} f_{NP}(\mathbf{x}_i^e) f_2(\mathbf{x}_i^e) & \dots & \sum_{i=1}^{NE} f_{NP}(\mathbf{x}_i^e) f_{NP}(\mathbf{x}_i^e) \end{bmatrix} \quad (4.17)$$

e

$$\mathbf{Y}_f = \begin{bmatrix} \sum_{i=1}^{NE} y_i^e f_1(\mathbf{x}_i^e) \\ \sum_{i=1}^{NE} y_i^e f_2(\mathbf{x}_i^e) \\ \dots \\ \sum_{i=1}^{NE} y_i^e f_{NP}(\mathbf{x}_i^e) \end{bmatrix} \quad (4.18)$$

então

$$\mathbf{M} \boldsymbol{\alpha} = \mathbf{Y}_f \Rightarrow \boldsymbol{\alpha} = \mathbf{M}^{-1} \mathbf{Y}_f \quad (4.19)$$

que é uma solução de enorme importância prática para a teoria de estimação de parâmetros e planejamento de experimentos. É muito conveniente ainda definir a matriz

$$\mathbf{G}_Y = \begin{bmatrix} f_1(\mathbf{x}_1^e) & f_1(\mathbf{x}_2^e) & \dots & f_1(\mathbf{x}_{NE}^e) \\ f_2(\mathbf{x}_1^e) & f_2(\mathbf{x}_2^e) & \dots & f_2(\mathbf{x}_{NE}^e) \\ \dots & \dots & \dots & \dots \\ f_{NP}(\mathbf{x}_1^e) & f_{NP}(\mathbf{x}_2^e) & \dots & f_{NP}(\mathbf{x}_{NE}^e) \end{bmatrix} \quad (4.20)$$

pois assim é possível escrever a solução diretamente em termos das variáveis medidas na forma

$$\boldsymbol{\alpha} = \mathbf{M}^{-1} \mathbf{G}_Y \mathbf{Y}^e, \quad \mathbf{Y}^e = \begin{bmatrix} y_1^e \\ y_2^e \\ \vdots \\ y_{NE}^e \end{bmatrix} \quad (4.21)$$

que indica que existe uma relação linear direta entre a medida experimental da variável dependente e o valor estimado para o parâmetro.

**Exemplo 4.18** - Suponha que o seguinte modelo está sendo usado para interpretar um problema físico

$$y^m = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2 + \alpha_4$$

onde  $y$  representa uma variável dependente que depende de duas outras variáveis independentes  $x_1$  e  $x_2$ . No problema proposto,

$$f_1(\mathbf{x}^e) = x_1^e, \quad f_2(\mathbf{x}^e) = x_2^e, \quad f_3(\mathbf{x}^e) = x_1^e x_2^e, \quad f_4(\mathbf{x}^e) = 1$$

Dessa forma,



$$\mathbf{G}_Y = \begin{bmatrix} x_{1,1}^e & x_{1,2}^e & x_{1,3}^e & \cdots & x_{1,NE}^e \\ x_{2,1}^e & x_{2,2}^e & x_{2,3}^e & \cdots & x_{2,NE}^e \\ x_{1,1}^e x_{2,1}^e & x_{1,2}^e x_{2,2}^e & x_{1,3}^e x_{2,3}^e & \cdots & x_{1,NE}^e x_{2,NE}^e \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$$

$$\mathbf{M} = \begin{bmatrix} \sum_{k=1}^{NE} (x_{1,k}^e)^2 & \sum_{k=1}^{NE} x_{1,k}^e x_{2,k}^e & \sum_{k=1}^{NE} (x_{1,k}^e)^2 x_{2,k}^e & \sum_{k=1}^{NE} x_{1,k}^e \\ \sum_{k=1}^{NE} x_{1,k}^e x_{2,k}^e & \sum_{k=1}^{NE} (x_{2,k}^e)^2 & \sum_{k=1}^{NE} x_{1,k}^e (x_{2,k}^e)^2 & \sum_{k=1}^{NE} x_{2,k}^e \\ \sum_{k=1}^{NE} (x_{1,k}^e)^2 x_{2,k}^e & \sum_{k=1}^{NE} x_{1,k}^e (x_{2,k}^e)^2 & \sum_{k=1}^{NE} (x_{1,k}^e)^2 (x_{2,k}^e)^2 & \sum_{k=1}^{NE} x_{1,k}^e x_{2,k}^e \\ \sum_{k=1}^{NE} x_{1,k}^e & \sum_{k=1}^{NE} x_{2,k}^e & \sum_{k=1}^{NE} x_{1,k}^e x_{2,k}^e & \sum_{i=1}^{NE} 1 = NE \end{bmatrix}$$

onde  $x_{i,k}^e$  representa a medida da variável independente  $x_i$  no experimento  $k$ . Portanto, de acordo com a Equação (4.21), existe uma solução analítica explícita para o problema de estimação de parâmetros, uma vez conhecidos os dados experimentais. Modelos como esse são muito úteis para interpretação quantitativa de dados experimentais, como discutido no Volume II dessa série de publicações.

Suponha que as variáveis independentes não contêm erros e que toda a flutuação experimental é devida aos erros de medição das variáveis dependentes  $\mathbf{Y}^e$ . Suponha ainda que dois conjuntos de dados obtidos em condições análogas são comparados entre si. Nesse caso,

$$\Delta \boldsymbol{\alpha} = \boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2 = \mathbf{M}^{-1} \mathbf{G}_Y \mathbf{Y}_1^e - \mathbf{M}^{-1} \mathbf{G}_Y \mathbf{Y}_2^e = \mathbf{M}^{-1} \mathbf{G}_Y \Delta \mathbf{Y}^e \quad (4.22)$$

Portanto, os valores dos parâmetros flutuam, à medida que flutuam os valores experimentais obtidos. Se a única fonte de flutuação é o erro experimental

$$\Delta \boldsymbol{\alpha} = \mathbf{M}^{-1} \mathbf{G}_Y \boldsymbol{\varepsilon} \quad (4.23)$$

que relaciona os erros paramétricos com os erros experimentais. Portanto, se os erros flutuam em torno dos valores verdadeiros com média igual a zero,

$$E \{ \Delta \boldsymbol{\alpha} \} = E \{ \mathbf{M}^{-1} \mathbf{G}_Y \boldsymbol{\varepsilon} \} = \mathbf{M}^{-1} \mathbf{G}_Y E \{ \boldsymbol{\varepsilon} \} = \mathbf{0} \quad (4.24)$$

os parâmetros também flutuam em torno dos valores verdadeiros, o que mostra que o procedimento proposto para estimação de parâmetros é **consistente**. Mais importante ainda é perceber que a matriz de covariâncias dos parâmetros pode ser calculada como

$$\begin{aligned} \mathbf{V}_a &= E\{\Delta\mathbf{a} \Delta\mathbf{a}^T\} = E\left\{ \begin{bmatrix} (\Delta\alpha_1)^2 & \Delta\alpha_1 \Delta\alpha_2 & \cdots & \Delta\alpha_1 \Delta\alpha_{NP} \\ \Delta\alpha_2 \Delta\alpha_1 & (\Delta\alpha_2)^2 & \cdots & \Delta\alpha_2 \Delta\alpha_{NP} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta\alpha_{NP} \Delta\alpha_1 & \Delta\alpha_{NP} \Delta\alpha_2 & \cdots & (\Delta\alpha_{NP})^2 \end{bmatrix} \right\} = \\ &= E\left\{ \mathbf{M}^{-1} \mathbf{G}_Y \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \mathbf{G}_Y^T (\mathbf{M}^{-1})^T \right\} = \mathbf{M}^{-1} \mathbf{G}_Y E\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\} \mathbf{G}_Y^T (\mathbf{M}^{-1})^T \quad (4.25) \\ &= \mathbf{M}^{-1} \mathbf{G}_Y \mathbf{V}_y \mathbf{G}_Y^T (\mathbf{M}^{-1})^T \end{aligned}$$

Observe que a matriz  $\mathbf{M}$  é simétrica, de maneira que ela é igual a sua transposta ( $\mathbf{M} = \mathbf{M}^T$ ,  $\mathbf{M}^{-1} = (\mathbf{M}^{-1})^T$ ). Observe ainda que

$$\mathbf{G}_Y \mathbf{G}_Y^T = \begin{bmatrix} \sum_{i=1}^{NE} f_1(\mathbf{x}_i^e) f_1(\mathbf{x}_i^e) & \sum_{i=1}^{NE} f_1(\mathbf{x}_i^e) f_2(\mathbf{x}_i^e) & \cdots & \sum_{i=1}^{NE} f_1(\mathbf{x}_i^e) f_{NP}(\mathbf{x}_i^e) \\ \sum_{i=1}^{NE} f_2(\mathbf{x}_i^e) f_1(\mathbf{x}_i^e) & \sum_{i=1}^{NE} f_2(\mathbf{x}_i^e) f_2(\mathbf{x}_i^e) & \cdots & \sum_{i=1}^{NE} f_2(\mathbf{x}_i^e) f_{NP}(\mathbf{x}_i^e) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{NE} f_{NP}(\mathbf{x}_i^e) f_1(\mathbf{x}_i^e) & \sum_{i=1}^{NE} f_{NP}(\mathbf{x}_i^e) f_2(\mathbf{x}_i^e) & \cdots & \sum_{i=1}^{NE} f_{NP}(\mathbf{x}_i^e) f_{NP}(\mathbf{x}_i^e) \end{bmatrix} = \mathbf{M} \quad (4.26)$$

Dessa forma, se as medidas experimentais são independentes e os erros de medição são constantes e iguais em todas as condições experimentais, então a matriz de erros experimentais pode ser escrita como

$$\mathbf{V}_y = \begin{bmatrix} \sigma_y^2 & 0 & \cdots & 0 \\ 0 & \sigma_y^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_y^2 \end{bmatrix} = \sigma_y^2 \mathbf{I} \quad (4.27)$$

onde  $\mathbf{I}$  é a matriz identidade. Nesse caso, a matriz de covariâncias dos parâmetros ganha a forma bastante simples

$$\mathbf{V}_a = \sigma_y^2 \mathbf{M}^{-1} \mathbf{G}_Y \mathbf{G}_Y^T (\mathbf{M}^{-1})^T = \sigma_y^2 \mathbf{M}^{-1} \mathbf{M} \mathbf{M}^{-1} = \sigma_y^2 \mathbf{M}^{-1} \quad (4.28)$$

que é utilizada para a interpretação e solução de um grande número de problemas práticos. A Equação (4.28) mostra que a incerteza dos parâmetros depende dos erros experimentais e das condições de experimentação; portanto, é possível alterar os erros paramétricos através de manipulação apropriadas das condições de experimentação. A Equação (4.28) constitui a base fundamental sobre a qual foi erigida boa parte dos procedimentos de planejamento estatístico de experimentos. É importante observar, no entanto, que a Equação (4.28) é rigorosamente válida apenas quando uma longa série de condições é satisfeita: o modelo é perfeito, os experimentos são bem feitos, a função objetivo é dada pela função de mínimos quadrados, o modelo é linear nos parâmetros,

os experimentos são independentes e os erros experimentais são constantes na região de experimentação.

Finalmente, os parâmetros são usados para fazer previsões com o modelo em qualquer condição  $\mathbf{x}$ . Usando a notação matricial, o modelo pode ser escrito na forma

$$y^m(\mathbf{x}, \boldsymbol{\alpha}) = \mathbf{B}(\mathbf{x})^T \boldsymbol{\alpha} \quad (4.29)$$

onde

$$\mathbf{B}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_{NP}(\mathbf{x}) \end{bmatrix} \quad (4.30)$$

que também estabelece uma relação linear direta entre o valor dos parâmetros e a previsão da variável dependente em um ponto qualquer  $\mathbf{x}$  da região experimental. A matriz  $\mathbf{B}$  é chamada de **matriz de sensibilidades** do modelo em relação aos parâmetros. De forma análoga à realizada anteriormente,

$$\begin{aligned} \hat{\sigma}_y^2 &= E\{(\Delta y^c)^2\} = E\{\mathbf{B}^T \Delta \boldsymbol{\alpha} \Delta \boldsymbol{\alpha}^T \mathbf{B}\} \\ \hat{\sigma}_y^2 &= \mathbf{B}^T E\{\Delta \boldsymbol{\alpha} \Delta \boldsymbol{\alpha}^T\} \mathbf{B} = \mathbf{B}^T \mathbf{V}_\alpha \mathbf{B} = \sigma_y^2 \mathbf{B}^T \mathbf{M}^{-1} \mathbf{B} \end{aligned} \quad (4.31)$$

que estabelece o vínculo entre os erros de predição, a qualidade dos dados experimentais e o procedimento de estimação de parâmetros. Todas essas expressões são fundamentais para a perfeita compreensão dos procedimentos clássicos de planejamento experimental, como discutido no Volume II dessa série de publicações.

**Exemplo 4.19** – No problema tratado no Exemplo 4.18, a matriz de sensibilidades do modelo em relação aos parâmetros é dada por

$$\mathbf{B}(\mathbf{x}) = \begin{bmatrix} x_1 \\ x_2 \\ x_1 x_2 \\ 1 \end{bmatrix}$$

Repare que a matriz  $\mathbf{B}$  depende da condição de experimentação considerada durante os cálculos.

A Equação (4.31) é muito importante para considerações filosóficas a respeito da propagação de erros e da caracterização da natureza dos erros cometidos em trabalhos de engenharia. Observe que a Equação (4.31) vincula os erros dos parâmetros, resultantes do procedimento de estimação de parâmetros e dos erros experimentais

cometidos no passado, com os erros de predição ou de simulação, que dizem respeito a experimentos ainda não realizados (ou a serem realizados no futuro). Portanto, os erros experimentais cometidos nas medições experimentais passadas se propagam para o futuro indefinidamente, uma vez que os parâmetros são usados para o projeto de novas unidades e simulações (extrapolações) de novas condições experimentais. Portanto, parece muito claro que a correta caracterização dos erros experimentais é fundamental para a interpretação do desempenho dos modelos de simulação e das rotinas de projeto, que usarão os parâmetros estimados a partir desses mesmos dados experimentais em outras situações.

**Exemplo 4.20** - Os problemas tratados nos Exemplos 4.15 e 4.16 podem ser analisados no contexto desenvolvido de forma generalizada para modelos lineares nos parâmetros. Por exemplo, a reta pode ser escrita como

$$y^m = \alpha_1 f_1(x) + \alpha_2 f_2(x)$$

com  $f_1(x) = x$  e  $f_2(x) = 1$ . Nesse caso, conforme a Equação (4.17), a matriz  $\mathbf{M}$  pode ser escrita como:

$$\mathbf{M} = \begin{bmatrix} \sum_{i=1}^{NE} (x_i^e)^2 & \sum_{i=1}^{NE} (x_i^e) \\ \sum_{i=1}^{NE} (x_i^e) & NE \end{bmatrix}$$

e a sua inversa é

$$\mathbf{M}^{-1} = \frac{1}{NE \left[ \sum_{i=1}^{NE} (x_i^e)^2 \right] - \left[ \sum_{i=1}^{NE} (x_i^e) \right]^2} \begin{bmatrix} NE & -\sum_{i=1}^{NE} (x_i^e) \\ -\sum_{i=1}^{NE} (x_i^e) & \sum_{i=1}^{NE} (x_i^e)^2 \end{bmatrix}$$

Logo, a variância do parâmetro  $\alpha_1$  é dada por

$$\sigma_{11}^2 = \sigma_y^2 \frac{NE}{NE \left[ \sum_{i=1}^{NE} (x_i^e)^2 \right] - \left[ \sum_{i=1}^{NE} (x_i^e) \right]^2}$$

enquanto a variância do parâmetro  $\alpha_2$  é dada por

$$\sigma_{22}^2 = \sigma_y^2 \frac{\left[ \sum_{i=1}^{NE} (x_i^e)^2 \right]}{NE \left[ \sum_{i=1}^{NE} (x_i^e)^2 \right] - \left[ \sum_{i=1}^{NE} (x_i^e) \right]^2}$$

É muito curioso observar que os parâmetros  $\alpha_1$  e  $\alpha_2$  não são necessariamente independentes, apresentando uma covariância igual a

$$\sigma_{12}^2 = \sigma_y^2 \frac{-\left[\sum_{i=1}^{NE} (x_i^e)\right]}{NE \left[\sum_{i=1}^{NE} (x_i^e)^2\right] - \left[\sum_{i=1}^{NE} (x_i^e)\right]^2}$$

e um coeficiente de correlação igual a

$$\rho_{12} = \frac{-\left[\sum_{i=1}^{NE} (x_i^e)\right]}{\sqrt{NE \left[\sum_{i=1}^{NE} (x_i^e)^2\right]}}$$

Isso quer dizer que um parâmetro influencia o outro; ou seja, se um dos parâmetros mudar um pouco, o outro também muda. Isso em geral é ruim, pois mistura a importância dos diferentes efeitos considerados por cada um dos parâmetros do modelo. O ideal seria obter parâmetros independentes, embora isso raramente seja possível. A necessidade de obter parâmetros independentes durante a análise de modelos de simulação é um ponto central dos procedimentos de planejamento experimental discutidos no Volume II dessa série de publicações.

Com relação ao erro de predição, definido pela Equação (4.31), pode-se escrever:

$$\hat{\sigma}_y^2(x) = \sigma_y^2 \mathbf{B}^T \mathbf{M}^{-1} \mathbf{B} = \sigma_y^2 \begin{bmatrix} x & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} = \sigma_y^2 (\sigma_{11}^2 x^2 + 2\sigma_{12}^2 x + \sigma_{22}^2)$$

$$\hat{\sigma}_y^2(x) = \sigma_y^2 \frac{x^2 NE - 2x \left[\sum_{i=1}^{NE} (x_i^e)\right] + \left[\sum_{i=1}^{NE} (x_i^e)^2\right]}{NE \left[\sum_{i=1}^{NE} (x_i^e)^2\right] - \left[\sum_{i=1}^{NE} (x_i^e)\right]^2}$$

onde pode ser observado que mesmo para um modelo linear, o erro de predição é uma função quadrática com relação à condição experimental  $x$ .

É interessante observar nas expressões dos erros apresentadas acima que os erros paramétricos aumentam sempre com o aumento dos erros experimentais, o que poderia já ser esperado, e diminuem sempre com o aumento do número de experimentos. Portanto, à medida que o número de experimentos aumenta indefinidamente, os erros paramétricos se aproximam continuamente de zero.

**Exemplo 4.21** – Suponha que um conjunto de dados experimentais  $\mathbf{y} = [y_1^e \ y_2^e \ \dots \ y_{NE}^e]$  está disponível e que se deseja representar o sistema pelo modelo constante  $y^m = \alpha$ . Nesse caso, segundo a Equação (4.17),  $\mathbf{M} = [NE]$ . Segundo a Equação (4.20),  $\mathbf{G}_Y = [1 \ 1 \ \dots \ 1]$ . Portanto, segundo a Equação (4.21)

$$\alpha = \frac{\sum_{k=1}^{NE} y_k^e}{NE}$$

Logo, a média amostral pode ser interpretada como a melhor inferência de um modelo constante para um conjunto de dados, quando se supõe que o modelo constante é perfeito, os experimentos são bem feitos, a função objetivo é dada pela função de mínimos quadrados, os experimentos são independentes e os erros experimentais são constantes na região de experimentação. Vê-se, portanto, que o contexto de validade da média amostral proposta no Capítulo 3 pode ser bastante questionado, em bases técnicas absolutamente legítimas.

**Exemplo 4.22** – Uma pergunta pertinente diz respeito à variância das diferenças entre os dados experimentais e as predições feitas com o modelo. Suponha um modelo linear na forma

$$y = \mathbf{B}(\mathbf{x})^T \boldsymbol{\alpha}$$

como discutido nessa seção. Nesse caso, as respostas do modelo são obtidas na forma

$$y^m = \mathbf{B}(\mathbf{x})^T \mathbf{M}^{-1} \mathbf{G}_Y \mathbf{Y}^e$$

Em particular, a diferença entre os dados experimentais e calculados no ponto experimental  $k$  pode ser dada na forma:

$$y_k^e - y_k^m = y_k^e - \mathbf{B}_k^T \mathbf{M}^{-1} \mathbf{G}_Y \mathbf{Y}^e$$

Supondo que o modelo é perfeito e que os experimentos são bem feitos, é possível escrever

$$y_k^e - y_k^m = y_k^e - \mathbf{B}_k^T \mathbf{M}^{-1} \mathbf{G}_Y (\boldsymbol{\mu}_Y + \boldsymbol{\varepsilon}) = (y_k^e - \mu_k) - \mathbf{B}_k^T \mathbf{M}^{-1} \mathbf{G}_Y \boldsymbol{\varepsilon}$$

onde  $\boldsymbol{\mu}_Y$  representa os valores verdadeiros e desconhecidos das medidas experimentais, por causa do erro experimental  $\boldsymbol{\varepsilon}$ . Portanto,  $\mu_k$  representa o valor verdadeiro e desconhecido da medida experimental  $y_k^e$ . Portanto, a variância entre a medida experimental e a previsão do modelo no ponto experimental  $k$  pode ser dada por

$$\begin{aligned} \text{Var}\{y_k^e - y_k^m\} &= E\left\{\left[(y_k^e - \mu_k) - \mathbf{B}_k^T \mathbf{M}^{-1} \mathbf{G}_Y \boldsymbol{\varepsilon}\right]^2\right\} = \\ &E\left\{\left(y_k^e - \mu_k\right)^2 - 2\mathbf{B}_k^T \mathbf{M}^{-1} \mathbf{G}_Y \boldsymbol{\varepsilon} (y_k^e - \mu_k) + \mathbf{B}_k^T \mathbf{M}^{-1} \mathbf{G}_Y \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \mathbf{G}_Y^T (\mathbf{M}^{-1})^T \mathbf{B}_k\right\} \end{aligned}$$

Se as medidas experimentais são independentes e os erros a que as medidas estão sujeitas são os mesmos,

$$\begin{aligned} \text{Var}\{y_k^e - y_k^m\} &= \\ &E\left\{\left(y_k^e - \mu_k\right)^2\right\} - 2\mathbf{B}_k^T \mathbf{M}^{-1} \mathbf{G}_Y E\left\{\boldsymbol{\varepsilon} (y_k^e - \mu_k)\right\} + \mathbf{B}_k^T \mathbf{M}^{-1} \mathbf{G}_Y E\left\{\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T\right\} \mathbf{G}_Y^T (\mathbf{M}^{-1})^T \mathbf{B}_k = \\ &\sigma^2 - \sigma^2 \mathbf{B}_k^T \mathbf{M}^{-1} \mathbf{B}_k = \sigma^2 \left[1 - \mathbf{B}_k^T \mathbf{M}^{-1} \mathbf{B}_k\right] < \sigma^2 = E\left\{\left(y_k^e - \mu_k\right)^2\right\} \end{aligned}$$

onde  $\sigma^2$  é a variância dos erros experimentais. Conclui-se, portanto, que a variância das diferenças entre os dados experimentais e as predições feitas com o modelo é menor que a variância dos erros experimentais. Assim, se a variância dos erros experimentais for inferida pela diferença existente entre os dados experimentais e as predições do modelo, é necessário levar esse fato em consideração, para que não se subestime a variância experimental (ver Equações 3.7-3.9).

De forma semelhante

$$\begin{aligned} E\left\{\sum_{k=1}^{NE} (y_k^e - y_k^m)^2\right\} &= \sum_{k=1}^{NE} E\left\{(y_k^e - y_k^m)^2\right\} = \sum_{k=1}^{NE} \text{Var}\{y_k^e - y_k^m\} = \\ &\sigma^2 \left[NE - \sum_{k=1}^{NE} \mathbf{B}_k^T \mathbf{M}^{-1} \mathbf{B}_k\right] = \sigma^2 [NE - NP] \end{aligned}$$

Portanto, uma inferência consistente da variância experimental pode ser dada pela equação

$$s^2 = \frac{\sum_{k=1}^{NE} (y_k^e - y_k^m)^2}{NE - NP}$$

de maneira que se diz que o sistema perde  $NP$  graus de liberdade quando se estimam  $NP$  parâmetros de um modelo. No caso particular do modelo do Exemplo 4.21, obtém-se a mesma expressão definida na Equação (3.7) para a variância amostral, mostrando a consistência interna da análise efetuada.

## 4.6. O Método da Máxima Verossimilhança

Apesar de ser bastante útil e permitir a solução de uma série de problemas práticos, como nos Exemplos 4.15 a 4.22, a função de mínimos quadrados definida pela Equação (4.12) é bastante limitada porque admite implicitamente que todas as variáveis

analisadas pertencem a um mesmo conjunto amostral; ou seja, são medidas de uma mesma variável, obtidas com a mesma precisão em qualquer condição experimental. Nem uma coisa nem outra são necessariamente verdadeiras. Por exemplo, durante a análise de dados de reação, podem ser medidas temperaturas e pressões como variáveis dependentes. Obviamente, não faz qualquer sentido misturar os dados de temperatura e pressão, como definido pelas Equações (4.11) e (4.12). Além disso, os erros de medição de cada uma das variáveis podem mudar de ponto para ponto, em virtude de mudanças de desempenho dos instrumentos de medida e das técnicas experimentais. É importante observar que nem todo instrumento tem o desempenho de uma régua, que fornece um erro de medida aproximadamente constante em toda a faixa de utilização. Por exemplo, a incerteza de medição de termopares usados para medir temperaturas costuma aumentar com o aumento da temperatura. Dessa forma, também não faz sentido juntar as medidas de temperatura obtidas em condições distintas nas Equações (4.11) e (4.12). Finalmente, as medidas experimentais não são necessariamente independentes. Uma medida pode influenciar a outra de maneira direta (por exemplo, flutuações de temperatura influenciam o desempenho do medidor de pressão) ou de maneira indireta (flutuações dos níveis de impureza resultantes de certas corridas experimentais podem afetar os resultados obtidos nas corridas seguintes). Portanto, é necessária uma maneira alternativa de formular a função objetivo, que contemple a possibilidade de levar todos esses fatos em consideração.

Uma maneira bastante comum de se efetuar a estimação de parâmetros é fazer uso do método da máxima verossimilhança. O método da máxima verossimilhança está baseado em pressupostos relativamente simples e permite a análise de virtualmente qualquer problema experimental de forma rigorosa, desde que se conheça de forma apropriada como se comportam os erros de medição na região experimental. Os pressupostos para construção do método da máxima verossimilhança são apresentados a seguir.

**Pressuposto 1** – Admite-se que as distribuições dos erros experimentais na região de experimentação são conhecidas.

Nesse caso, suponha que a curva de densidade de probabilidades  $\wp(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_z)$  descreve as probabilidades de se encontrarem as medidas experimentais  $\mathbf{z}^e$ , dado os valores reais (e desconhecidos)  $\mathbf{z}$  e uma medida da variância dos erros experimentais  $\mathbf{V}_z$ . O vetor  $\mathbf{z}$  contém as variáveis independentes e as dependentes; ou seja,  $\mathbf{z}^T = [\mathbf{x}^T \ \mathbf{y}^T]$ . Obviamente, os desvios  $\boldsymbol{\varepsilon}_z = (\mathbf{z}^e - \mathbf{z})$  são os erros experimentais. Por exemplo, no caso dos desvios experimentais apresentarem uma distribuição de probabilidades normal, tem-se, de acordo com a Equação (2.72), a seguinte curva de densidade de probabilidade:

$$\wp(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_z) = \frac{1}{\sqrt{2\pi \det(\mathbf{V}_z)}} \exp \left[ -\frac{1}{2} (\mathbf{z}^e - \mathbf{z})^T \mathbf{V}_z^{-1} (\mathbf{z}^e - \mathbf{z}) \right] \quad (4.32a)$$

Deve ficar claro que diferentes experimentos podem apresentar diferentes distribuições de erros experimentais, de maneira que a Equação (4.32a) apenas ilustra um caso, em que as flutuações ocorrem de acordo com o que prevê a curva de distribuição normal. Por exemplo, se a distribuição de erros experimentais puder ser



descrita por uma distribuição exponencial, a curva de densidade de probabilidade dos desvios experimentais poderia ser descrita na forma:

$$\wp(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_z) = \frac{1}{2\alpha(\mathbf{V}_z)} \exp \left[ -\frac{\|\mathbf{z}^e - \mathbf{z}\|}{\alpha(\mathbf{V}_z)} \right] \quad (4.32b)$$

onde  $\|\mathbf{z}^e - \mathbf{z}\|$  representa uma norma apropriada dos desvios experimentais e  $\alpha(\mathbf{V}_z)$  representa um escalar que pondera a magnitude dos desvios experimentais.

Não é possível escolher qual das duas equações (Equação 4.32a-b) é melhor para representar os erros experimentais sem que se faça uma correta caracterização dos erros de medição no laboratório, como discutido no Capítulo 3 (e no Volume II dessa série de publicações). Na verdade, outras funções de distribuição, como aquelas apresentadas no Capítulo 2, podem também ser usadas para descrever de forma apropriada os erros experimentais. Assim, as Equações (4.32a-b) são apenas dois exemplos possíveis de comportamento em um universo virtualmente infinito de possibilidades.

Freqüentemente os experimentos são realizados de forma independente. Nesse caso, a curva de densidade de probabilidades que descreve o conjunto de observações experimentais pode ser descrita na forma

$$\wp(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_z) = \prod_{i=1}^{NE} \wp_i(\mathbf{z}_i^e; \mathbf{z}_i, \mathbf{V}_{z_i}) \quad (4.33a)$$

que explicita o fato de que a probabilidade conjunta das observações é resultado da composição das diferentes probabilidades de cada um dos resultados obtidos. Dessa maneira, as Equações (4.32a-b) podem ser rescritas na forma:

$$\wp(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_z) = \prod_{i=1}^{NE} \left\{ \frac{1}{\sqrt{2\pi \det(\mathbf{V}_{z_i})}} \exp \left[ -\frac{1}{2} (\mathbf{z}_i^e - \mathbf{z}_i)^T \mathbf{V}_{z_i}^{-1} (\mathbf{z}_i^e - \mathbf{z}_i) \right] \right\} \quad (4.33b)$$

$$\wp(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_z) = \prod_{i=1}^{NE} \left\{ \frac{1}{2\alpha_i(\mathbf{V}_{z_i})} \exp \left[ -\frac{\|\mathbf{z}_i^e - \mathbf{z}_i\|}{\alpha_i(\mathbf{V}_{z_i})} \right] \right\} \quad (4.33c)$$

Muito freqüentemente, o experimentador consegue controlar com bastante eficiência a precisão das medidas experimentais independentes  $\mathbf{x}^e$ . Além disso, técnicas de planejamento experimental (ver Volume II dessa série de publicações) podem ser utilizadas para minimizar o efeito dos erros experimentais das variáveis independentes  $\mathbf{x}^e$  sobre as medidas das variáveis dependentes  $\mathbf{y}^e$ . Por isso, pode ser conveniente tratar as variáveis dependentes e independentes de forma distinta. Admitindo-se que as medições das variáveis independentes não estão correlacionadas com as medições das variáveis dependentes, chega-se a uma nova expressão para a curva de densidade de probabilidades:

$$\wp(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_Z) = \prod_{i=1}^{NE} [\wp_{x_i}(\mathbf{x}_i^e; \mathbf{x}_i, \mathbf{V}_{X_i}) \wp_{y_i}(\mathbf{y}_i^e; \mathbf{y}_i, \mathbf{V}_{Y_i})] \quad (4.34a)$$

$$\wp(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_Z) = \prod_{i=1}^{NE} \left\{ \frac{1}{\sqrt{2\pi \det(\mathbf{V}_{X_i})}} \exp \left[ -\frac{1}{2} (\mathbf{x}_i^e - \mathbf{x}_i)^T \mathbf{V}_{X_i}^{-1} (\mathbf{x}_i^e - \mathbf{x}_i) \right] \cdot \frac{1}{\sqrt{2\pi \det(\mathbf{V}_{Y_i})}} \exp \left[ -\frac{1}{2} (\mathbf{y}_i^e - \mathbf{y}_i)^T \mathbf{V}_{Y_i}^{-1} (\mathbf{y}_i^e - \mathbf{y}_i) \right] \right\} \quad (4.34b)$$

$$\wp(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_Z) = \prod_{i=1}^{NE} \left\{ \frac{1}{2\alpha_{x_i}(\mathbf{V}_{X_i})} \exp \left[ -\frac{\|\mathbf{x}_i^e - \mathbf{x}_i\|}{\alpha_{x_i}(\mathbf{V}_{X_i})} \right] \frac{1}{2\alpha_{y_i}(\mathbf{V}_{Y_i})} \exp \left[ -\frac{\|\mathbf{y}_i^e - \mathbf{y}_i\|}{\alpha_{y_i}(\mathbf{V}_{Y_i})} \right] \right\} \quad (4.34c)$$

Em experimentos realizados sob condições controladas, como geralmente acontece em laboratórios de pesquisa, os valores das variáveis independentes são conhecidos com grande precisão. Nesse caso, parece razoável considerar que  $(\mathbf{x}^e - \mathbf{x}) \approx \mathbf{0}$ . Admitindo-se como válida essa hipótese, é possível reescrever as Equações (4.34a-c) na forma:

$$\wp(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_Z) = \prod_{i=1}^{NE} [\wp_{y_i}(\mathbf{y}_i^e; \mathbf{y}_i, \mathbf{V}_{Y_i})] \quad (4.35a)$$

$$\wp(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_Z) = \prod_{i=1}^{NE} \left\{ \frac{1}{\sqrt{2\pi \det(\mathbf{V}_{Y_i})}} \exp \left[ -\frac{1}{2} (\mathbf{y}_i^e - \mathbf{y}_i)^T \mathbf{V}_{Y_i}^{-1} (\mathbf{y}_i^e - \mathbf{y}_i) \right] \right\} \quad (4.35b)$$

$$\wp(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_Z) = \prod_{i=1}^{NE} \left\{ \frac{1}{2\alpha_{y_i}(\mathbf{V}_{Y_i})} \exp \left[ -\frac{\|\mathbf{y}_i^e - \mathbf{y}_i\|}{\alpha_{y_i}(\mathbf{V}_{Y_i})} \right] \right\} \quad (4.35c)$$

Deve ficar claro que em muitos problemas o controle sobre as variáveis independentes não é tão rígido, de maneira que nem sempre é razoável representar os erros experimentais na forma proposta pelas Equações (4.35a-c). Em experimentos realizados em unidades pilotos e unidades industriais, os desvios experimentais nas variáveis independentes não podem ser geralmente descartados, como discutido nas próximas seções desse capítulo.

Por fim, admitindo-se que todas as medições experimentais podem ser realizadas de forma independente, a curva de densidade de probabilidades pode ser expressa na forma:

$$\wp(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_Z) = \prod_{i=1}^{NE} \left\{ \prod_{j=1}^{NX} \wp_{x_{ij}}(x_{ij}^e; x_{ij}, \sigma_{x_{ij}}^2) \prod_{j=1}^{NY} \wp_{y_{ij}}(y_{ij}^e; y_{ij}, \sigma_{y_{ij}}^2) \right\} \quad (4.36a)$$

$$\phi(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_z) = \prod_{i=1}^{NE} \prod_{j=1}^{NX} \left\{ \frac{1}{\sqrt{2\pi} \sigma_{xij}^2} \exp \left[ -\frac{1}{2} \frac{(x_{ij}^e - x_{ij})^2}{\sigma_{xij}^2} \right] \right\} \prod_{j=1}^{NY} \left\{ \frac{1}{\sqrt{2\pi} \sigma_{yij}^2} \exp \left[ -\frac{1}{2} \frac{(y_{ij}^e - y_{ij})^2}{\sigma_{yij}^2} \right] \right\} \quad (4.36b)$$

$$\phi(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_z) = \prod_{i=1}^{NE} \left\{ \prod_{j=1}^{NX} \frac{1}{2\alpha_{xij}(\sigma_{xij}^2)} \exp \left[ -\frac{\|x_{ij}^e - x_{ij}\|}{\alpha_{xij}(\sigma_{xij}^2)} \right] \prod_{j=1}^{NY} \frac{1}{2\alpha_{yij}(\sigma_{yij}^2)} \exp \left[ -\frac{\|y_{ij}^e - y_{ij}\|}{\alpha_{yij}(\sigma_{yij}^2)} \right] \right\} \quad (4.36c)$$

ou ainda

$$\phi(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_z) = \prod_{i=1}^{NE} \left\{ \prod_{j=1}^{NY} \phi_{yij}(y_{ij}^e; y_{ij}, \sigma_{yij}^2) \right\} \quad (4.36d)$$

$$\phi(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_z) = \prod_{i=1}^{NE} \prod_{j=1}^{NY} \left\{ \frac{1}{\sqrt{2\pi} \sigma_{yij}^2} \exp \left[ -\frac{1}{2} \frac{(y_{ij}^e - y_{ij})^2}{\sigma_{yij}^2} \right] \right\} \quad (4.36e)$$

$$\phi(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_z) = \prod_{i=1}^{NE} \left\{ \prod_{j=1}^{NY} \frac{1}{2\alpha_{yij}(\sigma_{yij}^2)} \exp \left[ -\frac{\|y_{ij}^e - y_{ij}\|}{\alpha_{yij}(\sigma_{yij}^2)} \right] \right\} \quad (4.36f)$$

se as medidas das variáveis independentes não estão sujeitas a erros experimentais. Nas Equações (4.36a-f),  $\sigma_{xij}^2$  e  $\sigma_{yij}^2$  correspondem às variâncias de cada medição  $x_{ij}^e$  e  $y_{ij}^e$ .

É importante observar que as diferentes simplificações introduzidas permitem que a densidade de probabilidades das flutuações experimentais seja reescrita de formas distintas, como mostrado nas Equações (4.32-4.36), a depender da natureza dos erros experimentais. A escolha de uma das muitas formas propostas para a distribuição dos erros experimentais só é possível depois da caracterização apropriada desses erros. Como visto a seguir, essa caracterização dos erros é fundamental para a proposição da técnica de máxima verossimilhança para estimação de parâmetros.

**Pressuposto 2** – Admite-se como válida a hipótese do modelo perfeito.

Considerando que a natureza das flutuações experimentais é conhecida e que existe um modelo eficiente para descrever as relações existentes entre as diferentes variáveis do problema, pode-se admitir que os valores reais e desconhecidos de  $\mathbf{x}$  e  $\mathbf{y}$  ( $\mathbf{z}$ ) nas Equações (4.32-4.36) são aqueles previstos pelo modelo. Em outras palavras, se o modelo é capaz de representar os dados medidos, parece razoável admitir que  $\mathbf{x} = \mathbf{x}^m$  e  $\mathbf{y} = \mathbf{y}^m$  ( $\mathbf{z} = \mathbf{z}^m$ ), onde  $m$  denota que o valor é calculado a partir de um modelo. Fica implícito, assim, que as medidas experimentais flutuam em torno dos valores previstos pelo modelo, já que se admite que o modelo é perfeitamente capaz de descrever as relações existentes entre as diversas variáveis do problema. Essa é certamente uma hipótese bastante forte, já que nenhum modelo consegue de fato capturar todos os detalhes de um problema real. Sob outro ponto de vista, no entanto, parece pouco

producente admitir que um usuário queira utilizar um modelo sabidamente ruim para representar um problema experimental qualquer.

O modelo pode ser definido genericamente na forma:

$$\mathbf{y}^m = \mathbf{f}(\mathbf{x}^m, \boldsymbol{\alpha}) \quad (4.37)$$

onde  $\mathbf{f}$  indica as equações do modelo e  $\boldsymbol{\alpha}$  é o vetor dos parâmetros do modelo. Como já discutido, admite-se que os valores das variáveis dependentes  $\mathbf{y}^m$  podem ser calculados (de forma explícita ou numérica) a partir dos valores das variáveis independentes  $\mathbf{x}^m$  e dos parâmetros  $\boldsymbol{\alpha}$ . Portanto, o vetor  $\mathbf{z}$  que contém as variáveis independentes e dependentes do problema pode ser dado na forma  $\mathbf{z}^{mT} = [\mathbf{x}^{mT} \ \mathbf{y}^{mT}] = [\mathbf{x}^{mT} \ \mathbf{f}^T(\mathbf{x}^m; \boldsymbol{\alpha})]$ . Logo, a distribuição dos erros experimentais pode ser redefinida na forma  $\wp(\mathbf{z}^e; \mathbf{z}^m, \mathbf{V}_z)$ , onde os valores reais desconhecidos da variável  $\mathbf{z}$  foram substituídos pelos valores  $\mathbf{z}^m$  calculados com o auxílio do modelo perfeito. Dessa maneira, a curva de densidade de probabilidades pode ser escrita de forma explicitar o fato de que existe um modelo que descreve como as variáveis dependentes respondem a mudanças das condições experimentais e dos parâmetros. Por exemplo, no problema particular definido pela Equação (4.36e):

$$\wp(\mathbf{z}^e; \mathbf{z}^m, \mathbf{V}_z) = \prod_{i=1}^{NE} \prod_{j=1}^{NY} \left\{ \frac{1}{\sqrt{2\pi} \sigma_{yij}^2} \exp \left[ -\frac{1}{2} \frac{(y_{ij}^e - y_{ij}^m(\mathbf{x}_i, \boldsymbol{\alpha}))^2}{\sigma_{yij}^2} \right] \right\} \quad (4.38)$$

Portanto, a hipótese do modelo perfeito permite introduzir os parâmetros desconhecidos do modelo na função probabilística que descreve as flutuações experimentais.

**Pressuposto 3** – Admite-se como válida a hipótese do experimento bem feito.

Quando os experimentos são conduzidos de forma apropriada, parece razoável admitir que os valores experimentais obtidos representam adequadamente a realidade experimental estudada, a despeito das pequenas e inevitáveis flutuações experimentais. Portanto, parece também razoável admitir que os dados experimentais obtidos são altamente prováveis, uma vez que não são obtidos de forma aleatória, mas como resultado de um processo experimental cuidadoso e reproduzível. Sendo assim, parece também razoável admitir que a repetição das medidas experimentais conduziria a resultados experimentais semelhantes, tendo em vista a pertinência e esmero das observações realizadas. Levada ao extremo, essa argumentação permite considerar como válida a hipótese de que os dados experimentais obtidos não apresentam apenas alta probabilidade de ocorrência – são aqueles que apresentam **máxima probabilidade** de ocorrência. Portanto, parece razoável admitir que os pontos experimentais obtidos devem estar situados na região de máxima probabilidade de  $\wp(\mathbf{z}^e; \mathbf{z}^m, \mathbf{V}_z)$ .

A argumentação proposta deposita enorme confiança nas observações experimentais realizadas. Isso pode ser certamente questionado, já que medidas

experimentais são freqüentemente corrompidas por erros grosseiros e sistemáticos que não se enquadram na argumentação desenvolvida no parágrafo anterior. Sob outro ponto de vista, parece também ser contra-producente acreditar que um experimentador use medidas experimentais sabidamente corrompidas para compreender um problema real. Se o experimento é de fato bem feito, não parece haver razão para acreditar que as observações experimentais sejam pouco prováveis e não representem satisfatoriamente a realidade experimental.

A conseqüência fundamental da hipótese do experimento bem feito é a definição do problema de estimação de parâmetros como o problema de maximização da função  $\wp(\mathbf{z}^e; \mathbf{z}^m, \mathbf{V}_z)$ ; ou seja, de maximização da probabilidade de encontrar os dados experimentais obtidos, que flutuam de forma aleatória ao redor das previsões do modelo, segundo a distribuição conhecida dos erros experimentais. Para que essa tarefa seja possível, manipulam-se os valores desconhecidos  $\mathbf{x}^m$  e  $\boldsymbol{\alpha}$  (quase sempre através de procedimentos numéricos, discutidos no Capítulo 5), que constituem os parâmetros do problema. Esse problema de estimação de parâmetros é conhecido como problema de estimação de parâmetros por máxima verossimilhança.

No caso da distribuição normal definida na Equação (4.38), as variáveis independentes  $\mathbf{x}^e$  não estão sujeitas a erro, de maneira que o vetor  $\mathbf{x}^m = \mathbf{x}^e$  é conhecido. Para maximizar a Equação (4.38), é conveniente perceber que o ponto de máximo da distribuição também coincide com o ponto de máximo do logaritmo dessa função, dado que o  $\ln(\wp)$  é uma função monotônica crescente de  $\wp$ . Por isso, é conveniente escrever

$$\ln[\wp(\mathbf{z}^e; \mathbf{z}^m, \mathbf{V}_z)] = \sum_{i=1}^{NE} \sum_{j=1}^{NY} \ln \left[ \frac{1}{\sqrt{2\pi\sigma_{yij}^2}} \right] - \frac{1}{2} \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{(y_{ij}^e - y_{ij}^m(\mathbf{x}_i, \boldsymbol{\alpha}))^2}{\sigma_{yij}^2} \quad (4.39)$$

Repare que o primeiro termo do somatório do lado direito da equação é constante e não depende do valor dos parâmetros. Logo, procurar o ponto de máximo da função acima é o mesmo que procurar o ponto de máximo da função

$$F = -\frac{1}{2} \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{(y_{ij}^e - y_{ij}^m(\mathbf{x}_i, \boldsymbol{\alpha}))^2}{\sigma_{yij}^2} \quad (4.40)$$

que equivale a procurar o ponto de mínimo da função

$$F_{Obj} = \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{(y_{ij}^e - y_{ij}^m(\mathbf{x}_i, \boldsymbol{\alpha}))^2}{\sigma_{yij}^2} \quad (4.41)$$

que é uma métrica para o problema de estimação de parâmetros. A função acima é usualmente chamada de **função de mínimos quadrados ponderados**. Repare que a Equação (4.41) tem um significado estatístico preciso e profundo, sendo a métrica natural quando os erros experimentais são distribuídos normalmente, não estão correlacionados e quando as variáveis independentes não estão sujeitas a erro, desde que as hipóteses de experimentos bem feitos e modelo perfeito sejam aceitáveis. Observe que o fator de ponderação é o inverso da variância do erro de medida nesse caso; logo,

quanto maiores os erros experimentais, maiores também são os desvios aceitos entre as medidas experimentais e os valores calculados com o modelo. Além disso, a função objetivo da Equação (4.41) permite misturar diferentes conjuntos de dados, desde que os erros de medida sejam conhecidos. É curioso observar que a variância do erro experimental é o fator de normalização natural das variáveis do problema. Observe ainda que a Equação (4.41) converge naturalmente para a Equação (4.12), quando os erros de medição são iguais e constantes em toda a região experimental. Portanto, a função de mínimos quadrados é também uma função de máxima verossimilhança, quando os erros são normalmente distribuídos, não estão relacionados, são constantes e quando as variáveis independentes não estão sujeitas a erro, desde que as hipóteses de experimentos bem feitos e modelo perfeito sejam aceitáveis.

Um dos grandes méritos do método da máxima verossimilhança é permitir a extensão natural da função objetivo para distintas condições de experimentação, de acordo com a estrutura da matriz de covariância. Por exemplo, admitindo agora que os erros são normalmente distribuídos e independentes, mas que as variáveis independentes também estão sujeitas a erro, a função objetivo toma a seguinte forma:

$$\varphi(\mathbf{z}^e; \mathbf{z}, \mathbf{V}_z) = \prod_{i=1}^{NE} \left\{ \prod_{j=1}^{NX} \frac{1}{\sqrt{2\pi} \sigma_{xij}^2} \exp\left(-\frac{(x_{ij}^e - x_{ij})^2}{2\sigma_{xij}^2}\right) \right\} \left\{ \prod_{j=1}^{NY} \frac{1}{\sqrt{2\pi} \sigma_{yij}^2} \exp\left(-\frac{(y_{ij}^e - y_{ij})^2}{2\sigma_{yik}^2}\right) \right\} \quad (4.42)$$

que corresponde à Equação (4.34) após ser introduzida a hipótese de que todas as medições são feitas de forma independente, de maneira que as matrizes de covariâncias dos erros experimentais são diagonais. O ponto de máximo da Equação (4.42) corresponde ao ponto de mínimo da seguinte expressão (similar à Equação (4.41)):

$$F_{Obj} = \sum_{i=1}^{NE} \left[ \sum_{j=1}^{NY} \frac{(y_{ij}^e - y_{ij}^m(\mathbf{x}_i^m; \boldsymbol{\alpha}))^2}{\sigma_{yij}^2} + \sum_{j=1}^{NX} \frac{(x_{ij}^e - x_{ij}^m)^2}{\sigma_{xij}^2} \right] \quad (4.43)$$

que é a métrica natural para o problema com erro também na variável independente, quando as distribuições de erro são normais e os experimentos não estão correlacionados. Observe que na Equação (4.43) só é possível calcular o *valor verdadeiro* de  $\mathbf{x}^m$  se esses valores são incluídos no conjunto de parâmetros a serem estimados pelo problema. Esse é o problema normalmente designado de **reconciliação de dados**, ilustrado no Exemplo 4.23 e apresentado com mais detalhes no Volume III dessa série de publicações.

**Exemplo 4.23** - No problema linear do Exemplo 4.20, admite-se que as variáveis independentes também estão sujeitas a erros de medição. Nesse caso,

$$y^m = \alpha x^m + \beta$$

A função de máxima verossimilhança fica na forma

$$F_{Obj} = \sum_{i=1}^{NE} \left[ \frac{(y_i^e - \alpha x_i^m - \beta)^2}{\sigma_{yi}^2} + \frac{(x_i^e - x_i^m)^2}{\sigma_{xi}^2} \right]$$

Os valores de  $\alpha$  e  $\beta$  e os valores desconhecidos de  $x_i^m$  são então obtidos minimizando-se o valor da função objetivo

$$\frac{\partial F_{Obj}}{\partial \alpha} = \sum_{i=1}^{NE} 2 \frac{(y_i^e - \alpha x_i^m - \beta)}{\sigma_{yi}^2} (-x_i^m) = 0$$

$$\frac{\partial F_{Obj}}{\partial \beta} = \sum_{i=1}^{NE} 2 \frac{(y_i^e - \alpha x_i^m - \beta)}{\sigma_{yi}^2} (-1) = 0$$

$$\frac{\partial F_{Obj}}{\partial x_k^m} = 2 \frac{(y_k^e - \alpha x_k^m - \beta)}{\sigma_{yk}^2} (-\alpha) + 2 \frac{(x_k^e - x_k^m)}{\sigma_{xk}^2} (-1) = 0, \quad k = 1 \dots NE$$

Dessa última equação, é possível concluir que

$$\frac{\frac{\alpha(y_k^e - \beta)}{\sigma_{yk}^2} + \frac{x_k^e}{\sigma_{xk}^2}}{\frac{\alpha^2}{\sigma_{yk}^2} + \frac{1}{\sigma_{xk}^2}} = x_k^m, \quad k = 1 \dots NE$$

que pode ser substituído nas duas equações anteriores, permitindo a solução do problema. No entanto, apenas uma solução numérica é possível, já que não se consegue derivar uma solução analítica para o problema. Dessa forma, o problema de reconciliação de dados, mesmo para o caso mais trivial da reta, requer o uso de rotinas computacionais para resolução adequada do problema. É conveniente observar na expressão acima que, se os erros de medida da variável independente vão a zero,

$$x_k^m = \frac{\frac{\alpha(y_k^e - \beta)}{\sigma_{yk}^2} + \frac{x_k^e}{\sigma_{xk}^2}}{\frac{\alpha^2}{\sigma_{yk}^2} + \frac{1}{\sigma_{xk}^2}} \approx \frac{x_k^e}{\frac{1}{\sigma_{xk}^2}} = x_k^e, \quad k = 1 \dots NE$$

de maneira que o valor calculado coincide com o valor experimental, como admitido anteriormente.

Como discutido na Seção 4.2.2, modelos matemáticos podem ser classificados como lineares ou não-lineares, dependendo do conjunto de variáveis consideradas. Mesmo para um modelo simples, como no caso de uma reta, é importante avaliar a questão da linearidade. Durante a estimação de parâmetros, quando as variáveis

independentes não estão sujeitas a erros, o modelo é linear nos parâmetros e a solução do problema pode ser facilmente obtida. Contudo, durante a reconciliação dos dados, quando se consideram os erros inerentes às medidas das variáveis independentes, mesmo uma reta deixa de ser um modelo linear. Nesse caso, como as variáveis independentes devem ser estimadas em conjunto com os demais parâmetros do modelo, o problema de estimação ganha maior complexidade e a solução do problema passa a requerer ferramentas numéricas mais sofisticadas, como observado no Exemplo 4.23. Para tornar o problema ainda mais envolvente, a maioria absoluta dos modelos fenomenológicos que representam processos reais é constituída por modelos não lineares nos parâmetros. Dessa forma, independentemente da natureza das medidas efetuadas nas variáveis independentes, a solução do problema só pode ser obtida com o auxílio de ferramentas numéricas, como as que serão discutidos no Capítulo 5.

A seguir (Seção 4.6.1) será discutida a aplicação do método da Máxima Verossimilhança para a solução de problemas que usam modelos lineares nos parâmetros e admitem a distribuição normal dos erros experimentais. Nesses casos, algumas soluções analíticas podem ser derivadas para o problema, de maneira que a análise estatística dos resultados pode ser conduzida mais facilmente. Na seção seguinte (Seção 4.6.2) serão considerados alguns modelos não-lineares nos parâmetros, mostrando-se as dificuldades numéricas existentes para a solução do problema de estimação, assim como as aproximações usualmente realizadas para permitir a análise estatística dos resultados finais.

#### 4.6.1 O Método da Máxima Verossimilhança Aplicado a Modelos Lineares

O procedimento clássico de mínimos quadrados usado para estimação de parâmetros de modelos lineares (Equações 4.13-31) é reavaliado agora com o auxílio do procedimento mais geral de máxima verossimilhança, proposto na seção anterior. Suponha que o modelo, contendo  $NY$  respostas, seja definido pelo seguinte conjunto de equações:

$$\begin{aligned}
 y_1^m(\mathbf{x}, \boldsymbol{\alpha}) &= \sum_{p=1}^{NP} \alpha_p f_{1,p}(\mathbf{x}) \\
 y_2^m(\mathbf{x}, \boldsymbol{\alpha}) &= \sum_{p=1}^{NP} \alpha_p f_{2,p}(\mathbf{x}) \\
 &\vdots \\
 y_{NY}^m(\mathbf{x}, \boldsymbol{\alpha}) &= \sum_{p=1}^{NP} \alpha_p f_{NY,p}(\mathbf{x})
 \end{aligned} \tag{4.44}$$

O modelo proposto tem  $NP$  parâmetros,  $\boldsymbol{\alpha}^T = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_{NP}]$ , associados a  $NP \cdot NY$  funções  $f_{i,j}(\mathbf{x})$  que transformam as  $NX$  variáveis independentes,  $\mathbf{x}^T = [x_1 \ x_2 \ \dots \ x_{NX}]$ , nas  $NY$  variáveis dependentes  $\mathbf{y}^T = [y_1 \ y_2 \ \dots \ y_{NY}]$ . Cada uma das variáveis independentes e dependentes é medida em cada uma das  $NE$  condições experimentais. A função de máxima verossimilhança, admitindo-se flutuação normal, medidas independentes de cada variável e ausência de erro significativo nas variáveis independentes, fica então na forma da Equação (4.41). A inclusão das equações do modelo na função objetivo definida na Equação (4.41) leva a:



$$F_{Obj} = \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{\left( y_{ij}^e - \sum_{p=1}^{NP} \alpha_p f_{j,p}(\mathbf{x}_i^e) \right)^2}{\sigma_{ij}^2} \quad (4.45)$$

Os valores de  $\alpha$  são obtidos minimizando-se o valor da função objetivo em relação a cada um de seus componentes, na forma

$$\frac{\partial F_{Obj}}{\partial \alpha_k} = \sum_{i=1}^{NE} \sum_{j=1}^{NY} 2 \frac{\left( y_{ij}^e - \sum_{p=1}^{NP} \alpha_p f_{j,p}(\mathbf{x}_i^e) \right)^2}{\sigma_{ij}^2} \left( -f_{j,k}(\mathbf{x}_i^e) \right) = 0, \quad k = 1 \dots NP \quad (4.46)$$

que resulta no seguinte sistema de equações

$$\sum_{p=1}^{NP} \alpha_p \left[ \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{f_{j,p}(\mathbf{x}_i^e) f_{j,k}(\mathbf{x}_i^e)}{\sigma_{ij}^2} \right] = \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{\left( y_{ij}^e f_{j,k}(\mathbf{x}_i^e) \right)}{\sigma_{ij}^2}, \quad k = 1 \dots NP \quad (4.47)$$

Há, portanto,  $NP$  equações a resolver e  $NP$  incógnitas a determinar, uma vez que os dados experimentais são conhecidos. A solução analítica para esse problema pode ser facilmente derivada se a notação matricial é utilizada. Sejam

$$\mathbf{M} = \begin{bmatrix} \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{f_{j,1}(\mathbf{x}_i^e) f_{j,1}(\mathbf{x}_i^e)}{\sigma_{ij}^2} & \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{f_{j,1}(\mathbf{x}_i^e) f_{j,2}(\mathbf{x}_i^e)}{\sigma_{ij}^2} & \dots & \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{f_{j,1}(\mathbf{x}_i^e) f_{j,NP}(\mathbf{x}_i^e)}{\sigma_{ij}^2} \\ \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{f_{j,2}(\mathbf{x}_i^e) f_{j,1}(\mathbf{x}_i^e)}{\sigma_{ij}^2} & \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{f_{j,2}(\mathbf{x}_i^e) f_{j,2}(\mathbf{x}_i^e)}{\sigma_{ij}^2} & \dots & \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{f_{j,2}(\mathbf{x}_i^e) f_{j,NP}(\mathbf{x}_i^e)}{\sigma_{ij}^2} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{f_{j,NP}(\mathbf{x}_i^e) f_{j,1}(\mathbf{x}_i^e)}{\sigma_{ij}^2} & \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{f_{j,NP}(\mathbf{x}_i^e) f_{j,2}(\mathbf{x}_i^e)}{\sigma_{ij}^2} & \dots & \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{f_{j,NP}(\mathbf{x}_i^e) f_{j,NP}(\mathbf{x}_i^e)}{\sigma_{ij}^2} \end{bmatrix} \quad (4.48)$$

e

$$\mathbf{Y}_f = \begin{bmatrix} \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{\left( y_{ij}^e f_{j,1}(\mathbf{x}_i^e) \right)}{\sigma_{ij}^2} \\ \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{\left( y_{ij}^e f_{j,2}(\mathbf{x}_i^e) \right)}{\sigma_{ij}^2} \\ \vdots \\ \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{\left( y_{ij}^e f_{j,NP}(\mathbf{x}_i^e) \right)}{\sigma_{ij}^2} \end{bmatrix} \quad (4.49)$$

então, reescrevendo a Equação (4.47) com a notação matricial, chega-se a

$$\mathbf{M}\mathbf{a} = \mathbf{Y}_f \Rightarrow \mathbf{a} = \mathbf{M}^{-1}\mathbf{Y}_f \quad (4.50)$$

que é uma solução de enorme importância prática para a teoria de estimação de parâmetros e planejamento de experimentos, muito similar à solução obtida anteriormente com a técnica de mínimos quadrados. Como feito antes, é muito conveniente ainda definir as matrizes

$$\mathbf{G}_Y^T(\mathbf{x}^e) = \begin{bmatrix} f_{1,1}(\mathbf{x}_1^e) & \cdots & f_{1,NP}(\mathbf{x}_1^e) \\ \vdots & \ddots & \vdots \\ f_{NY,1}(\mathbf{x}_1^e) & \cdots & f_{NY,NP}(\mathbf{x}_1^e) \\ f_{1,1}(\mathbf{x}_2^e) & \cdots & f_{1,NP}(\mathbf{x}_2^e) \\ \vdots & \ddots & \vdots \\ f_{NY,1}(\mathbf{x}_2^e) & \cdots & f_{NY,NP}(\mathbf{x}_2^e) \\ \vdots & \vdots & \vdots \\ f_{1,1}(\mathbf{x}_{NE}^e) & \cdots & f_{1,NP}(\mathbf{x}_{NE}^e) \\ \vdots & \ddots & \vdots \\ f_{NY,1}(\mathbf{x}_{NE}^e) & \cdots & f_{NY,NP}(\mathbf{x}_{NE}^e) \end{bmatrix} \quad (4.45)$$

$$\mathbf{V}_y = \begin{bmatrix} \sigma_{1,1}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{1,NY}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{2,1}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{2,NY}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{NE,1}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{NE,NY}^2 \end{bmatrix} \quad (4.46)$$

pois assim é possível escrever a solução diretamente em termos das variáveis medidas na forma

$$\mathbf{a} = \mathbf{M}^{-1}\mathbf{G}_Y^T\mathbf{V}_y^{-1}\mathbf{Y}^e \quad (4.47)$$

que indica que existe uma relação linear direta entre a medida experimental da variável dependente e o valor estimado para o parâmetro (ver Exemplo 4.22).

Como já dito, suponha que as variáveis independentes não contêm erros e que toda a flutuação experimental é devida aos erros de medição das variáveis dependentes  $\mathbf{Y}^e$ . Suponha ainda que dois conjuntos de dados obtidos em condições análogas são comparados entre si. Nesse caso,

$$\Delta\alpha = \alpha_1 - \alpha_2 = \mathbf{M}^{-1}\mathbf{G}_Y^T\mathbf{V}_y^{-1}\mathbf{Y}_1^e - \mathbf{M}^{-1}\mathbf{G}_Y^T\mathbf{V}_y^{-1}\mathbf{Y}_2^e = \mathbf{M}^{-1}\mathbf{G}_Y^T\mathbf{V}_y^{-1}\Delta\mathbf{Y}^e \quad (4.48)$$

Portanto, os valores dos parâmetros fluuam, à medida que fluuam os valores experimentais obtidos. Se a única fonte de flutuação é o erro experimental

$$\Delta\alpha = \mathbf{M}^{-1}\mathbf{G}_Y^T\mathbf{V}_y^{-1}\boldsymbol{\varepsilon} \quad (4.49)$$

que relaciona os erros paramétricos com os erros experimentais. Portanto, se os erros fluuam em torno dos valores verdadeiros com média igual a zero,

$$\mathbb{E}\{\Delta\alpha\} = \mathbb{E}\{\mathbf{M}^{-1}\mathbf{G}_Y^T\mathbf{V}_y^{-1}\boldsymbol{\varepsilon}\} = \mathbf{M}^{-1}\mathbf{G}_Y^T\mathbf{V}_y^{-1}\mathbb{E}\{\boldsymbol{\varepsilon}\} = \mathbf{0} \quad (4.50)$$

os parâmetros também fluuam em torno dos valores verdadeiros, o que mostra que o procedimento de estimação de parâmetros é consistente. Isso é muito bom e confere um certo grau de robustez ao procedimento de estimação de parâmetros proposto. Além disso, a matriz de covariâncias dos parâmetros pode ser calculada como

$$\begin{aligned} \mathbf{V}_\alpha &= \mathbb{E}\{\Delta\alpha\Delta\alpha^T\} = \mathbb{E}\left\{\mathbf{M}^{-1}\mathbf{G}_Y^T\mathbf{V}_y^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T(\mathbf{V}_y^{-1})^T\mathbf{G}_Y(\mathbf{M}^{-1})^T\right\} \\ \mathbf{V}_\alpha &= \mathbf{M}^{-1}\mathbf{G}_Y^T\mathbf{V}_y^{-1}\mathbb{E}\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\}(\mathbf{V}_y^{-1})^T\mathbf{G}_Y(\mathbf{M}^{-1})^T \\ \mathbf{V}_\alpha &= \mathbf{M}^{-1}\mathbf{G}_Y^T\mathbf{V}_y^{-1}\mathbf{V}_y(\mathbf{V}_y^{-1})^T\mathbf{G}_Y(\mathbf{M}^{-1})^T \\ \mathbf{V}_\alpha &= \mathbf{M}^{-1}\mathbf{G}_Y^T(\mathbf{V}_y^{-1})^T\mathbf{G}_Y(\mathbf{M}^{-1})^T \end{aligned} \quad (4.51)$$

Observe que as matrizes  $\mathbf{M}$  e  $\mathbf{V}_y$  são simétricas, de maneira que elas são iguais a suas transpostas. Observe ainda que

$$\mathbf{M} = \mathbf{G}_Y^T\mathbf{V}_y^{-1}\mathbf{G}_Y \quad (4.52)$$

Nesse caso, a matriz de covariâncias dos parâmetros ganha a forma bastante simples

$$\mathbf{V}_\alpha = \mathbf{M}^{-1}\mathbf{G}_Y^T\mathbf{V}_y^{-1}\mathbf{G}_Y(\mathbf{M}^{-1})^T = \mathbf{M}^{-1} \quad (4.53)$$

que é utilizada para a interpretação e solução de um grande número de problemas práticos. Como na Equação (4.28), observe que os erros paramétricos dependem tanto da estrutura do modelo quanto dos dados experimentais disponíveis, já que tanto o modelo quanto os dados experimentais precisam ser definidos para que seja possível calcular a matriz  $\mathbf{M}$ . Portanto, cada modelo e cada conjunto de dados experimentais resulta em um conjunto distinto de incertezas paramétricas, não sendo possível avaliar a estrutura das incertezas paramétricas de forma arbitrária e isolada do contexto de estimação considerado.

Finalmente, os parâmetros são usados para fazer previsões com o modelo em qualquer condição  $\mathbf{x}$ . Usando a notação matricial, o modelo pode ser escrito na forma

$$\mathbf{y}^m(\mathbf{x}, \boldsymbol{\alpha}) = \mathbf{B}(\mathbf{x})\boldsymbol{\alpha} \quad (4.54)$$

onde

$$\mathbf{B}(\mathbf{x}) = \begin{bmatrix} f_{1,1}(\mathbf{x}) & f_{1,1}(\mathbf{x}) & \cdots & f_{1,NP}(\mathbf{x}) \\ f_{2,1}(\mathbf{x}) & f_{2,2}(\mathbf{x}) & \cdots & f_{2,NP}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ f_{NY,1}(\mathbf{x}) & f_{NY,2}(\mathbf{x}) & \cdots & f_{NY,NP}(\mathbf{x}) \end{bmatrix} \quad (4.55)$$

que também estabelece uma relação linear direta entre o valor dos parâmetros e a previsão da variável dependente em um ponto qualquer  $\mathbf{x}$  da região experimental. De forma análoga à realizada anteriormente,

$$\begin{aligned} \hat{\mathbf{V}}_y &= E\{\Delta\mathbf{y}\Delta\mathbf{y}^T\} = E\{\mathbf{B}\Delta\boldsymbol{\alpha}\Delta\boldsymbol{\alpha}^T\mathbf{B}^T\} \\ \hat{\mathbf{V}}_y &= \mathbf{B}E\{\Delta\boldsymbol{\alpha}\Delta\boldsymbol{\alpha}^T\}\mathbf{B}^T = \mathbf{B}\mathbf{V}_\alpha\mathbf{B}^T = \mathbf{B}\mathbf{M}^{-1}\mathbf{B}^T \end{aligned} \quad (4.56)$$

onde  $\hat{\mathbf{V}}_y$  é a matriz de covariância dos erros de predição, que estabelece o vínculo entre os erros de predição, a qualidade dos dados experimentais e o procedimento de estimação de parâmetros. Observe que a Equação (4.56) leva em consideração somente o erro relacionado ao modelo, sendo necessário somar à esta equação a matriz de covariância experimental. Como acima, todas essas expressões são fundamentais para a perfeita compreensão dos procedimentos clássicos de planejamento experimental, como discutido no Volume II dessa série de publicações.

#### 4.6.2 O Método da Máxima Verossimilhança Aplicado a Modelos Não-Lineares

Considerando um problema mais geral, suponha que o modelo não linear contendo  $NY$  respostas, seja definido pelo seguinte conjunto de equações:

$$\begin{aligned} y_1^m(\mathbf{x}, \boldsymbol{\alpha}) &= f_1(\mathbf{x}, \boldsymbol{\alpha}) \\ y_2^m(\mathbf{x}, \boldsymbol{\alpha}) &= f_2(\mathbf{x}, \boldsymbol{\alpha}) \\ &\vdots \\ y_{NY}^m(\mathbf{x}, \boldsymbol{\alpha}) &= f_{NY}(\mathbf{x}, \boldsymbol{\alpha}) \end{aligned} \quad (4.57)$$

Este modelo tem  $NP$  parâmetros,  $\boldsymbol{\alpha}^T = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_{NP}]$ , associados a  $NY$  funções não lineares  $f_j(\mathbf{x})$  que transformam as  $NX$  variáveis independentes,  $\mathbf{x}^T = [x_1 \ x_2 \ \dots \ x_{NX}]$ , nas  $NY$  variáveis dependentes  $\mathbf{y}^T = [y_1 \ y_2 \ \dots \ y_{NY}]$ . Cada uma das variáveis independentes e dependentes é medida em cada uma das  $NE$  condições experimentais. A princípio, qualquer forma da função de máxima verossimilhança pode ser usada para a análise proposta nas seções subseqüentes, seja como uma maximização da função  $\wp(\mathbf{z}^e; \mathbf{z}^m, \mathbf{V}_z)$  ou como uma minimização da função  $F_{Obj}$ .

Admitindo-se flutuação normal, medidas independentes de cada variável e ausência de erro significativo nas variáveis independentes, a função de máxima

verossimilhança fica então na forma da Equação (4.41) (embora pudesse ser qualquer outra). A inclusão das equações do modelo na função objetivo da Equação (4.41) leva a:

$$F_{Obj} = \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{\left( y_{ij}^e - f_j(\mathbf{x}_i^e, \boldsymbol{\alpha}) \right)^2}{\sigma_{ij}^2} \quad (4.58)$$

Quando os erros nas variáveis independentes não podem ser desprezados, a função objetivo deve levar em consideração estes desvios, podendo ser está escrita na forma da Equação (4.43), como mostrado abaixo:

$$F_{Obj} = \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{\left( y_{ij}^e - f_j(\mathbf{x}_i^m, \boldsymbol{\alpha}) \right)^2}{\sigma_{ij}^2} + \sum_{i=1}^{NE} \sum_{k=1}^{NX} \frac{\left( x_{ik}^e - x_{ik}^m \right)^2}{\sigma_{ik}^2} \quad (4.59)$$

Observe que na Equação (4.58) as funções  $f_j$  são funções dos valores experimentais das variáveis independentes  $\mathbf{x}^e$  e dos parâmetros  $\boldsymbol{\alpha}$  que estão sendo estimados. Uma vez medidos, os dados experimentais são fixos e, assim, a função objetivo (Equação 4.58) depende somente dos parâmetros que estão sendo estimados. Como o número de variáveis experimentais que estão sendo previstas pelo modelo é igual a  $NE.NY$  e o número de variáveis que estão sendo calculadas para a minimização da função objetivo é igual a  $NP$  (lembre que as variáveis independentes estão fixas nos valores experimentais), o número de graus de liberdade é definido como:

$$GL = NE.NY - NP \quad (4.60)$$

Já na Equação (4.59) as funções  $f_j$  são funções dos valores estimados das variáveis independentes  $\mathbf{x}^m$  e dos parâmetros  $\boldsymbol{\alpha}$ , que também estão sendo estimados. Neste caso, a função objetivo (Equação 4.59) depende das variáveis independentes e dos parâmetros que estão sendo estimados. Com relação ao número de graus de liberdade, o número de variáveis experimentais que estão sendo previstas pelo modelo é igual a  $NE.NY + NE.NX$  e o número de variáveis que estão sendo calculadas para a minimização da função objetivo é igual a  $NE.NX + NP$ , de forma que o número de graus de liberdade é:

$$\begin{aligned} GL &= NE.NY + NE.NX - NE.NX - NP \\ GL &= NE.NY - NP \end{aligned} \quad (4.61)$$

Assim, o número de graus de liberdade é o mesmo para os dois casos (estimação e reconciliação), apesar do número de variáveis estimadas ser maior no caso da reconciliação.

Considere a função objetivo definida na Equação (4.58). A minimização da função objetivo, nesse caso, deve satisfazer as seguintes equações:

$$\frac{\partial F_{Obj}}{\partial \alpha_p} = 2 \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{\left( y_{ij}^e - f_j(\mathbf{x}_i^e, \boldsymbol{\alpha}) \right)}{\sigma_{ij}^2} \left( - \frac{\partial f_j(\mathbf{x}_i^e, \boldsymbol{\alpha})}{\partial \alpha_p} \right) = 0, \quad p = 1..NP \quad (4.62a)$$

$$\nabla_{\mathbf{a}} F_{Obj} = \begin{bmatrix} \frac{\partial F_{Obj}}{\partial \alpha_1} \\ \frac{\partial F_{Obj}}{\partial \alpha_2} \\ \vdots \\ \frac{\partial F_{Obj}}{\partial \alpha_{NP}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0} \quad (4.62b)$$

ou seja, o **vetor gradiente** da função objetivo em relação aos parâmetros do modelo deve ser nulo.

Quando as variáveis independentes estão também sujeitas a erros, a função objetivo considerada é aquela definida na Equação (4.59). Nesse caso, como as variáveis independentes devem ser estimadas simultaneamente com os demais parâmetros do modelo, o vetor gradiente da função objetivo em relação às variáveis independentes também deve ser nulo, isto é:

$$\frac{\partial F_{Obj}}{\partial x_{i,k}^m} = 2 \sum_{j=1}^{NY} \frac{(y_{i,j}^e - f_j(\mathbf{x}_i^m, \mathbf{a}))}{\sigma_{i,j}^2} \left( -\frac{\partial f_j(\mathbf{x}_i^m, \mathbf{a})}{\partial x_{i,k}^m} \right) - 2 \frac{(x_{i,k}^e - x_{i,k}^m)}{\sigma_{i,k}^2} = 0, \quad i = 1..NE \quad (4.63a)$$

$$k = 1..NX$$

$$\nabla_{\mathbf{x}} F_{Obj} = \begin{bmatrix} \frac{\partial F_{Obj}}{\partial x_{11}} \\ \frac{\partial F_{Obj}}{\partial x_{12}} \\ \vdots \\ \frac{\partial F_{Obj}}{\partial x_{NE,NX}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0} \quad (4.63b)$$

Deve ser observado que o segundo termo do lado direito da Equação (4.59) não depende dos parâmetros  $\mathbf{a}$ , de forma que os vetores gradientes em relação aos parâmetros  $\mathbf{a}$  das funções objetivos definidas nas Equações (4.58) e (4.59) são semelhantes. Apenas deve-se usar os valores de  $\mathbf{x}^m$  ao invés de  $\mathbf{x}^e$  para o cálculo das funções  $f_j$ , no caso em que as variáveis independentes estão sujeitas a erros experimentais que não podem ser desprezados.

**Exemplo 4.24** - Considere o modelo não linear definido pelas equações:

$$y_1 = \alpha_1 x_1^2 + x_2^{\alpha_2}$$

$$y_2 = \alpha_1 \alpha_2 x_1 x_2$$

Utilizando-se a função de mínimos quadrados ponderados para análise do problema, na forma definida pela Equação (4.41) para o caso de estimação e pela

Equação (4.43) para o caso de reconciliação, o gradiente da função objetivo com relação aos parâmetros  $\alpha_1$  e  $\alpha_2$  pode ser definido pelas seguintes equações:

$$\frac{\partial F_{Obj}}{\partial \alpha_1} = 2 \sum_{i=1}^{NE} \left[ \frac{(y_{i,1}^e - \alpha_1 x_{i,1}^2 - x_{i,2}^{\alpha_2})}{\sigma_{i,1}^2} (-x_{i,1}^2) \right] + \left[ \frac{(y_{i,2}^e - \alpha_1 \alpha_2 x_{i,1} x_{i,2})}{\sigma_{i,2}^2} (-\alpha_2 x_{i,1} x_{i,2}) \right] = 0$$

$$\frac{\partial F_{Obj}}{\partial \alpha_2} = 2 \sum_{i=1}^{NE} \left[ \frac{(y_{i,1}^e - \alpha_1 x_{i,1}^2 - x_{i,2}^{\alpha_2})}{\sigma_{i,1}^2} (-x_{i,2}^{\alpha_2} \ln(x_{i,2})) \right] + \left[ \frac{(y_{i,2}^e - \alpha_1 \alpha_2 x_{i,1} x_{i,2})}{\sigma_{i,2}^2} (-\alpha_1 x_{i,1} x_{i,2}) \right] = 0$$

Observe que não é possível obter uma solução analítica para os valores dos parâmetros a partir das duas equações acima, sendo necessária a utilização de um método numérico específico para esse fim. Quando são consideradas as variáveis independentes, as seguintes equações devem ser adicionadas ao problema:

$$\frac{\partial F_{Obj}}{\partial x_{i,1}} = 2 \left\{ \frac{(y_{i,1}^e - \alpha_1 x_{i,1}^2 - x_{i,2}^{\alpha_2})}{\sigma_{i,1}^2} (-2\alpha_1 x_{i,1}) + \frac{(y_{i,2}^e - \alpha_1 \alpha_2 x_{i,1} x_{i,2})}{\sigma_{i,2}^2} (-\alpha_1 \alpha_2 x_{i,2}) - \frac{(x_{i,1}^e - x_{i,1})}{\sigma_{i,1}^2} \right\} = 0$$

$$\frac{\partial F_{Obj}}{\partial x_{i,2}} = 2 \left\{ \frac{(y_{i,1}^e - \alpha_1 x_{i,1}^2 - x_{i,2}^{\alpha_2})}{\sigma_{i,1}^2} (-\alpha_1 x_{i,2}^{\alpha_2-1}) + \frac{(y_{i,2}^e - \alpha_1 \alpha_2 x_{i,1} x_{i,2})}{\sigma_{i,2}^2} (-\alpha_1 \alpha_2 x_{i,1}) - \frac{(x_{i,2}^e - x_{i,2})}{\sigma_{i,2}^2} \right\} = 0$$

onde  $i = 1..NE$ .

Quando somente a estimação dos parâmetros  $\alpha$  é considerada, a solução do problema é equivalente à solução de um sistema de duas equações algébricas com duas incógnitas. Quando o problema de reconciliação é considerado, o sistema de equações algébricas passa a ser constituído por  $2.NE + 2$  equações e  $2.NE + 2$  incógnitas ( $2.NE$  variáveis independentes e 2 parâmetros). Assim, a dimensão do problema numérico que precisa ser resolvido aumenta consideravelmente no problema de reconciliação. A depender do problema de estimação proposto, a dimensão do sistema de equações que deve ser resolvido pode ser bastante elevada.

Além da maior dificuldade computacional associada à obtenção de uma solução para o problema não linear de estimação de parâmetros, uma outra questão relevante que se coloca diz respeito à obtenção das matrizes de covariância dos parâmetros e dos erros de predição; ou seja, a como caracterizar estatisticamente a qualidade da solução encontrada. No caso de modelos não lineares, não é possível derivar uma solução analítica similar à Equação (4.47), usada para interpretar as incertezas paramétricas de modelos lineares.

Para que se compreendam as soluções apresentadas a seguir, é necessário lembrar que a matriz de covariâncias pode ser definida na forma:

$$\mathbf{V} = E\{\Delta\mathbf{x}\Delta\mathbf{x}^T\} = E\left\{\begin{bmatrix} \Delta x_1\Delta x_1 & \Delta x_1\Delta x_2 & \cdots & \Delta x_1\Delta x_N \\ \Delta x_2\Delta x_1 & \Delta x_2\Delta x_2 & \cdots & \Delta x_2\Delta x_N \\ \vdots & \vdots & \ddots & \vdots \\ \Delta x_N\Delta x_1 & \Delta x_N\Delta x_2 & \cdots & \Delta x_N\Delta x_N \end{bmatrix}\right\} \quad (4.64)$$

Considerando que os desvios experimentais e paramétricos não são muito grandes (o que pode **não** ser uma hipótese muito boa em alguns problemas!), a matriz de covariância pode ser aproximada com auxílio da seguinte forma diferencial da equação (4.64):

$$\mathbf{V} = E\{\Delta\mathbf{x}\Delta\mathbf{x}^T\} \cong E\{d\mathbf{x}d\mathbf{x}^T\} \quad (4.65)$$

Considerando que o modelo é perfeito, como já discutido nas seções anteriores, o único motivo de não serem observados experimentalmente os valores preditos pelo modelo é o erro experimental. Assim, dependendo da existência ou não de desvios experimentais significativos nas variáveis independentes, podem-se escrever as seguintes equações:

$$\mathbf{y}^e = \mathbf{y}^m(\mathbf{x}^e, \boldsymbol{\alpha}) + \boldsymbol{\varepsilon}_y = \mathbf{f}(\mathbf{x}^e, \boldsymbol{\alpha}) + \boldsymbol{\varepsilon}_y \quad (4.66a)$$

$$\mathbf{y}^e = \mathbf{y}^m(\mathbf{x}^m + \boldsymbol{\varepsilon}_x, \boldsymbol{\alpha}) + \boldsymbol{\varepsilon}_y = \mathbf{f}(\mathbf{x}^m + \boldsymbol{\varepsilon}_x, \boldsymbol{\alpha}) + \boldsymbol{\varepsilon}_y \quad (4.66b)$$

onde  $\boldsymbol{\varepsilon}$  representa os desvios observados entre os valores experimentais e os valores preditos pelo modelo, confundidos com os próprios erros experimentais (uma vez que o modelo é perfeito). Considerando que as variáveis independentes não estão sujeitas a erros, a aproximação linear das equações do modelo, válida quando os desvios experimentais e paramétricos são pequenos, leva a:

$$\mathbf{y}^e = \mathbf{f}(\mathbf{x}^e, \boldsymbol{\alpha}_0) + \frac{\partial \mathbf{f}(\mathbf{x}^e, \boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} \Delta \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_y \quad (4.67a)$$

sendo possível escrever:

$$\Delta \mathbf{y} = \frac{\partial \mathbf{f}(\mathbf{x}^e, \boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} \Delta \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_y \quad (4.68a)$$

De forma similar, se as variáveis independentes estão sujeitas a erros experimentais

$$\mathbf{y}^e = \mathbf{f}(\mathbf{x}^m, \boldsymbol{\alpha}_0) + \frac{\partial \mathbf{f}(\mathbf{x}^m, \boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} \Delta \boldsymbol{\alpha} + \frac{\partial \mathbf{f}(\mathbf{x}^m, \boldsymbol{\alpha}_0)}{\partial \mathbf{x}} \boldsymbol{\varepsilon}_x + \boldsymbol{\varepsilon}_y \quad (4.67b)$$

sendo possível escrever:



$$\Delta \mathbf{y} = \frac{\partial \mathbf{f}(\mathbf{x}^m, \mathbf{a}_0)}{\partial \mathbf{a}} \Delta \mathbf{a} + \frac{\partial \mathbf{f}(\mathbf{x}^m, \mathbf{a}_0)}{\partial \mathbf{x}} \boldsymbol{\varepsilon}_x + \boldsymbol{\varepsilon}_y \quad (4.68b)$$

A Equação (4.68) mostra, de forma aproximada, como os desvios entre os dados experimentais e os preditos pelo modelo variam em função dos desvios paramétricos e dos erros experimentais.

Por outro lado, o vetor gradiente da função objetivo em relação aos parâmetros do modelo é nulo quando a função objetivo está no ponto mínimo, como imposto pelo procedimento de minimização executado durante a estimação de parâmetros. Logo,

$$\nabla_{\mathbf{a}} F_{Obj}(\mathbf{y}^e, \mathbf{x}^e, \mathbf{a}) = \mathbf{0} \quad (4.69)$$

Considerando que as variáveis independentes não estão sujeitas a desvios experimentais significativos, a aproximação linear do vetor gradiente apresentado na Equação (4.69) pode ser escrita na forma:

$$\begin{aligned} \nabla_{\mathbf{a}} F_{Obj}(\mathbf{y}^e + \boldsymbol{\varepsilon}_y, \mathbf{x}^e, \mathbf{a} + \Delta \mathbf{a}) &\cong \nabla_{\mathbf{a}} F_{Obj}(\mathbf{y}^e, \mathbf{x}^e, \mathbf{a}) + \frac{\partial [\nabla_{\mathbf{a}} F_{Obj}(\mathbf{y}^e, \mathbf{x}^e, \mathbf{a})]}{\partial \mathbf{y}^e} \boldsymbol{\varepsilon}_y + \\ &+ \frac{\partial [\nabla_{\mathbf{a}} F_{Obj}(\mathbf{y}^e, \mathbf{x}^e, \mathbf{a})]}{\partial \mathbf{a}} \Delta \mathbf{a} = \mathbf{0} \end{aligned} \quad (4.70a)$$

que descreve como as incertezas experimentais e paramétricas afetam o cálculo do vetor gradiente da função objetivo nas proximidades dos valores estimados para os parâmetros do modelo. De forma similar, se as variáveis independentes estão sujeitas a erros:

$$\begin{aligned} \nabla_{\mathbf{a}} F_{Obj}(\mathbf{y}^e + \boldsymbol{\varepsilon}_y, \mathbf{x}^e + \boldsymbol{\varepsilon}_x, \mathbf{a} + \Delta \mathbf{a}) &\cong \nabla_{\mathbf{a}} F_{Obj}(\mathbf{y}^e, \mathbf{x}^e, \mathbf{a}) + \frac{\partial [\nabla_{\mathbf{a}} F_{Obj}(\mathbf{y}^e, \mathbf{x}^e, \mathbf{a})]}{\partial \mathbf{y}^e} \boldsymbol{\varepsilon}_y + \\ &\frac{\partial [\nabla_{\mathbf{a}} F_{Obj}(\mathbf{y}^e, \mathbf{x}^e, \mathbf{a})]}{\partial \mathbf{x}^e} \boldsymbol{\varepsilon}_x + \frac{\partial [\nabla_{\mathbf{a}} F_{Obj}(\mathbf{y}^e, \mathbf{x}^e, \mathbf{a})]}{\partial \mathbf{a}} \Delta \mathbf{a} = \mathbf{0} \end{aligned} \quad (4.70b)$$

O zero do lado direito da Equação (4.70a-b) é imposto pelo procedimento de estimação de parâmetros; ou seja, independentemente de como os dados experimentais mudem, os novos parâmetros estimados ( $\mathbf{a} + \Delta \mathbf{a}$ ) sempre fazem com que o vetor gradiente da função objetivo seja igual a zero. A Equação (4.70a-b) indica, portanto, como as incertezas experimentais ( $\boldsymbol{\varepsilon}_x, \boldsymbol{\varepsilon}_y$ ) provocam mudanças nos valores dos parâmetros ( $\Delta \mathbf{a}$ ). O primeiro termo do lado direito da Equação (4.70a-b) é o próprio vetor gradiente da função objetivo em relação aos parâmetros no ponto considerado. Por definição, esse gradiente também é nulo, já que os parâmetros são sempre estimados através da minimização da função objetivo.

A Equação (4.70a-b) sugere a definição das seguintes matrizes:

$$\mathbf{G}_Y = \frac{\partial \left[ \nabla_{\mathbf{a}} F_{Obj}(\mathbf{y}^e, \mathbf{x}^e, \mathbf{a}) \right]}{\partial \mathbf{y}^e} = \begin{bmatrix} \frac{\partial^2 F_{Obj}}{\partial \alpha_1 \partial y_{1,1}^e} & \frac{\partial^2 F_{Obj}}{\partial \alpha_1 \partial y_{1,2}^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial \alpha_1 \partial y_{NE,NY}^e} \\ \frac{\partial^2 F_{Obj}}{\partial \alpha_2 \partial y_{1,1}^e} & \frac{\partial^2 F_{Obj}}{\partial \alpha_2 \partial y_{1,2}^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial \alpha_2 \partial y_{NE,NY}^e} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 F_{Obj}}{\partial \alpha_{NP} \partial y_{1,1}^e} & \frac{\partial^2 F_{Obj}}{\partial \alpha_{NP} \partial y_{1,2}^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial \alpha_{NP} \partial y_{NE,NY}^e} \end{bmatrix} \quad (4.71a)$$

$$\mathbf{G}_X = \frac{\partial \left[ \nabla_{\mathbf{a}} F_{Obj}(\mathbf{y}^e, \mathbf{x}^e, \mathbf{a}) \right]}{\partial \mathbf{x}^e} = \begin{bmatrix} \frac{\partial^2 F_{Obj}}{\partial \alpha_1 \partial x_{1,1}^e} & \frac{\partial^2 F_{Obj}}{\partial \alpha_1 \partial x_{1,2}^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial \alpha_1 \partial x_{NE,NX}^e} \\ \frac{\partial^2 F_{Obj}}{\partial \alpha_2 \partial x_{1,1}^e} & \frac{\partial^2 F_{Obj}}{\partial \alpha_2 \partial x_{1,2}^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial \alpha_2 \partial x_{NE,NX}^e} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 F_{Obj}}{\partial \alpha_{NP} \partial x_{1,1}^e} & \frac{\partial^2 F_{Obj}}{\partial \alpha_{NP} \partial x_{1,2}^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial \alpha_{NP} \partial x_{NE,NX}^e} \end{bmatrix} \quad (4.71b)$$

$$\mathbf{H}_a = \frac{\partial \left[ \nabla_{\mathbf{a}} F_{Obj}(\mathbf{y}^e, \mathbf{x}^e, \mathbf{a}) \right]}{\partial \mathbf{a}} = \begin{bmatrix} \frac{\partial^2 F_{Obj}}{\partial \alpha_1 \partial \alpha_1} & \frac{\partial^2 F_{Obj}}{\partial \alpha_1 \partial \alpha_2} & \dots & \frac{\partial^2 F_{Obj}}{\partial \alpha_1 \partial \alpha_{NP}} \\ \frac{\partial^2 F_{Obj}}{\partial \alpha_2 \partial \alpha_1} & \frac{\partial^2 F_{Obj}}{\partial \alpha_2 \partial \alpha_2} & \dots & \frac{\partial^2 F_{Obj}}{\partial \alpha_2 \partial \alpha_{NP}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 F_{Obj}}{\partial \alpha_{NP} \partial \alpha_1} & \frac{\partial^2 F_{Obj}}{\partial \alpha_{NP} \partial \alpha_2} & \dots & \frac{\partial^2 F_{Obj}}{\partial \alpha_{NP} \partial \alpha_{NP}} \end{bmatrix} \quad (4.71c)$$

de maneira que a Equação (4.70) pode ser escrita na forma

$$\mathbf{G}_Y \boldsymbol{\varepsilon}_Y + \mathbf{H}_a \Delta \mathbf{a} = \mathbf{0} \quad (4.72a)$$

$$\mathbf{G}_Y \boldsymbol{\varepsilon}_Y + \mathbf{G}_X \boldsymbol{\varepsilon}_X + \mathbf{H}_a \Delta \mathbf{a} = \mathbf{0} \quad (4.72b)$$

a depender da existência ou não de erros de medida nas variáveis independentes. Observe que a matriz  $\mathbf{H}_a$ , chamada de **matriz Hessiana** ou **matriz de derivadas segundas**, é simétrica ( $\mathbf{H}_a = \mathbf{H}_a^T$ ). Assim, a partir da Equação (4.72) é possível obter uma expressão que mostra como os desvios experimentais se propagam até os parâmetros, na forma:

$$\Delta \mathbf{a} = -\mathbf{H}_a^{-1} \mathbf{G}_Y \boldsymbol{\varepsilon}_Y \quad (4.73a)$$

$$\Delta \mathbf{a} = -\mathbf{H}_a^{-1} [\mathbf{G}_Y \boldsymbol{\varepsilon}_Y + \mathbf{G}_X \boldsymbol{\varepsilon}_X] \quad (4.73b)$$

Inserindo a Equação (4.73) na Equação (4.65), chega-se finalmente a:

$$\begin{aligned} \mathbf{V}_a &= E\{\Delta\boldsymbol{\alpha} \Delta\boldsymbol{\alpha}^T\} = E\left\{\left(-\mathbf{H}_a^{-1} \mathbf{G}_Y \boldsymbol{\varepsilon}_Y\right)\left(-\mathbf{H}_a^{-1} \mathbf{G}_Y \boldsymbol{\varepsilon}_Y\right)^T\right\} \\ \mathbf{V}_a &= E\left\{\mathbf{H}_a^{-1} \mathbf{G}_Y \boldsymbol{\varepsilon}_Y \boldsymbol{\varepsilon}_Y^T \mathbf{G}_Y^T \mathbf{H}_a^{-1}\right\} = \mathbf{H}_a^{-1} \mathbf{G}_Y E\left\{\boldsymbol{\varepsilon}_Y \boldsymbol{\varepsilon}_Y^T\right\} \mathbf{G}_Y^T \mathbf{H}_a^{-1} \quad (4.74a) \\ \mathbf{V}_a &= \mathbf{H}_a^{-1} \mathbf{G}_Y \mathbf{V}_Y \mathbf{G}_Y^T \mathbf{H}_a^{-1} \end{aligned}$$

ou, de forma análoga

$$\begin{aligned} \mathbf{V}_a &= E\{\Delta\boldsymbol{\alpha} \Delta\boldsymbol{\alpha}^T\} = E\left\{\left(-\mathbf{H}_a^{-1} [\mathbf{G}_Y \boldsymbol{\varepsilon}_Y + \mathbf{G}_X \boldsymbol{\varepsilon}_X]\right)\left(-\mathbf{H}_a^{-1} [\mathbf{G}_Y \boldsymbol{\varepsilon}_Y + \mathbf{G}_X \boldsymbol{\varepsilon}_X]\right)^T\right\} \\ \mathbf{V}_a &= \mathbf{H}_a^{-1} \left[\mathbf{G}_Y \mathbf{V}_Y \mathbf{G}_Y^T + 2 \mathbf{G}_X \mathbf{V}_{XY} \mathbf{G}_Y^T + \mathbf{G}_X \mathbf{V}_X \mathbf{G}_X^T\right] \mathbf{H}_a^{-1} \quad (4.74b) \\ \mathbf{V}_a &= \mathbf{H}_a^{-1} \left[\mathbf{G}_Z \mathbf{V}_Z \mathbf{G}_Z^T\right] \mathbf{H}_a^{-1} \end{aligned}$$

onde  $\mathbf{V}_X$ ,  $\mathbf{V}_Y$  e  $\mathbf{V}_{XY}$  são as matrizes de covariâncias dos desvios experimentais nas variáveis independentes, dependentes e das covariâncias entre elas. Se as medidas experimentais das variáveis independentes e dependentes não estão correlacionadas entre si, a Equação (4.74b) ganha a forma particular

$$\mathbf{V}_a = \mathbf{H}_a^{-1} \left[\mathbf{G}_Y \mathbf{V}_Y \mathbf{G}_Y^T + \mathbf{G}_X \mathbf{V}_X \mathbf{G}_X^T\right] \mathbf{H}_a^{-1} \quad (4.74c)$$

As Equações (4.74a-c) mostram de forma explícita como a incerteza experimental se transforma em incerteza nos parâmetros, através da função objetivo e do modelo utilizados (cujas derivadas estão em  $\mathbf{H}_a$ ,  $\mathbf{G}_Y$  e  $\mathbf{G}_X$ ), durante o procedimento de estimação de parâmetros. A validade das Equações (4.70-74) pressupõe que os erros experimentais e paramétricos são pequenos.

**Exemplo 4.25** - No problema linear do Exemplo 4.23, admite-se que as variáveis independentes estão sujeitas a erros de medição e que as medidas são obtidas de forma independente. Nesse caso,

$$y^m = \alpha x^m + \beta$$

A função de máxima verossimilhança fica na forma

$$F_{Obj} = \sum_{i=1}^{NE} \left[ \frac{(y_i^e - \alpha x_i^m - \beta)^2}{\sigma_{yi}^2} + \frac{(x_i^e - x_i^m)^2}{\sigma_{xi}^2} \right]$$

O vetor gradiente da função objetivo fica então na forma

$$\nabla F_{Obj} = \begin{bmatrix} \frac{\partial F_{Obj}}{\partial \alpha} \\ \frac{\partial F_{Obj}}{\partial \beta} \\ \frac{\partial F_{Obj}}{\partial x_1^m} \\ \vdots \\ \frac{\partial F_{Obj}}{\partial x_{NE}^m} \end{bmatrix} = \begin{bmatrix} 2 \sum_{i=1}^{NE} \left[ \frac{(y_i^e - \alpha x_i^m - \beta)(-x_i^m)}{\sigma_{yi}^2} \right] \\ 2 \sum_{i=1}^{NE} \left[ \frac{(y_i^e - \alpha x_i^m - \beta)(-1)}{\sigma_{yi}^2} \right] \\ 2 \frac{(y_1^e - \alpha x_1^m - \beta)(-\alpha)}{\sigma_{y1}^2} + 2 \frac{(x_1^e - x_1^m)(-1)}{\sigma_{x1}^2} \\ \vdots \\ 2 \frac{(y_{NE}^e - \alpha x_{NE}^m - \beta)(-\alpha)}{\sigma_{yNE}^2} + 2 \frac{(x_{NE}^e - x_{NE}^m)(-1)}{\sigma_{xNE}^2} \end{bmatrix}$$

As matrizes  $\mathbf{H}_\alpha$ ,  $\mathbf{G}_X$  e  $\mathbf{G}_Y$  podem então ser calculadas como

$$\mathbf{H}_\alpha = \begin{bmatrix} \frac{\partial^2 F_{Obj}}{\partial \alpha^2} & \frac{\partial^2 F_{Obj}}{\partial \alpha \partial \beta} & \frac{\partial^2 F_{Obj}}{\partial \alpha \partial x_1^m} & \dots & \frac{\partial^2 F_{Obj}}{\partial \alpha \partial x_{NE}^m} \\ \frac{\partial^2 F_{Obj}}{\partial \alpha \partial \beta} & \frac{\partial^2 F_{Obj}}{\partial \beta^2} & \frac{\partial^2 F_{Obj}}{\partial \beta \partial x_1^m} & \dots & \frac{\partial^2 F_{Obj}}{\partial \beta \partial x_{NE}^m} \\ \frac{\partial^2 F_{Obj}}{\partial \alpha \partial x_1^m} & \frac{\partial^2 F_{Obj}}{\partial \beta \partial x_1^m} & \frac{\partial^2 F_{Obj}}{\partial (x_1^m)^2} & \dots & \frac{\partial^2 F_{Obj}}{\partial x_1^m \partial x_{NE}^m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 F_{Obj}}{\partial \alpha \partial x_{NE}^m} & \frac{\partial^2 F_{Obj}}{\partial \beta \partial x_{NE}^m} & \frac{\partial^2 F_{Obj}}{\partial x_1^m \partial x_{NE}^m} & \dots & \frac{\partial^2 F_{Obj}}{\partial (x_{NE}^m)^2} \end{bmatrix}$$

$$\mathbf{H}_\alpha = \begin{bmatrix} 2 \sum_{i=1}^{NE} \left[ \frac{(x_i^m)^2}{\sigma_{yi}^2} \right] & 2 \sum_{i=1}^{NE} \left[ \frac{(x_i^m)}{\sigma_{yi}^2} \right] & 2 \frac{-(y_1^e - \alpha x_1^m - \beta) + \alpha x_1^m}{\sigma_{y1}^2} & \dots & 2 \frac{-(y_{NE}^e - \alpha x_{NE}^m - \beta) + \alpha x_{NE}^m}{\sigma_{yNE}^2} \\ 2 \sum_{i=1}^{NE} \left[ \frac{(x_i^m)}{\sigma_{yi}^2} \right] & 2 \sum_{i=1}^{NE} \left[ \frac{1}{\sigma_{yi}^2} \right] & 2 \frac{\alpha}{\sigma_{y1}^2} & \dots & 2 \frac{\alpha}{\sigma_{yNE}^2} \\ 2 \frac{-(y_1^e - \alpha x_1^m - \beta) + \alpha x_1^m}{\sigma_{y1}^2} & 2 \frac{\alpha}{\sigma_{y1}^2} & 2 \frac{\alpha^2}{\sigma_{y1}^2} + 2 \frac{1}{\sigma_{x1}^2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 2 \frac{-(y_{NE}^e - \alpha x_{NE}^m - \beta) + \alpha x_{NE}^m}{\sigma_{yNE}^2} & 2 \frac{\alpha}{\sigma_{yNE}^2} & 0 & \dots & 2 \frac{\alpha^2}{\sigma_{yNE}^2} + 2 \frac{1}{\sigma_{xNE}^2} \end{bmatrix}$$

$$\mathbf{G}_x = \left[ \frac{\partial^2 F_{Obj}}{\partial \alpha_i \partial x_j^e} \right] = \begin{bmatrix} \frac{\partial^2 F_{Obj}}{\partial \alpha \partial x_1^e} & \frac{\partial^2 F_{Obj}}{\partial \alpha \partial x_2^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial \alpha \partial x_{NE}^e} \\ \frac{\partial^2 F_{Obj}}{\partial \beta \partial x_1^e} & \frac{\partial^2 F_{Obj}}{\partial \beta \partial x_2^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial \beta \partial x_{NE}^e} \\ \frac{\partial^2 F_{Obj}}{\partial x_1^m \partial x_1^e} & \frac{\partial^2 F_{Obj}}{\partial x_1^m \partial x_2^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial x_1^m \partial x_{NE}^e} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 F_{Obj}}{\partial x_{NE}^m \partial x_1^e} & \frac{\partial^2 F_{Obj}}{\partial x_{NE}^m \partial x_2^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial x_{NE}^m \partial x_{NE}^e} \end{bmatrix}$$

$$\mathbf{G}_x = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ -\frac{2}{\sigma_{x1}^2} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -\frac{2}{\sigma_{xNE}^2} \end{bmatrix}$$

$$\mathbf{G}_y = \left[ \frac{\partial^2 F_{Obj}}{\partial \alpha_i \partial y_j^e} \right] = \begin{bmatrix} \frac{\partial^2 F_{Obj}}{\partial \alpha \partial y_1^e} & \frac{\partial^2 F_{Obj}}{\partial \alpha \partial y_2^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial \alpha \partial y_{NE}^e} \\ \frac{\partial^2 F_{Obj}}{\partial \beta \partial y_1^e} & \frac{\partial^2 F_{Obj}}{\partial \beta \partial y_2^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial \beta \partial y_{NE}^e} \\ \frac{\partial^2 F_{Obj}}{\partial x_1^m \partial y_1^e} & \frac{\partial^2 F_{Obj}}{\partial x_1^m \partial y_2^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial x_1^m \partial y_{NE}^e} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 F_{Obj}}{\partial x_{NE}^m \partial y_1^e} & \frac{\partial^2 F_{Obj}}{\partial x_{NE}^m \partial y_2^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial x_{NE}^m \partial y_{NE}^e} \end{bmatrix}$$

$$\mathbf{G}_y = \begin{bmatrix} 2 \frac{(-x_1^m)}{\sigma_{y1}^2} & 2 \frac{(-x_2^m)}{\sigma_{y2}^2} & \dots & 2 \frac{(-x_{NE}^m)}{\sigma_{yNE}^2} \\ 2 \frac{(-1)}{\sigma_{y1}^2} & 2 \frac{(-1)}{\sigma_{y2}^2} & \dots & 2 \frac{(-1)}{\sigma_{yNE}^2} \\ 2 \frac{(-\alpha)}{\sigma_{y1}^2} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 2 \frac{(-\alpha)}{\sigma_{yNE}^2} \end{bmatrix}$$

Considerando que não há erro nas variáveis independentes  $\mathbf{x}$ , a função objetivo fica na forma:

$$F_{Obj} = \sum_{i=1}^{NE} \left[ \frac{(y_i^e - \alpha x_i^e - \beta)^2}{\sigma_{yi}^2} \right]$$

O vetor gradiente da função objetivo fica então na forma

$$\nabla F_{Obj} = \begin{bmatrix} \frac{\partial F_{Obj}}{\partial \alpha} \\ \frac{\partial F_{Obj}}{\partial \beta} \end{bmatrix} = \begin{bmatrix} 2 \sum_{i=1}^{NE} \left[ \frac{(y_i^e - \alpha x_i^e - \beta)(-x_i^e)}{\sigma_{yi}^2} \right] \\ 2 \sum_{i=1}^{NE} \left[ \frac{(y_i^e - \alpha x_i^e - \beta)(-1)}{\sigma_{yi}^2} \right] \end{bmatrix}$$

de maneira que as matrizes  $\mathbf{H}_\alpha$  e  $\mathbf{G}_Y$  podem ser calculadas como

$$\mathbf{H}_\alpha = \begin{bmatrix} \frac{\partial^2 F_{Obj}}{\partial \alpha_i \partial \alpha_j} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 F_{Obj}}{\partial \alpha^2} & \frac{\partial^2 F_{Obj}}{\partial \alpha \partial \beta} \\ \frac{\partial^2 F_{Obj}}{\partial \alpha \partial \beta} & \frac{\partial^2 F_{Obj}}{\partial \beta^2} \end{bmatrix}$$

$$\mathbf{H}_\alpha = \begin{bmatrix} 2 \sum_{i=1}^{NE} \left[ \frac{(x_i^e)^2}{\sigma_{yi}^2} \right] & 2 \sum_{i=1}^{NE} \left[ \frac{(x_i^e)}{\sigma_{yi}^2} \right] \\ 2 \sum_{i=1}^{NE} \left[ \frac{(x_i^e)}{\sigma_{yi}^2} \right] & 2 \sum_{i=1}^{NE} \left[ \frac{1}{\sigma_{yi}^2} \right] \end{bmatrix}$$

$$\mathbf{G}_Y = \begin{bmatrix} \frac{\partial^2 F_{Obj}}{\partial \alpha_i \partial y_j^e} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 F_{Obj}}{\partial \alpha \partial y_1^e} & \frac{\partial^2 F_{Obj}}{\partial \alpha \partial y_2^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial \alpha \partial y_{NE}^e} \\ \frac{\partial^2 F_{Obj}}{\partial \beta \partial y_1^e} & \frac{\partial^2 F_{Obj}}{\partial \beta \partial y_2^e} & \dots & \frac{\partial^2 F_{Obj}}{\partial \beta \partial y_{NE}^e} \end{bmatrix}$$

$$\mathbf{G}_Y = \begin{bmatrix} 2 \frac{(-x_1^e)}{\sigma_{y1}^2} & 2 \frac{(-x_2^e)}{\sigma_{y2}^2} & \dots & 2 \frac{(-x_{NE}^e)}{\sigma_{yNE}^2} \\ 2 \frac{(-1)}{\sigma_{y1}^2} & 2 \frac{(-1)}{\sigma_{y2}^2} & \dots & 2 \frac{(-1)}{\sigma_{yNE}^2} \end{bmatrix}$$

Nesse caso, de acordo com a Equação (4.74), é possível escrever

$$\mathbf{G}_Y \mathbf{V}_Y \mathbf{G}_Y^T = \begin{bmatrix} 4 \sum_{i=1}^{NE} \frac{x_i^2}{\sigma_i^2} & 4 \sum_{i=1}^{NE} \frac{x_i}{\sigma_i^2} \\ 4 \sum_{i=1}^{NE} \frac{x_i}{\sigma_i^2} & 4 \sum_{i=1}^{NE} \frac{1}{\sigma_i^2} \end{bmatrix} = 2 \mathbf{H}_\alpha$$

$$\mathbf{V}_\alpha = \mathbf{H}_\alpha^{-1} \mathbf{G}_Y \mathbf{V}_Y \mathbf{G}_Y^T \mathbf{H}_\alpha^{-1} = 2 \mathbf{H}_\alpha^{-1} = \frac{1}{\left[ \left( \sum_{i=1}^{NE} \frac{x_i^2}{\sigma_i^2} \right) \left( \sum_{i=1}^{NE} \frac{1}{\sigma_i^2} \right) - \left( \sum_{i=1}^{NE} \frac{x_i}{\sigma_i} \right)^2 \right]} \begin{bmatrix} \sum_{i=1}^{NE} \frac{1}{\sigma_i^2} & - \sum_{i=1}^{NE} \frac{x_i}{\sigma_i^2} \\ - \sum_{i=1}^{NE} \frac{x_i}{\sigma_i^2} & \sum_{i=1}^{NE} \frac{x_i^2}{\sigma_i^2} \end{bmatrix}$$

No caso particular em que os erros experimentais são constantes, a expressão pode ser reduzida a

$$\mathbf{V}_\alpha = \frac{1}{\left[ \left( \sum_{i=1}^{NE} x_i^2 \right) (NE) - \left( \sum_{i=1}^{NE} x_i \right)^2 \right]} \begin{bmatrix} NE & - \sum_{i=1}^{NE} x_i \\ - \sum_{i=1}^{NE} x_i & \sum_{i=1}^{NE} x_i^2 \end{bmatrix}$$

já apresentada anteriormente no Exemplo 4.20 e obtida através de outro tipo de argumentação. Isso mostra claramente a consistência da discussão e das equações apresentadas.

Uma aproximação muito usada para a estimação de parâmetros de modelos não lineares é a aproximação de Gauss. Quando a função objetivo é definida na seguinte forma:

$$F_{Obj} = (\mathbf{y}^e - \mathbf{y}^m)^T \mathbf{V}_y^{-1} (\mathbf{y}^e - \mathbf{y}^m) \quad (4.75)$$

os elementos  $[h_{r,s}]$  da matriz Hessiana (definida na Equação (4.72)) podem ser escritos como:

$$h_{r,s} = \frac{\partial^2 F_{Obj}}{\partial \alpha_r \partial \alpha_s} = 2 \left( \frac{\partial \mathbf{y}^m}{\partial \alpha_r} \right)^T \mathbf{V}_y^{-1} \left( \frac{\partial \mathbf{y}^m}{\partial \alpha_s} \right) - 2 \left( \frac{\partial^2 \mathbf{y}^m}{\partial \alpha_r \partial \alpha_s} \right)^T \mathbf{V}_y^{-1} (\mathbf{y}^e - \mathbf{y}^m) \quad (4.76)$$

Admitindo-se que a diferença entre os valores experimentais e os calculados pelo modelo são pequenos e flutuam aleatoriamente ao redor do valor zero, conforme as hipóteses do modelo perfeito e do experimento bem feito, o segundo termo do lado direito da Equação (4.76) pode ser desprezado. Nesse caso, a matriz Hessiana pode ser aproximada por:

$$\mathbf{H}_a = \frac{\partial^2 F_{Obj}}{\partial \mathbf{a} \partial \mathbf{a}} \cong 2 \mathbf{B}^T \mathbf{V}_y^{-1} \mathbf{B} \quad (4.77)$$

onde  $\mathbf{B}$  é a **matriz de sensitividades**, definida como:

$$\mathbf{B} = \left[ \frac{\partial \mathbf{y}^m}{\partial \mathbf{a}} \right] = \begin{bmatrix} \frac{\partial y_1^m}{\partial \alpha_1} & \frac{\partial y_1^m}{\partial \alpha_2} & \dots & \frac{\partial y_1^m}{\partial \alpha_{NP}} \\ \frac{\partial y_2^m}{\partial \alpha_1} & \frac{\partial y_2^m}{\partial \alpha_2} & \dots & \frac{\partial y_2^m}{\partial \alpha_{NP}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{NY}^m}{\partial \alpha_1} & \frac{\partial y_{NY}^m}{\partial \alpha_2} & \dots & \frac{\partial y_{NY}^m}{\partial \alpha_{NP}} \end{bmatrix} \quad (4.78)$$

De acordo com a Equação (4.71),  $\mathbf{G}_Y$  pode ser calculado como:

$$\mathbf{G}_Y = \frac{\partial^2 F_{Obj}}{\partial \mathbf{a} \partial \mathbf{y}_e} = 2 \mathbf{B}^T \mathbf{V}_y^{-1} \quad (4.79)$$

de maneira que o termo  $\mathbf{G}_Y \mathbf{V}_y \mathbf{G}_Y^T$  que aparece na Equação (4.74) fica reduzido a:

$$\mathbf{G}_Y \mathbf{V}_y \mathbf{G}_Y^T = 4 \mathbf{B}^T \mathbf{V}_y^{-1} \mathbf{V}_y \mathbf{V}_y^{-1} \mathbf{B} = 4 \mathbf{B}^T \mathbf{V}_y^{-1} \mathbf{B} = 2 \mathbf{H}_a \quad (4.80)$$

Assim, a matriz de covariâncias dos parâmetros pode ser reescrita como

$$\mathbf{V}_a = \mathbf{H}_a^{-1} (\mathbf{G}_Y \mathbf{V}_y \mathbf{G}_Y^T) \mathbf{H}_a^{-1} = \mathbf{H}_a^{-1} (2 \mathbf{H}_a) \mathbf{H}_a^{-1} = 2 \mathbf{H}_a^{-1} \quad (4.81)$$

A Equação (4.81) é freqüentemente usada para interpretar as incertezas paramétricas, embora só seja rigorosamente válida quando a função objetivo tem a forma da Equação (4.75). Usando finalmente a Equação (4.77), que define a matriz Hessiana após a aproximação de Gauss, a matriz de covariâncias dos parâmetros fica na forma

$$\mathbf{V}_a = [\mathbf{B}^T \mathbf{V}_y^{-1} \mathbf{B}]^{-1} \quad (4.82)$$

A Equação (4.82) é usada com muita freqüência para a interpretação de incertezas paramétricas, mas deve ficar claro que ela representa um conjunto muito restrito de condições. Por outro lado, esta forma da matriz de covariâncias dos parâmetros é muito útil, já que para seu cálculo são necessárias somente as derivadas primeiras das respostas do modelo com relação aos parâmetros. Além disso, deve ser observado que, para modelos lineares nos parâmetros, o termo desprezado da Equação (4.76) é nulo, de modo que a Equação (4.82) se torna exata.

**Exemplo 4.26** – Considerando o modelo definido como:

$$y^m = e^{ax}$$



e considerando que não há erros nas variáveis independentes  $\mathbf{x}$ , a função de máxima verossimilhança pode ser definida na forma

$$F_{Obj} = \sum_{i=1}^{NE} \frac{(y_i^e - e^{\alpha x_i^e})^2}{\sigma_{yi}^2}$$

O gradiente da função objetivo fica então na forma

$$\frac{\partial F_{Obj}}{\partial \alpha} = \sum_{i=1}^{NE} \frac{2}{\sigma_{yi}^2} \left[ (y_i^e - e^{\alpha x_i^e}) (-x_i^e) e^{\alpha x_i^e} \right]$$

A matriz Hessiana  $\mathbf{H}_\alpha$  (que neste caso específico tem dimensão 1x1) pode então ser calculada como

$$\mathbf{H}_\alpha = \sum_{i=1}^{NE} \frac{2}{\sigma_{yi}^2} \left[ (x_i^e e^{\alpha x_i^e})^2 + (y_i^e - e^{\alpha x_i^e}) (x_i^e)^2 e^{\alpha x_i^e} \right]$$

A matriz de covariâncias dos parâmetros é definida na Equação (4.81):

$$\mathbf{V}_\alpha = 2\mathbf{H}_\alpha^{-1} = \frac{1}{\sum_{i=1}^{NE} \frac{1}{\sigma_{yi}^2} \left[ (x_i^e e^{\alpha x_i^e})^2 + (y_i^e - e^{\alpha x_i^e}) (x_i^e)^2 e^{\alpha x_i^e} \right]}$$

Utilizando agora a aproximação de Gauss, a matriz de covariâncias dos parâmetros pode ser representada como mostrado na Equação (4.82):

$$\mathbf{V}_\alpha = \left[ \mathbf{B}^T \mathbf{V}_y^{-1} \mathbf{B} \right]^{-1} = \left[ \begin{array}{cccc} \frac{\partial y_1}{\partial \alpha} & \frac{\partial y_2}{\partial \alpha} & \dots & \frac{\partial y_{NE}}{\partial \alpha} \end{array} \right] \left[ \begin{array}{cccc} \sigma_{y1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{y2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{yNE}^2 \end{array} \right]^{-1} \left[ \begin{array}{c} \frac{\partial y_1}{\partial \alpha} \\ \frac{\partial y_2}{\partial \alpha} \\ \vdots \\ \frac{\partial y_{NE}}{\partial \alpha} \end{array} \right]^{-1}$$

$$\mathbf{V}_\alpha = \left[ \begin{array}{cccc} x_1 e^{(\alpha x_1)} & x_2 e^{(\alpha x_2)} & \dots & x_{NE} e^{(\alpha x_{NE})} \end{array} \right] \left[ \begin{array}{cccc} \sigma_{y1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{y2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{yNE}^2 \end{array} \right]^{-1} \left[ \begin{array}{c} x_1 e^{(\alpha x_1)} \\ x_2 e^{(\alpha x_2)} \\ \vdots \\ x_{NE} e^{(\alpha x_{NE})} \end{array} \right]^{-1}$$

$$\mathbf{V}_\alpha = \left[ \sum_{i=1}^{NE} \frac{(x_i e^{(\alpha x_i)})^2}{\sigma_{yi}^2} \right]^{-1} = \frac{1}{\sum_{i=1}^{NE} \frac{1}{\sigma_{yi}^2} (x_i e^{(\alpha x_i)})^2}$$

Pode ser observado que a diferença entre as matrizes de covariâncias das incertezas paramétricas obtidas com as Equações (4.81) e (4.82) é justamente o termo que contém a soma dos resíduos  $(y_i^e - e^{\alpha x_i^e})$ . Segundo a aproximação de Gauss, baseada nas hipóteses do experimento bem feito e do modelo perfeito, esse termo tende a zero e pode, assim, ser desprezado.

É interessante observar como os desvios paramétricos estão correlacionados com os desvios experimentais, na forma:

$$\begin{aligned} \text{Covar}(\Delta\boldsymbol{\alpha}, \boldsymbol{\varepsilon}_Y) &= E\{\Delta\boldsymbol{\alpha}\boldsymbol{\varepsilon}_Y^T\} = E\left\{\left(-\mathbf{H}_a^{-1}[\mathbf{G}_Y\boldsymbol{\varepsilon}_Y + \mathbf{G}_X\boldsymbol{\varepsilon}_X]\right)(\boldsymbol{\varepsilon}_Y^T)\right\} \\ \text{Covar}(\Delta\boldsymbol{\alpha}, \boldsymbol{\varepsilon}_Y) &= E\left\{-\mathbf{H}_a^{-1}\left[\mathbf{G}_Y\boldsymbol{\varepsilon}_Y\boldsymbol{\varepsilon}_Y^T + \mathbf{G}_X\boldsymbol{\varepsilon}_X\boldsymbol{\varepsilon}_Y^T\right]\right\} \\ \text{Covar}(\Delta\boldsymbol{\alpha}, \boldsymbol{\varepsilon}_Y) &= -\mathbf{H}_a^{-1}\left[\mathbf{G}_YE\{\boldsymbol{\varepsilon}_Y\boldsymbol{\varepsilon}_Y^T\} + \mathbf{G}_XE\{\boldsymbol{\varepsilon}_X\boldsymbol{\varepsilon}_Y^T\}\right] \\ \text{Covar}(\Delta\boldsymbol{\alpha}, \boldsymbol{\varepsilon}_Y) &= -\mathbf{H}_a^{-1}\mathbf{G}_Y\mathbf{V}_y - \mathbf{H}_a^{-1}\mathbf{G}_X\mathbf{V}_{yx} \end{aligned} \quad (4.83a)$$

$$\begin{aligned} \text{Covar}(\Delta\boldsymbol{\alpha}, \boldsymbol{\varepsilon}_X) &= E\{\Delta\boldsymbol{\alpha}\boldsymbol{\varepsilon}_X^T\} = E\left\{\left(-\mathbf{H}_a^{-1}[\mathbf{G}_Y\boldsymbol{\varepsilon}_Y + \mathbf{G}_X\boldsymbol{\varepsilon}_X]\right)(\boldsymbol{\varepsilon}_X^T)\right\} \\ \text{Covar}(\Delta\boldsymbol{\alpha}, \boldsymbol{\varepsilon}_X) &= E\left\{-\mathbf{H}_a^{-1}\left[\mathbf{G}_Y\boldsymbol{\varepsilon}_Y\boldsymbol{\varepsilon}_X^T + \mathbf{G}_X\boldsymbol{\varepsilon}_X\boldsymbol{\varepsilon}_X^T\right]\right\} \\ \text{Covar}(\Delta\boldsymbol{\alpha}, \boldsymbol{\varepsilon}_X) &= -\mathbf{H}_a^{-1}\left[\mathbf{G}_YE\{\boldsymbol{\varepsilon}_Y\boldsymbol{\varepsilon}_X^T\} + \mathbf{G}_XE\{\boldsymbol{\varepsilon}_X\boldsymbol{\varepsilon}_X^T\}\right] \\ \text{Covar}(\Delta\boldsymbol{\alpha}, \boldsymbol{\varepsilon}_X) &= -\mathbf{H}_a^{-1}\mathbf{G}_Y\mathbf{V}_{yx} - \mathbf{H}_a^{-1}\mathbf{G}_X\mathbf{V}_x \end{aligned} \quad (4.83b)$$

Quando os erros nas variáveis independentes são pequenos e podem ser desprezados, a Equação (4.83a) fica:

$$\begin{aligned} \text{Covar}(\Delta\boldsymbol{\alpha}, \boldsymbol{\varepsilon}_Y) &= E\{\Delta\boldsymbol{\alpha}\boldsymbol{\varepsilon}_Y^T\} = E\left\{\left(-\mathbf{H}_a^{-1}\mathbf{G}_Y\boldsymbol{\varepsilon}_Y\right)(\boldsymbol{\varepsilon}_Y^T)\right\} \\ \text{Covar}(\Delta\boldsymbol{\alpha}, \boldsymbol{\varepsilon}_Y) &= E\left\{-\mathbf{H}_a^{-1}\mathbf{G}_Y\boldsymbol{\varepsilon}_Y\boldsymbol{\varepsilon}_Y^T\right\} \\ \text{Covar}(\Delta\boldsymbol{\alpha}, \boldsymbol{\varepsilon}_Y) &= -\mathbf{H}_a^{-1}\mathbf{G}_YE\{\boldsymbol{\varepsilon}_Y\boldsymbol{\varepsilon}_Y^T\} \\ \text{Covar}(\Delta\boldsymbol{\alpha}, \boldsymbol{\varepsilon}_Y) &= -\mathbf{H}_a^{-1}\mathbf{G}_Y\mathbf{V}_y \end{aligned} \quad (4.83c)$$

De forma semelhante, é possível mostrar que:

$$\text{Covar}(\boldsymbol{\varepsilon}_Y, \Delta\boldsymbol{\alpha}) = \text{Covar}(\Delta\boldsymbol{\alpha}, \boldsymbol{\varepsilon}_Y)^T = \left(-\mathbf{H}_a^{-1}\mathbf{G}_Y\mathbf{V}_y\right)^T = -\mathbf{V}_y\mathbf{G}_Y^T\mathbf{H}_a^{-1} \quad (4.84)$$

Estes resultados, mostram que os erros paramétricos e os erros experimentais não são independentes. Essa noção de dependência é fundamental para correta compreensão dos procedimentos de planejamento experimental e interpretação de resultados. As distribuições de erros paramétricos dependem de forma complexa das distribuições dos erros experimentais, de maneira que diferentes incertezas paramétricas são obtidas para cada conjunto característico de dados experimentais.

**Exemplo 4.27** – Na segunda parte do Exemplo 4.25 foram calculadas as matrizes  $\mathbf{H}_\alpha$  e  $\mathbf{G}_Y$ , admitindo-se que as variáveis independentes não estão sujeitas a erros de medição. Nesse caso, a covariância entre os erros paramétricos e os erros experimentais pode ser calculada como:

$$\text{Covar}(\Delta\alpha, \varepsilon_Y) = -\mathbf{H}_\alpha^{-1} \mathbf{G}_Y \mathbf{V}_y$$

As matrizes  $\mathbf{H}_\alpha$  e  $\mathbf{G}_Y$  foram calculadas como

$$\mathbf{H}_\alpha = \begin{bmatrix} 2 \sum_{i=1}^{NE} \left[ \frac{(x_i^e)^2}{\sigma_{yi}^2} \right] & 2 \sum_{i=1}^{NE} \left[ \frac{(x_i^e)}{\sigma_{yi}^2} \right] \\ 2 \sum_{i=1}^{NE} \left[ \frac{(x_i^e)}{\sigma_{yi}^2} \right] & 2 \sum_{i=1}^{NE} \left[ \frac{1}{\sigma_{yi}^2} \right] \end{bmatrix}$$

$$\mathbf{G}_Y = \begin{bmatrix} 2 \frac{(-x_1^e)}{\sigma_{y1}^2} & 2 \frac{(-x_2^e)}{\sigma_{y2}^2} & \dots & 2 \frac{(-x_{NE}^e)}{\sigma_{yNE}^2} \\ 2 \frac{(-1)}{\sigma_{y1}^2} & 2 \frac{(-1)}{\sigma_{y2}^2} & \dots & 2 \frac{(-1)}{\sigma_{yNE}^2} \end{bmatrix}$$

de maneira que a inversa da matriz Hessiana tem a forma:

$$\mathbf{H}_\alpha^{-1} = \frac{1}{2 \left[ \left( \sum_{i=1}^{NE} \frac{x_i^2}{\sigma_i^2} \right) \left( \sum_{i=1}^{NE} \frac{1}{\sigma_i^2} \right) - \left( \sum_{i=1}^{NE} \frac{x_i}{\sigma_i^2} \right)^2 \right]} \begin{bmatrix} \sum_{i=1}^{NE} \frac{1}{\sigma_i^2} & - \sum_{i=1}^{NE} \frac{x_i}{\sigma_i^2} \\ - \sum_{i=1}^{NE} \frac{x_i}{\sigma_i^2} & \sum_{i=1}^{NE} \frac{x_i^2}{\sigma_i^2} \end{bmatrix}$$

A matriz das covariâncias experimentais foi definida como:

$$\mathbf{V}_y = \begin{bmatrix} \sigma_{y1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{y2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{yNE}^2 \end{bmatrix}$$

Fazendo inicialmente o produto  $\mathbf{G}_Y \mathbf{V}_Y$ :

$$\mathbf{G}_y \mathbf{V}_y = \begin{bmatrix} 2 \frac{(-x_1^e)}{\sigma_{y1}^2} & 2 \frac{(-x_2^e)}{\sigma_{y2}^2} & \dots & 2 \frac{(-x_{NE}^e)}{\sigma_{yNE}^2} \end{bmatrix} \begin{bmatrix} \sigma_{y1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{y2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{yNE}^2 \end{bmatrix}$$

$$\mathbf{G}_y \mathbf{V}_y = \begin{bmatrix} 2(-x_1^e) & 2(-x_2^e) & \dots & 2(-x_{NE}^e) \\ 2(-1) & 2(-1) & \dots & 2(-1) \end{bmatrix}$$

$$\text{Covar}(\Delta \alpha, \boldsymbol{\varepsilon}_y) = \frac{1}{\left[ \left( \sum_{i=1}^{NE} \frac{x_i^2}{\sigma_i^2} \right) \left( \sum_{i=1}^{NE} \frac{1}{\sigma_i^2} \right) - \left( \sum_{i=1}^{NE} \frac{x_i}{\sigma_i} \right)^2 \right]} \begin{bmatrix} x_1^e \sum_{i=1}^{NE} \frac{1}{\sigma_i^2} - \sum_{i=1}^{NE} \frac{x_i}{\sigma_i^2} & \dots & x_{NE}^e \sum_{i=1}^{NE} \frac{1}{\sigma_i^2} - \sum_{i=1}^{NE} \frac{x_i}{\sigma_i^2} \\ -x_1^e \sum_{i=1}^{NE} \frac{x_i}{\sigma_i^2} + \sum_{i=1}^{NE} \frac{x_i^2}{\sigma_i^2} & \dots & -x_{NE}^e \sum_{i=1}^{NE} \frac{x_i}{\sigma_i^2} + \sum_{i=1}^{NE} \frac{x_i^2}{\sigma_i^2} \end{bmatrix}$$

Resta ainda definir a matriz de covariâncias das incertezas da predição do modelo. Substituindo a matriz de sensibilidades na Equação (4.68), a matriz de covariâncias das incertezas da predição pode ser definida como:

$$\hat{\mathbf{V}}_y = E \{ \Delta y \Delta y^T \} = E \{ (\mathbf{B} \Delta \alpha + \boldsymbol{\varepsilon}_y) (\mathbf{B} \Delta \alpha + \boldsymbol{\varepsilon}_y)^T \}$$

$$\hat{\mathbf{V}}_y = E \{ \mathbf{B} \Delta \alpha \Delta \alpha^T \mathbf{B}^T + \mathbf{B} \Delta \alpha \boldsymbol{\varepsilon}_y^T + \boldsymbol{\varepsilon}_y \mathbf{B}^T \Delta \alpha^T + \boldsymbol{\varepsilon}_y \boldsymbol{\varepsilon}_y^T \} \quad (4.85)$$

$$\hat{\mathbf{V}}_y = \mathbf{B} \mathbf{V}_\alpha \mathbf{B}^T + \mathbf{B} \text{Covar}(\Delta \alpha, \boldsymbol{\varepsilon}_y) + \text{Covar}(\Delta \alpha, \boldsymbol{\varepsilon}_y) \mathbf{B}^T + \mathbf{V}_y$$

(Deve ser observado que, no caso da estimação simultânea das variáveis independentes  $\mathbf{x}^m$ , como no problema de reconciliação de dados, esses valores devem ser interpretados como os demais parâmetros estimados do problema e incluídos no vetor  $\boldsymbol{\alpha}$ .)

Repare que a Equação (4.85) gera dois cenários: interpretação das incertezas de predição dos pontos experimentais usados para a estimação de parâmetros e a interpretação de incertezas de predição associadas a um novo ponto experimental. No primeiro caso, a matriz de covariâncias das incertezas da predição é a Equação (4.85). Já no segundo caso, os desvios  $\boldsymbol{\varepsilon}_y$  correspondem a desvios de condições que não foram usadas para a estimação dos parâmetros. Nesse caso, a correlação entre  $\Delta \alpha$  e  $\boldsymbol{\varepsilon}_y$  é nula, de forma que a matriz de covariâncias dos erros de predição fica na forma mais simples:

$$\hat{\mathbf{V}}_y = \mathbf{B} \mathbf{V}_\alpha \mathbf{B}^T + \mathbf{V}_y \quad (4.86)$$

A Equação (4.86) mostra que o erro de predição tem dois componentes. O primeiro é um componente relativo aos erros cometidos no passado (na forma dos erros dos parâmetros  $\mathbf{V}_\alpha$ , lembrando que a matriz de sensibilidades na Equação (4.82) deve ser calculada nas condições experimentais onde os experimentos já foram realizados). O segundo é um componente relativo aos erros futuros (na forma do erro de experimentação  $\mathbf{V}_y$  e na matriz de sensibilidade  $\mathbf{B}$  calculada, agora, nos pontos onde se

deseja calcular o erro de predição, onde ainda não foram necessariamente feitas observações experimentais). Essa equação mostra que nunca conseguimos nos livrar dos erros cometidos no passado, já que eles se propagam indefinidamente para o futuro na matriz  $V_{\alpha}$ . Portanto, há que se ter sempre muito cuidado com os erros cometidos no laboratório para estimar parâmetros, pois eles se propagarão indefinidamente para o futuro, sempre que o modelo for usado para simulação, projeto, otimização, etc. Esse é um aspecto operacional que não tem sido explorado suficientemente no ambiente da pesquisa.

As expressões desenvolvidas acima para a definição das incertezas paramétricas e de predição são fundamentais para a correta interpretação estatística dos resultados, como será discutido na próxima seção. Além disso, essas expressões são de fundamental importância para a perfeita compreensão dos procedimentos clássicos de planejamento experimental, como discutido no Volume II dessa série de publicações.

## 4.7. Interpretação Estatística dos Dados Estimados

Admite-se freqüentemente que o procedimento de estimação de parâmetros termina após a minimização da função objetivo. Como discutido nas seções anteriores, essa presunção está equivocada. Terminado o procedimento de minimização da função objetivo e de obtenção dos parâmetros do modelo, é imperioso analisar a qualidade dos resultados obtidos com as ferramentas estatísticas apropriadas. Algumas informações relevantes devem ser obtidas em relação à qualidade do modelo e dos parâmetros estimados, para que seja possível reavaliar a necessidade de modificar o modelo ou de continuar o processo de experimentação e coleta dos dados experimentais.

### 4.7.1. A Qualidade do Ajuste Obtido

Uma das hipóteses fundamentais usadas para formulação do procedimento de estimação de parâmetros é a hipótese do modelo perfeito. A adequação dessa hipótese pode ser checada através da comparação dos erros de modelagem obtidos com os desvios experimentais admissíveis. Se esses erros forem compatíveis, não há como rejeitar a hipótese do modelo perfeito; caso contrário, há argumento suficiente para o usuário continuar investindo no aperfeiçoamento da modelagem do sistema.

No caso da estimação de mínimos quadrados, as Equações (4.11) e (4.12) mostram que a variância final dos desvios de predição é dada por

$$\hat{\sigma}_y^2 = \frac{F_{Obj}}{NE - NP} \quad (4.87)$$

onde  $NE - NP$  é o número de graus de liberdade  $\nu$ . Portanto, a variância de predição do modelo ( $\hat{\sigma}_y^2$ ) pode ser comparada aos erros experimentais ( $\sigma_y^2$ ) com o teste F, descrito na Seção 3.3. Se as duas variâncias são semelhantes estatisticamente, o modelo deve ser considerado satisfatório e não há motivos aparentes para descartar o modelo (nem a hipótese do modelo perfeito). Caso contrário, dois cenários são possíveis:

- a)  $\hat{\sigma}_y^2 > \sigma_y^2$ : o modelo não é capaz de explicar os erros experimentais a contento, pois os erros de predição são significativamente maiores que os erros experimentais. Logo, esforços devem ser feitos para aperfeiçoar o modelo. Não deve também ser descartada a possibilidade dos erros experimentais estarem subestimados. Nesse caso, é também conveniente que o experimentador reavalie a precisão das medidas feitas e das informações usadas para fins de estimação de parâmetros.
  
- b)  $\hat{\sigma}_y^2 < \sigma_y^2$ : o modelo reproduz os dados experimentais muito melhor do que esperado. Esse cenário, ao contrário do que muitos acreditam, TAMBÉM indica que há algo errado, pois um modelo não pode levar a previsões melhores do que os dados usados para gerá-lo. Portanto, é bastante provável que o modelo esteja super parametrizado, indicando que o modelo talvez possa ser simplificado. (O uso de modelos super parametrizados pode resultar em risco muito grande para o usuário, como ilustrado na Figura 4.14 mostrada abaixo.) Não deve também ser descartada a possibilidade dos erros experimentais estarem superestimados. Como comentado anteriormente, é também conveniente que o experimentador reavalie a precisão das medidas feitas e das informações usadas para fins de estimação de parâmetros.

A Figura 4.14 apresentada abaixo mostra dois resultados advindos da estimação de parâmetros. Os pontos e respectivas barras verticais representam os valores medidos de uma certa variável ( $y$ ) e as respectivas incertezas, à medida que uma outra variável ( $x$ ) varia. A reta horizontal representa o resultado obtido com um modelo constante (a hipótese de que  $y$  não depende de  $x$ ), perfeitamente compatível com os erros experimentais. A curva que passa pelos pontos é resultante de um modelo cúbico de interpolação. O leitor atento deve reparar que o modelo cúbico **NÃO** deve ser considerado melhor do que o modelo constante, porque não consegue explicar os erros experimentais. Na verdade, as extrapolações feitas com o modelo cúbico podem ser muito piores do que as extrapolações feitas com o modelo constante, a despeito de o modelo cúbico passar por cima dos pontos. Diz-se que o modelo cúbico está super parametrizado e que está ajustando o ruído ou o erro experimental.

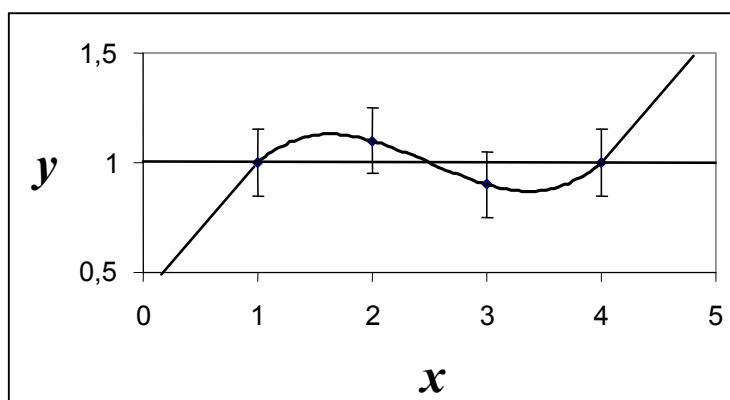


Figura 4.14 – Esquema ilustrativo dos problemas da super parametrização.

No caso das funções de máxima verossimilhança desenvolvidas a partir da hipótese de normalidade dos erros experimentais, as funções objetivos resultantes, como nas Equações (4.41) e (4.43), têm interpretação de  $\chi^2$  (ver Seção 3.2.2) com  $\nu = NE.NY-$

$NP$  graus de liberdade. Nesse caso, se o modelo consegue representar a contento os dados experimentais,

$$\chi_{min}^2 < F_{Obj} < \chi_{max}^2 \quad (4.88)$$

Caso contrário, como no parágrafo anterior, dois cenários são possíveis:

- a)  $F_{Obj} > \chi_{max}^2$  : o modelo não é capaz de explicar os erros experimentais a contento, pois os erros de predição são significativamente maiores que os erros experimentais. Logo, esforços devem ser feitos para aperfeiçoar o modelo. Não deve também ser descartada a possibilidade dos erros experimentais estarem subestimados. Nesse caso, é também conveniente que o experimentador reavalie a precisão das medidas feitas e das informações usadas para fins de estimação de parâmetros.
- b)  $F_{Obj} < \chi_{min}^2$  : o modelo reproduz os dados experimentais muito melhor do que esperado. Esse cenário, como já explicado, indica que há algo errado, pois um modelo não pode levar a previsões melhores do que os dados usados para gerá-lo. Portanto, é bastante provável que o modelo esteja super parametrizado, indicando que o modelo talvez possa ser simplificado. Não deve também ser descartada a possibilidade dos erros experimentais estarem superestimados. Como já comentado, é conveniente que o experimentador reavalie a precisão das medidas feitas e das informações usadas para fins de estimação de parâmetros.

Em ambos os casos discutidos, o leitor deve perceber a necessidade de avaliar os erros experimentais de forma independente, para que seja possível estabelecer uma opinião sobre a qualidade do ajuste. Se os erros experimentais não são conhecidos, é comum admitir que  $\hat{\sigma}_y^2 = \sigma_y^2$ , o que carece de qualquer rigor estatístico formal, embora muitos pacotes de simulação admitam como válida essa hipótese. O usuário deve estar atento para o uso dessa hipótese simplificadora e descartar o seu uso, sempre que possível. Deve-se inclusive ressaltar que esta definição do erro experimental não pode ser feita para modelos com múltiplas respostas. Por isso, o experimentador deve estar sempre preocupado com a caracterização precisa dos erros experimentais característicos do sistema em estudo, ao invés de deixar questão tão importante como essa sem uma avaliação formal.

Costuma-se ainda definir um coeficiente de correlação entre os valores experimentais e calculados na forma

$$\rho^m = \frac{\sum_{i=1}^{NE} (y_i^e - \bar{y}_i^e)(y_i^m - \bar{y}_i^m)}{\sqrt{\left[ \sum_{i=1}^{NE} (y_i^e - \bar{y}_i^e)^2 \right] \left[ \sum_{i=1}^{NE} (y_i^m - \bar{y}_i^m)^2 \right]}} \quad (4.89)$$

que indica quão proximamente os dados calculados acompanham os dados experimentais. Usualmente, se o coeficiente de correlação é superior a 0.9, o modelo é considerado satisfatório, indicando que os valores preditos pelo modelo variam de forma aproximadamente linear e proporcional com as medidas experimentais. Contudo,

é importante que o usuário perceba que valores inferiores a 0.9 podem indicar tanto desajuste do modelo (recomendando aperfeiçoamento da estrutura matemática usada para descrever os dados experimentais), quanto a existência de erros experimentais excessivos (recomendando o aperfeiçoamento das técnicas experimentais). (Recomenda-se que o leitor consulte a Seção 1.6 para observar que o coeficiente de correlação não pode ser tomado como uma medida absoluta da qualidade do ajuste do modelo.) A identificação do foco do problema pode ser feita com auxílio da função objetivo. Por exemplo, se a função objetivo recomenda o uso do modelo ( $\chi^2_{min} < F_{Obj} < \chi^2_{max}$ ) e o coeficiente de correlação é baixo ( $\rho^m < 0.9$ ), o problema central parece ser o excesso de erro de experimentação. Por outro lado, se a função objetivo não recomenda o uso do modelo ( $F_{Obj} > \chi^2_{max}$ ) e o coeficiente de correlação é baixo ( $\rho^m < 0.9$ ), o problema central parece ser a má qualidade do modelo. Idealmente, um bom modelo e um bom plano experimental vão levar simultaneamente a  $\chi^2_{min} < F_{Obj} < \chi^2_{max}$  e  $\rho^m > 0.9$ .

É importante que se perceba ainda que, uma vez obtidos os parâmetros do modelo, é possível montar uma tabela na forma da Tabela 4.1, onde são apresentados os dados obtidos, que pode ser transformada na Tabela 4.2, onde são explicitados os desvios experimentais.

**Tabela 4.1** – Dados experimentais e calculados com o modelo.

$x_{1,1}^e$	...	$x_{NX,1}^e$	$y_{1,1}^e$	...	$y_{NY,1}^e$	$x_{1,1}^m$	...	$x_{NX,1}^m$	$y_{1,1}^m$	...	$y_{NY,1}^m$
...	...		...	...		...	...		...	...	
$x_{1,NE}^e$	...	$x_{NX,NE}^e$	$y_{1,NE}^e$	...	$y_{NY,NE}^e$	$x_{1,NE}^m$	...	$x_{NX,NE}^m$	$y_{1,NE}^m$	...	$y_{NY,NE}^m$

**Tabela 4.2** – Desvios de modelagem.

$\epsilon x_{1,1} = x_{1,1}^e - x_{1,1}^m$	...	$\epsilon x_{NX,1} = x_{NX,1}^e - x_{NX,1}^m$	$\epsilon y_{1,1} = y_{1,1}^e - y_{1,1}^m$	...	$\epsilon y_{NY,1} = y_{NY,1}^e - y_{NY,1}^m$
...	...	...	...	...	...
$\epsilon x_{1,NE} = x_{1,NE}^e - x_{1,NE}^m$	...	$\epsilon x_{NX,NE} = x_{NX,NE}^e - x_{NX,NE}^m$	$\epsilon y_{1,NE} = y_{1,NE}^e - y_{1,NE}^m$	...	$\epsilon y_{NY,NE} = y_{NY,NE}^e - y_{NY,NE}^m$

Ora, se as hipóteses do experimento bem feito e do modelo perfeito são boas, a Tabela 4.2 contém amostras dos erros experimentais para cada uma das variáveis do problema. Portanto, as técnicas desenvolvidas no Capítulo 3 podem ser usadas agora para comparar essas amostras dos erros experimentais com os resultados independentes obtidos no laboratório a partir de réplicas, durante a fase de caracterização dos erros experimentais. Por exemplo, se os erros experimentais são normalmente distribuídos, cada uma das colunas da Tabela 4.2 deve ter média zero (pode ser usado o teste t para verificar a significância dessa hipótese, como ilustrado no Exemplo 3.18) e variância compatível com as variâncias experimentais (pode ser usado o teste F para verificar a significância dessa hipótese, como ilustrado no Exemplo 3.18). Se as medidas forem independentes, espera-se que as correlações existentes entre as medidas das várias colunas sejam fracas e pouco significativas (pode ser usada a ferramenta estatística mostrada na Seção 3.4.4, para caracterização da independência das medidas). Finalmente, espera-se ainda que os desvios sigam de forma aproximada a distribuição de probabilidades postulada para os erros experimentais (podem ser usadas as ferramentas estatísticas mostradas na Seção 3.4.3 e no Exemplo 3.20, para testes de aleatoriedade). Portanto, a organização dos dados na forma sugerida pelas Tabelas 4.1 e



4.2 permite a utilização de um grande arsenal de técnicas estatísticas para validação (ou não) das hipóteses utilizadas durante a construção do modelo e a aplicação do procedimento de estimação de parâmetros.

É conveniente comparar graficamente as distribuições esperadas para os desvios experimentais e as obtidas depois do procedimento de estimação de parâmetros, como discutido no Exemplo 3.20. Alguns padrões típicos são apresentados nas Figuras 4.15 a 4.17.

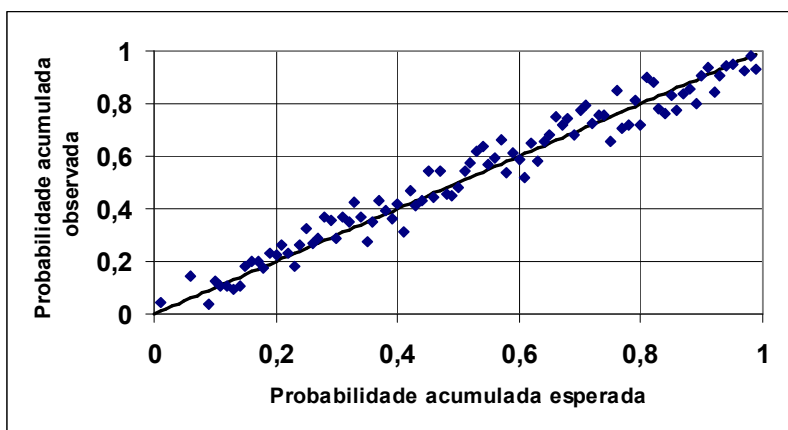


Figura 4.15 – Padrão de acúmulo de erros esperado para um bom modelo e boa caracterização de erros experimentais.

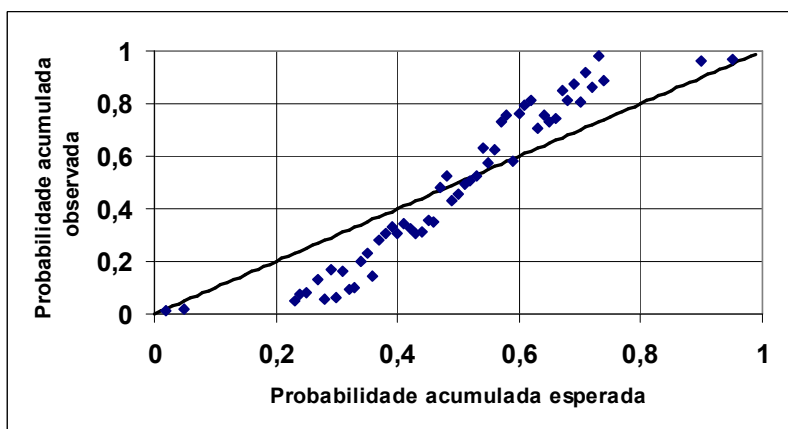


Figura 4.16 – Padrão de acúmulo de erros esperado quando existem outliers.

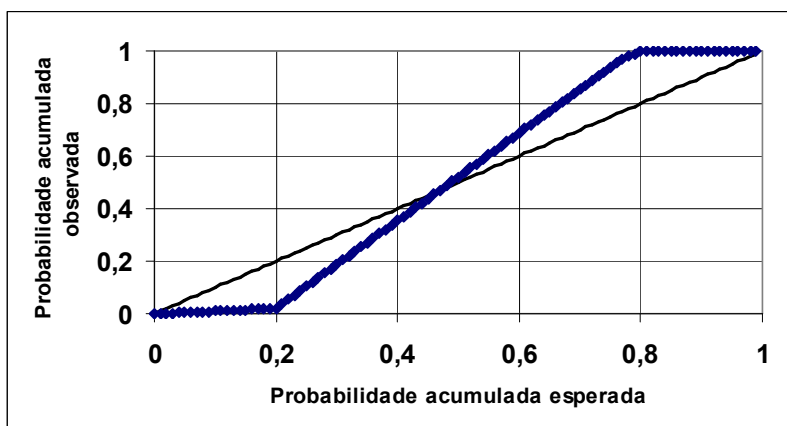


Figura 4.17 – Padrão de acúmulo de erros esperado quando o modelo é ruim ou existe má compreensão sobre a distribuição e natureza dos erros experimentais.

### 4.7.2. A Qualidade dos Parâmetros Obtidos

Para que a qualidade dos parâmetros obtidos possa ser avaliada, é fundamental calcular a **matriz de covariâncias dos parâmetros**  $V_{\alpha}$ , como nas Equações (4.28), (4.53), (4.74) e (4.82). Deve ser observado que soluções analíticas podem ser obtidas somente para modelos lineares. Na grande maioria das vezes os modelos são não lineares nos parâmetros e a matriz de incertezas paramétricas tem que ser construída numericamente, através de aproximações (por exemplo, a aproximação de Gauss), como mostrado na seção anterior.

Com a matriz de covariâncias dos parâmetros é possível obter várias informações relevantes sobre o problema analisado:

a) Definição dos intervalos de confiança dos parâmetros

Admitindo-se a distribuição normal e conhecendo-se as incertezas paramétricas, é possível construir os intervalos de confiança dos parâmetros como no Exemplo 3.2 do Capítulo 3. Nesse caso,

$$\hat{\alpha}_i - u\sigma_{\alpha_i} < \alpha_i < \hat{\alpha}_i + u\sigma_{\alpha_i} \tag{4.90a}$$

onde o valor de  $u$  depende do nível de confiança requerido,  $\hat{\alpha}_i$  é o valor estimado para o parâmetro e  $\sigma_{\alpha_i}$  é o desvio padrão associado à estimativa do parâmetro. No entanto, a definição dos intervalos de confiança para os parâmetros do modelo é um tema polêmico. Alguns textos sugerem o uso da distribuição  $t$  de Student para a definição do intervalo de confiança na forma:

$$\hat{\alpha}_i - t\sigma_{\alpha_i} < \alpha_i < \hat{\alpha}_i + t\sigma_{\alpha_i} \tag{4.90b}$$

onde  $t$  é obtido com o número de graus de liberdade da estimação e com o grau de confiança arbitrado pelo usuário. Quando o número de graus de liberdade é elevado (digamos, superior a 20), a Equação (4.90b) é essencialmente igual à Equação (4.90a). Para números de graus de liberdade inferiores a 20, a Equação (4.90b) leva a resultados

mais conservativos (intervalos de confiança mais largos). Para modelos lineares e sem erro na variável independente  $x$ , é possível associar a estimação de parâmetros feita por mínimos quadrados com uma operação de média amostral, que pode justificar o uso da distribuição  $t$  de Student na Equação (4.90b). Mas esse universo de condições é extremamente restritivo, de maneira que o intervalo de confiança dos parâmetros deveria ser construído com ferramentas numéricas mais poderosas, como as apresentadas no Capítulo 5. Por isso, alguns textos sugerem simplesmente que o intervalo de confiança do parâmetro seja construído na forma:

$$\hat{\alpha}_i - c\sigma_{\alpha_i} < \alpha_i < \hat{\alpha}_i + c\sigma_{\alpha_i} \quad (4.90c)$$

onde  $c$  é um número real positivo maior do que o valor sugerido pela curva normal para um determinado nível de confiança. Pelas razões apontadas, sugere-se que o usuário sempre defina de forma clara a forma com que se está calculando o intervalo de confiança dos parâmetros.

#### b) Grau de significância do parâmetro

O grau de significância é normalmente calculado em relação à referência zero, em que o efeito paramétrico do modelo desaparece. De forma simples, o grau de significância pode ser obtido a partir da Equação (4.90a-c), variando-se o grau de confiança até que o intervalo de confiança do parâmetro inclua o zero. Se o grau de confiança necessário para incluir o zero é inferior ao grau de confiança estabelecido pelo usuário, diz-se que o parâmetro é não significativo. Nesse caso, há argumentos estatísticos suficientes para remover o parâmetro (e o respectivo efeito) do modelo matemático. Caso contrário, diz-se que o parâmetro é significativo e há razões estatísticas para manter o parâmetro (e o respectivo efeito) no modelo.

#### c) Correlação paramétrica

De forma análoga à da Equação (1.50) da Seção 1.6, define-se o coeficiente de correlação paramétrica na forma:

$$\rho_{ij} = \frac{\sigma_{ij}^2}{\sigma_i \sigma_j} \quad (4.91)$$

Quanto mais próximos de zero estiverem os coeficientes de correlação paramétrica, mais eficientes serão os procedimentos de estimação dos parâmetros e mais precisa será a identificação dos diferentes efeitos no modelo. Quando a correlação paramétrica supera em módulo o valor de 0.9, é conveniente que o usuário reflita sobre a verdadeira necessidade de introduzir esses parâmetros no modelo, dado que pequenas mudanças no valor de um dos parâmetros podem ser compensadas com mudanças de um segundo parâmetro que está a ele relacionado. Dessa maneira, a correlação paramétrica indica que flutuações de alguns parâmetros podem ser acomodadas por variações de outros parâmetros, de forma que talvez seja possível reduzir o número de parâmetros do modelo. Por exemplo, o modelo apresentado na Equação (4.92a), muito utilizado para a descrição de modelos cinéticos, sugere a existência de 3 parâmetros:  $K_1$ ,  $K_2$  e  $K_3$ . No entanto, a Equação (4.92b) mostra que há apenas dois parâmetros no modelo:  $(K_1 / K_2)$  e  $(K_3 / K_2)$ . Portanto, a forma da Equação (4.92a) está errada, sob o

ponto de vista de estimação de parâmetros, dado que não é possível separar os efeitos paramétricos uns dos outros. Repare ainda que a definição dos parâmetros não é única, dado que qualquer um dos parâmetros poderia ser utilizado no denominador. Esse é o clássico exemplo de correlação paramétrica induzida pela formulação matemática do modelo.

$$y = \frac{K_1 x}{K_2 + K_3 x} \quad (4.92a)$$

$$y = \frac{\left(\frac{K_1}{K_2}\right) x}{1 + \left(\frac{K_3}{K_2}\right) x} = \frac{A_1 x}{1 + A_2 x} \quad (4.92b)$$

Correlações paramétricas elevadas às vezes são também geradas por planejamento experimental ineficiente, como ilustrado a seguir e discutido no Volume II dessa série de publicações. Por exemplo, suponha que um modelo pode ser escrito na forma

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 \quad (4.93a)$$

Não há nada de errado com a formulação do modelo apresentado na Equação (4.93). No entanto, suponha ainda que os dados experimentais são tais que  $x_1^e = x_2^e$ . Nesse caso, quando o modelo é aplicado à malha experimental, conclui-se que

$$y = \alpha_1 x_1^e + \alpha_2 x_2^e + \alpha_3 = (\alpha_1 + \alpha_2) x_1^e + \alpha_3 = A_1 x_1^e + \alpha_3 \quad (4.93b)$$

Portanto, apesar do modelo estar definido corretamente, parece claro que não é possível separar os efeitos de  $x_1$  e  $x_2$  na malha experimental proposta (ou seja, não é possível estimar  $\alpha_1$  e  $\alpha_2$  independentemente).

Para piorar, mesmo que não haja problemas nem com a formulação do modelo nem com a proposição da malha experimental, é possível que efeitos numéricos causem o aparecimento de correlações paramétricas e de problemas para a estimação independente dos parâmetros. Por exemplo, considere a Equação (4.92b). Suponha que  $A_2$  é muito grande. Nesse caso, a Equação (4.92b) ganha a forma:

$$y = \frac{A_1 x}{1 + A_2 x} \approx \left(\frac{A_1}{A_2}\right) = \left(\frac{K_1}{K_3}\right) \quad (4.92c)$$

de maneira que apenas um parâmetro está efetivamente presente no modelo. Suponha agora que  $A_2$  é muito pequeno. Nesse caso,

$$y = \frac{A_1 x}{1 + A_2 x} \approx A_1 x = \left(\frac{K_1}{K_2}\right) x \quad (4.92d)$$

e, uma vez mais, apenas um parâmetro está efetivamente presente no modelo. Esse tipo de correlação paramétrica é muito difícil de avaliar *a priori*, porque depende da magnitude relativa dos parâmetros. Na grande maioria das vezes, e em particular quando o modelo é não linear e contém muitos parâmetros, o usuário não conhece a magnitude relativa dos efeitos, de forma que não é possível eliminar esses efeitos antes de realizar a estimação. Isso torna o cômputo das correlações paramétricas fundamental para a correta avaliação da qualidade final dos resultados obtidos.

É importante ressaltar que correlações paramétricas elevadas às vezes não têm como ser evitadas, por resultarem da estrutura intrínseca do modelo matemático, o que é comum em modelos não lineares, como o modelo de Arrhenius (veja o Exemplo 4.10). Contudo, correlações elevadas sempre indicam problemas de estimação, que devem ser evitados e compreendidos. Uma das conseqüências práticas da existência de correlações paramétricas é o mau condicionamento da matriz  $\mathbf{H}_\alpha$ ; ou seja, em outras palavras, a matriz Hessiana usada amplamente nas seções anteriores pode ser não inversível (ou difícil de inverter numericamente). Obviamente, isso pode prejudicar toda a análise numérica proposta nas seções anteriores, já que a inversa de  $\mathbf{H}_\alpha$  é usada em vários procedimentos. (Isso indica que a invertibilidade da matriz Hessiana pode ser usada como ferramenta para identificação da existência de correlações paramétricas inaceitáveis no modelo.) Uma forma possível de minimizar os efeitos associados a correlações paramétricas é a reparametrização do modelo, como será discutido no Capítulo 5.

**Exemplo 4.28** - Considere o modelo linear

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3$$

e a função de mínimos quadrados

$$F_{Obj} = \sum_{i=1}^{NE} (y_i^e - \alpha_1 x_{1,i}^e - \alpha_2 x_{2,i}^e - \alpha_3)^2$$

Nesse caso, a matriz Hessiana pode ser dada na forma da Equação (4.17) como:

$$\mathbf{H}_\alpha = \begin{bmatrix} 2 \sum_{i=1}^{NE} (x_{1,i}^e)^2 & 2 \sum_{i=1}^{NE} x_{1,i}^e x_{2,i}^e & 2 \sum_{i=1}^{NE} x_{1,i}^e \\ 2 \sum_{i=1}^{NE} x_{1,i}^e x_{2,i}^e & 2 \sum_{i=1}^{NE} (x_{2,i}^e)^2 & 2 \sum_{i=1}^{NE} x_{2,i}^e \\ 2 \sum_{i=1}^{NE} x_{1,i}^e & 2 \sum_{i=1}^{NE} x_{2,i}^e & 2 NE \end{bmatrix}$$

Se ao longo dos experimentos os valores das variáveis  $x_{1,i}^e$  e  $x_{2,i}^e$  são iguais, então as linhas 1 e 2 da matriz Hessiana são iguais. Por conseguinte, a matriz Hessiana é não inversível.

### 4.7.3. A Qualidade das Previsões do Modelo

Para que a qualidade das previsões feitas com o modelo seja avaliada, é necessário calcular como os erros paramétricos se propagam através do modelo e viram erros de predição. Para alguns casos simples, onde o modelo é linear nos parâmetros, uma derivação teórica pode ser obtida, como pode ser observado nas Equações (4.31) e (4.56). Contudo, como no caso das incertezas paramétricas, os erros de predição de modelos não lineares têm que ser calculados com a ajuda de aproximações, como a obtida na Equação (4.86), ou de ferramentas numéricas.

Por exemplo, considere o caso apresentado no Exemplo 4.19, que trata de uma reta. Nesse caso, a variância de predição pode ser escrita como

$$\hat{\sigma}_y^2 = \mathbf{B}^T \mathbf{V}_a \mathbf{B} = [x \quad 1] \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} = \sigma_{11}^2 x^2 + 2\sigma_{12}^2 x + \sigma_{22}^2 \quad (4.94)$$

que mostra que os erros de predição crescem na forma de uma parábola, à medida que  $x$  cresce em valor absoluto. A Equação (4.94) é usada freqüentemente para justificar a frase de que a extrapolação é menos precisa que a interpolação. Contudo, é importante enfatizar que a Equação (4.94) é válida unicamente para a reta e não deve ser usada como argumento para outros modelos. Por exemplo, no caso do modelo na forma:

$$y = 1 - e^{-\alpha x} \quad (4.95)$$

é possível escrever

$$\hat{\sigma}_y^2 = \mathbf{B}^T \mathbf{V}_a \mathbf{B} = [x e^{-\alpha x}] [\sigma_\alpha^2] [x e^{-\alpha x}] = \sigma_\alpha^2 x^2 e^{-2\alpha x} \quad (4.96)$$

que tende a zero quando  $x$  cresce indefinidamente; ou seja, as incertezas das extrapolações diminuem. Portanto, não é possível garantir que as incertezas de predição crescem, quando nos afastamos da região experimental, sem que se diga claramente o modelo usado para fazer a extrapolação. O modelo da Equação (4.95) apresenta comportamento inverso.

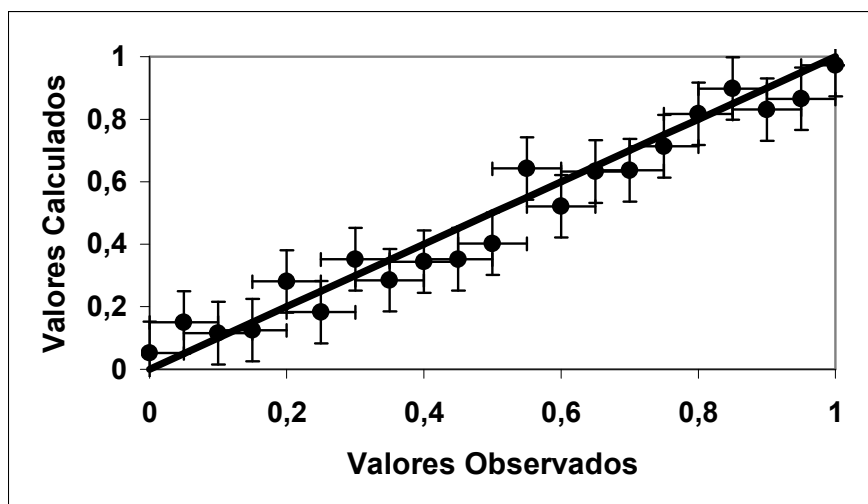
Expressões como a apresentada na Equação (4.85-86) permitem também definir os **intervalos de confiança de predição** do modelo, de forma similar à utilizada para definição dos intervalos de confiança dos parâmetros

$$y_i^m - u\hat{\sigma}_{yi} < y_i < y_i^m + u\hat{\sigma}_{yi} \quad (4.97)$$

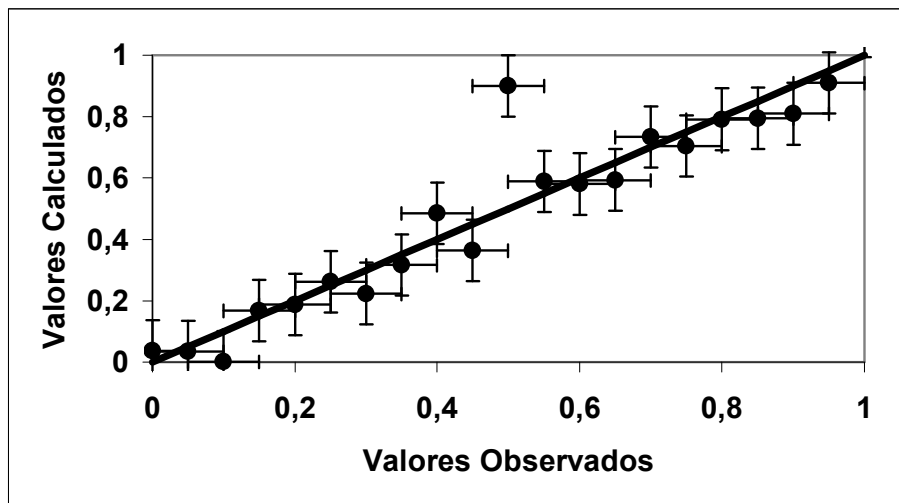
Esses intervalos de confiança podem ser usados para detectar pontos experimentais suspeitos, que não podem ser explicados pelo modelo. Como discutido na Seção 3.3.4, o procedimento de detecção de *outliers* é bastante controverso. Um ponto experimental não deve ser descartado, a não ser que a repetição do experimento confirme a existência de um erro grosseiro ou que haja razões muito específicas para condenar a observação experimental. Como já discutido, um ponto experimental que não se adequa ao modelo pode ser uma grande oportunidade científica.

Os erros de predição podem ser também usados para definir a região experimental onde modelo e medida experimental são equivalentes do ponto de vista estatístico, definindo assim a região de utilidade do modelo. Isso pode ser feito comparando-se o erro de predição, que é uma função do ponto experimental considerado, com o erro experimental através do teste F. Sempre que os dois erros puderem ser considerados equivalentes, o modelo pode ser considerado válido e útil naquela região.

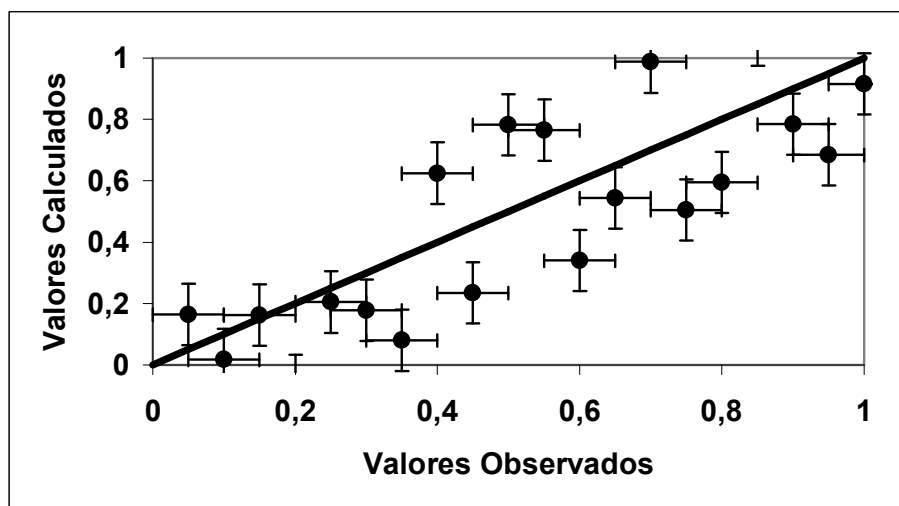
Como feito anteriormente com os desvios observados entre as predições do modelo e os dados experimentais, e ilustrado nas Figuras 4.15 a 4.17, é também conveniente comparar graficamente o desempenho dos modelos. Isso pode ser feito de forma muito simples com o auxílio de gráficos, como os ilustrados nas Figuras 4.18 a 4.21. Nesses gráficos, os valores previstos pelo modelo e observados experimentalmente são comparados visualmente, permitindo a obtenção de informações valiosas a respeito da qualidade da estimação realizada. Nos gráficos das Figuras 4.18 a 4.21, as barras verticais caracterizam os erros de predição do modelo, enquanto as barras horizontais caracterizam os erros das medidas experimentais.



**Figura 4.18** – Padrão típico de comparação entre os dados calculados pelo modelo e os dados observados experimentalmente, quando o modelo é bom e os experimentos são bem feitos.

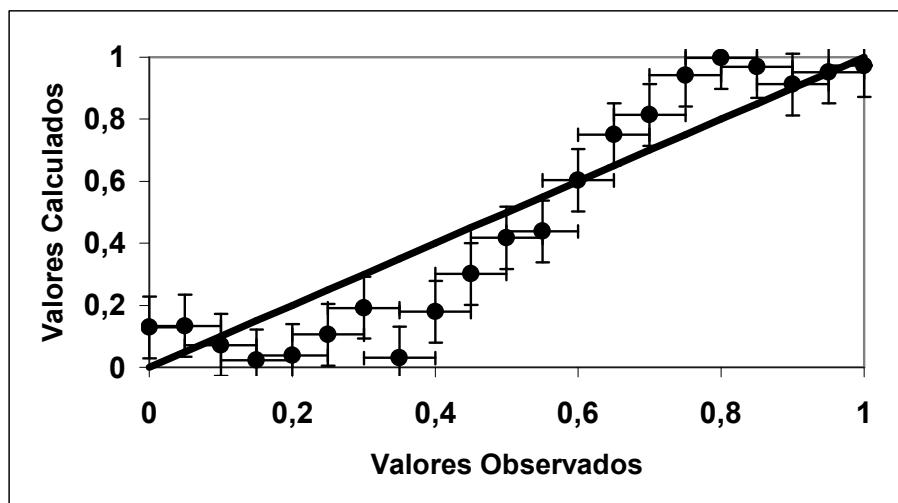


**Figura 4.19** – Padrão típico de comparação entre os dados calculados pelo modelo e os dados observados experimentalmente, quando há candidatos a *outliers* (o ponto isolado que não pode ser explicado pelos erros de medida).



**Figura 4.20** – Padrão típico de comparação entre os dados calculados pelo modelo e os dados observados experimentalmente, quando o modelo é ruim ou quando os desvios experimentais estão subestimados.





**Figura 4.21** – Padrão típico de comparação entre os dados calculados pelo modelo e os dados observados experimentalmente, quando o modelo não apresenta desempenho uniforme na região de experimentação (erra mais para um lado do que para outro).

**Exemplo 4.29** - Considere o seguinte conjunto de  $NE = 6$  dados experimentais

$x_i^e$	0	1	2	3	4	5	$\Rightarrow \sum_{i=1}^6 x_i^e = 15$
$y_i^e$	0.1	0.9	2.2	3.1	3.9	4.8	$\Rightarrow \sum_{i=1}^6 y_i^e = 15$
$(x_i^e)^2$	0	1	4	9	16	25	$\Rightarrow \sum_{i=1}^6 (x_i^e)^2 = 55$
$y_i^e x_i^e$	0	0.9	4.4	9.3	15.6	24	$\Rightarrow \sum_{i=1}^6 (y_i^e x_i^e) = 54.2$

aos quais se pretende ajustar uma reta através de procedimento padrão de mínimos quadrados. Nesse caso,

$$y^m = \alpha x + \beta$$

e o número de graus de liberdade é  $v = 4$ . Logo, de acordo com o Exemplo 4.20,

$$\alpha = \frac{6 \cdot 54.2 - 15 \cdot 15}{6 \cdot 55 - 15 \cdot 15} = 0.9542857143$$

$$\beta = \frac{15 - 0.9542857143 \cdot 15}{6} = 0.114285715$$

resultando no seguinte conjunto de predições feitas com o modelo

$$y_i^m : 0.1142857; 1.0685714; 2.0228571; 2.9771429; 3.9314287; 4.8857143$$

Portanto, o erro médio global de predição oriundo da estimação fica na forma

$$\hat{\sigma}_y^2 = \frac{\sum_{i=1}^{NE} (y_i^e - y_i^m)^2}{4} = 0.02085714281$$

$$\hat{\sigma}_y = \sqrt{\hat{\sigma}_y^2} = 0.1444200222$$

enquanto o coeficiente de correlação do modelo fica na forma

$$\rho^m = 0.9974$$

que indica fortíssima correlação entre os valores experimentais e calculados através do modelo, indicativo também de excelente grau de ajuste.

Admitindo-se que o erro experimental  $\sigma_y^2$  é igual ao erro oriundo do processo de estimação de parâmetros  $\hat{\sigma}_y^2$

$$\sigma_\alpha^2 = 0.020855714281 \frac{6}{6 \cdot 55 - 15 \cdot 15} = 0.00191836732$$

$$\sigma_\alpha = \sqrt{\sigma_\alpha^2} = 0.03452298846$$

$$\sigma_\beta^2 = 0.020855714281 \frac{55}{6 \cdot 55 - 15 \cdot 15} = 0.001546598622$$

$$\sigma_\beta = \sqrt{\sigma_\beta^2} = 0.1243623183$$

de maneira que com 95% de confiança,  $u = 1.96$ , resultando em

$$\alpha = 0.95 \pm 0.07$$

$$\beta = 0.11 \pm 0.25$$

Portanto, o parâmetro  $\alpha$  é significativo, enquanto o parâmetro  $\beta$  é não significativo. Assim, há justificativas para que se tente uma modelagem do tipo  $y^m = \alpha x$ .

A covariância entre os parâmetros  $\alpha$  e  $\beta$  é igual a

$$\sigma_{\alpha\beta}^2 = 0.020855714281 \frac{(-15)}{6 \cdot 55 - 15 \cdot 15} = -0.00297959183$$

resultando em um coeficiente de correlação paramétrica igual a

$$\rho_{\alpha\beta} = \frac{\sigma_{\alpha\beta}^2}{\sigma_{\alpha}\sigma_{\beta}} = -0.6940$$

indicando moderado grau de correlação entre os dois parâmetros. O sinal negativo indica que perturbações positivas em um dos parâmetros provocam perturbações negativas no outro e vice-versa.

O erro de predição exclusivo do modelo pode ser dado por

$$\hat{\sigma}_y^2 = 0.00191836732 \cdot x^2 - 2 \cdot 0.00297959183x + 0.01546598622$$

que se for somado ao erro experimental, tem a forma

$$\hat{\sigma}_y^2 = 0.00191836732 \cdot x^2 - 2 \cdot 0.00297959183x + 0.01546598622 + 0.02085714281$$

Aplicando-se o teste F para 4 graus de liberdade e grau de confiança de 95%, chega-se a

$$F = 9.6045 > \frac{\hat{\sigma}_y^2}{\sigma_y^2} = 0.092x^2 - 0.2857x + 1.742$$

de maneira que

$$0.092x^2 - 0.2857x - 7.863 < 0$$

$$-7.82 < x < 10.92$$

é a região de validade do modelo. Admitamos, por exemplo, que a realidade fosse  $y = x$ , plausível a partir da análise dos dados experimentais. Neste caso, para  $x = 10$ ,  $y = 10$  e  $y^m = 9.654$ , cujo desvio já é de cerca de 2.5 vezes o valor de  $\sigma_y$ .

A Figura 4.22 ilustra a qualidade do processo de estimação realizado

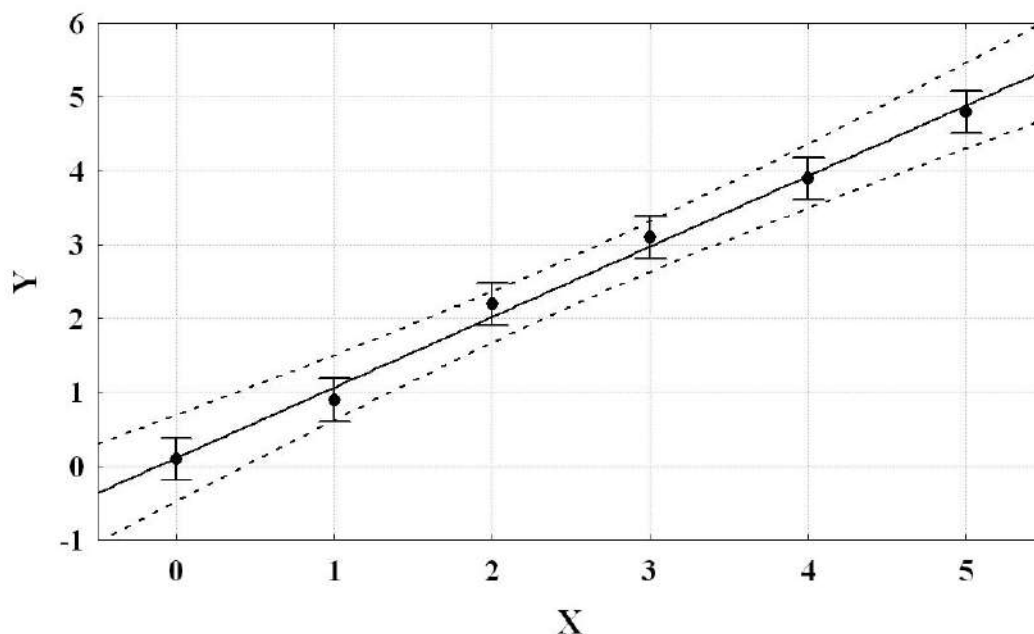


Figura 4.22 - Resultado do procedimento de estimação de parâmetros do Exemplo 4.29.

## 4.8. Conclusões

Foi definido nesse capítulo o problema de estimação de parâmetros. Um procedimento foi construído para inferir os valores de certas variáveis (os parâmetros) que não podem ser medidas, mas sem as quais os modelos matemáticos não têm utilidade. O problema é constituído por três etapas: uma etapa de formulação de uma função objetivo, uma etapa de minimização da função objetivo e uma etapa de interpretação dos resultados finais. Para formulação da função objetivo a ser minimizada durante a segunda etapa do procedimento proposto, foi criada uma metodologia de máxima verossimilhança. A metodologia propõe que os erros de modelagem sejam usados como amostras dos erros experimentais, o que está em consonância com as hipóteses de existência de um bom modelo e de bons dados experimentais. Essas hipóteses podem ser validadas (ou rejeitadas) *a posteriori*, depois de obtidos os resultados da estimação.

Mostrou-se através de exemplos que o usuário não deve ter a expectativa de gerar soluções analíticas para o problema de estimação de parâmetros para problemas genéricos, o que justifica o desenvolvimento dos métodos numéricos a serem apresentados no próximo capítulo. Apesar disso, várias expressões matemáticas úteis foram derivadas, para permitir a interpretação matemática (estatística) dos resultados obtidos, quando os erros experimentais não são muito grandes.

## 4.9. Leitura Adicional

A literatura dedicada à apresentação do problema de estimação de parâmetros é muito vasta. Não cabe aqui, portanto, uma revisão dessa área. O leitor interessado

encontrará centenas de livros que abordam esse tema em qualquer biblioteca dedicada à Matemática e à Engenharia. Alguns textos clássicos que tratam do assunto são apresentados a seguir, para eventuais consultas e aprofundamento dos estudos.

“*Nonlinear Parameter Estimation*”, Y. Bard, Academic Press Inc, San Diego, **1974**.

“*Nonlinear Regression Analysis and Its Applications*”, D. M. Bates e D. G. Watts, John Wiley & Sons, New York, **1988**.

“*Applied Nonlinear Regression*”, N. R. Draper e H. Smith, John Wiley & Sons, New York, **1998**.

“*Applied Parameter Estimation for Chemical Engineers*”, P. Englezos e N. Kalogerakis, Marcel Dekker Inc, New York, **2001**.

#### 4.10. Exercícios Sugeridos

1- Os seguintes dados estão disponíveis:

NE	$x$	$y$
1	0.10	0.38
2	0.20	0.91
3	0.30	1.69
4	0.40	2.13
5	0.50	2.66
6	0.60	2.61
7	0.70	3.65
8	0.80	3.94
9	0.90	4.28
10	1.00	5.24

Admitindo-se que o modelo  $y = ax$  é válido, onde  $a$  é o parâmetro a ser determinado, estime o melhor valor de  $a$  e a respectiva incerteza paramétrica nos três casos abaixo:

a)  $F_a = \sum_{i=1}^{10} (y_i^e - y_i^c)^2$  ;

b)  $F_b = \sum_{i=1}^{10} (y_i^e - y_i^c)^8$  ;

c)  $F_c = \sum_{i=1}^{10} \left( \frac{y_i^e - y_i^c}{y_i^e} \right)^2$  ;

d) Compare os valores obtidos.

2- Supondo que os erros de uma certa medida são independentes e sigam a distribuição exponencial, definida como:

$$P(\varepsilon_i) = \frac{1}{2\sigma_i} \exp\left(-\frac{|\varepsilon_i|}{\sigma_i}\right), \quad -\infty < \varepsilon_i < \infty$$

onde  $\varepsilon_i$  são os desvios ou erros experimentais e  $\sigma_i$  é o desvio padrão, defina a função objetivo a ser usada em um procedimento de estimação de parâmetros que utiliza estes dados.

3- Determine as expressões para os parâmetros  $A$  e  $B$  do modelo linear:

$$y = Ax + B$$

para as seguintes funções objetivos.

a)  $F = \sum_{i=1}^{NE} (y_i^e - y_i^m)^2$

b)  $F = \sum_{i=1}^{NE} \frac{(y_i^e - y_i^m)^2}{\sigma_i^2}$

c)  $F = \sum_{i=1}^{NE} \sum_{j=1}^{NE} (y_i^e - y_i^m)(y_j^e - y_j^m)[v_{ij}]^{-1}$ , onde  $[v_{ij}]^{-1}$  o elemento  $ij$  da inversa da matriz  $\mathbf{V}_y$  de erros experimentais.

d) Em seguida, usando os dados abaixo, calcule os valores dos parâmetros, a matriz de covariância dos parâmetros e o erro de predição para cada função objetivo.

$i$	$x_i^e$	$y_i^e$
1	1	2
2	2	3
3	3	7

$$\mathbf{V} = \begin{bmatrix} 1.0 & 0.5 & 0.1 \\ 0.5 & 1.0 & 0.5 \\ 0.1 & 0.5 & 10.0 \end{bmatrix}$$