

Capítulo 5

Procedimentos Numéricos para Estimação de Parâmetros

Os procedimentos numéricos usados para fins de estimação de parâmetros são fundamentalmente procedimentos numéricos que permitem a otimização (minimização) de uma certa função objetivo. Como visto no Capítulo 4, a otimização (minimização) da função objetivo admite solução analítica raramente, como no caso em que o modelo é linear nos parâmetros. Entretanto, a maioria absoluta dos modelos encontrados no dia-a-dia de cientistas e engenheiros, em particular na área da Engenharia Química, é constituída por modelos não lineares, de maneira que não é possível de forma geral obter soluções analíticas em problemas de estimação de parâmetros. Por isso, são necessários métodos numéricos para encontrar a solução do problema.

É importante observar que, uma vez definidos a função objetivo, o modelo e os dados experimentais, o mínimo da função objetivo já está também definido. O papel do método de otimização é encontrar esta solução desconhecida pelo usuário. Porém, esta tarefa pode ser muito difícil: pode haver muitos dados experimentais, os modelos podem ter comportamento complexo, a função objetivo pode ter múltiplos mínimos locais, os parâmetros podem estar correlacionados (a matriz Hessiana pode ser mal condicionada e difícil de inverter), o número de parâmetros pode ser elevado, etc. Devido a estas dificuldades, foi proposta na literatura uma infinidade de métodos diferentes de otimização, cada qual com suas particularidades. Por isso, a eficiência desses métodos pode variar muito de problema para problema. Assim, não existe um único método de otimização que seja capaz de resolver todos os problemas de estimação. Alguns métodos são muito eficientes em certos problemas, mas não são capazes de solucionar um outro problema com características um pouco diferentes.

Nas próximas seções, será definido o problema numérico de otimização e serão apresentados alguns métodos numéricos comumente utilizados em problemas de estimação de parâmetros. O objetivo desse capítulo não é rever a vasta literatura da área, mas oferecer ao leitor perspectivas numéricas que permitam compreender pacotes comerciais e construir seu próprio ambiente de estimação.

5.1 Definição do Problema Numérico de Otimização

O problema de estimação de parâmetros consiste na minimização de uma função objetivo, que mede a distância entre os dados experimentais e os dados preditos (calculados) por um modelo, conforme amplamente discutido no Capítulo 4. Considerando que os erros das variáveis medidas têm distribuição normal e que as medidas das variáveis independentes \mathbf{x} e das variáveis dependentes \mathbf{y} são independentes, a seguinte função objetivo pode ser obtida:

$$F_{Obj} = \sum_{i=1}^{NE} \left\{ (\mathbf{y}_i^e - \mathbf{y}_i^m)^T \mathbf{V}_{\mathbf{y}_i}^{-1} (\mathbf{y}_i^e - \mathbf{y}_i^m) + (\mathbf{x}_i^e - \mathbf{x}_i^m)^T \mathbf{V}_{\mathbf{x}_i}^{-1} (\mathbf{x}_i^e - \mathbf{x}_i^m) \right\} \quad (5.1)$$

Quando as variáveis independentes são conhecidas com grande precisão, a Equação (5.1) fica reduzida a:

$$F_{Obj} = \sum_{i=1}^{NE} (\mathbf{y}_i^e - \mathbf{y}_i^m)^T \mathbf{V}_{\mathbf{y}_i}^{-1} (\mathbf{y}_i^e - \mathbf{y}_i^m) \quad (5.2)$$

É importante observar, como discutido no Capítulo 4, que as Equações (5.1-2) são apenas exemplos de um conjunto muito mais amplo de possibilidades, já que a função objetivo pode apresentar muitas formas diferentes, a depender do problema analisado. Apesar disso, não ser que seja dito explicitamente o contrário, as Equações (5.1-2) serão usadas para a formulação dos problemas de estimação de parâmetros propostos neste capítulo. Deve ser ainda observado que, quando todas as medidas são realizadas de forma independente, as matrizes $\mathbf{V}_{\mathbf{y}_i}$ e $\mathbf{V}_{\mathbf{x}_i}$ são diagonais e a Equação (5.1) fica na forma da Equação (4.43) definida no Capítulo 4.

Considera-se que o modelo matemático que relaciona as variáveis independentes (\mathbf{x}), os parâmetros ($\boldsymbol{\alpha}$) e as variáveis dependentes (\mathbf{y}) pode ser escrito na forma:

$$\mathbf{y}_i^m = \mathbf{f}(\mathbf{x}_i^m, \boldsymbol{\alpha}) \quad (5.3)$$

onde, para uma dada condição experimental i , os valores das variáveis dependentes \mathbf{y}_i^m possam ser calculados (explicitamente ou numericamente) a partir das variáveis independentes \mathbf{x}_i^m e dos parâmetros $\boldsymbol{\alpha}$. (Observe que os parâmetros não variam com a condição experimental i , constituindo assim uma amarração teórica entre os pontos experimentais disponíveis.) Assim, a minimização da função objetivo definida na Equação (5.1) consiste em procurar os valores dos parâmetros $\boldsymbol{\alpha}$ e das variáveis independentes \mathbf{x}_i^m ($i=1\dots NE$) que fazem com que função F_{Obj} atinja um valor mínimo, já que os valores das variáveis dependentes \mathbf{y}_i^m são obtidos a partir de $\boldsymbol{\alpha}$ e \mathbf{x}_i^m , dadas as restrições matemáticas impostas pelo modelo. (Por isso, o problema de estimação de parâmetros pode ser definido como um problema de otimização com restrições.) Observe que o número de variáveis do problema de otimização é igual a NP parâmetros e NX variáveis independentes em cada uma das NE condições experimentais, somando $NP + NE.NX$ variáveis de otimização. Quando as variáveis independentes estão isentas de erro (Equação (5.2)), o número de variáveis de otimização cai para NP ; isto é, apenas os parâmetros do modelo são manipulados para minimizar F_{Obj} .

Uma forma conceitualmente simples de obter a solução do problema consiste em buscar o ponto onde o gradiente da função objetivo seja nulo. (Esta é uma condição necessária, mas não suficiente, já que o ponto encontrado pode ser um mínimo, um máximo ou um ponto de sela.) Assim, derivando-se a função objetivo em relação a cada uma das variáveis de otimização, pode ser obtido o seguinte conjunto de equações:

$$\mathbf{g}(\boldsymbol{\beta}) = \begin{bmatrix} g_1(\mathbf{x}_1^m, \dots, \mathbf{x}_{NE}^m, \boldsymbol{\alpha}) \\ \vdots \\ g_{NP}(\mathbf{x}_1^m, \dots, \mathbf{x}_{NE}^m, \boldsymbol{\alpha}) \\ g_{NP+1}(\mathbf{x}_1^m, \dots, \mathbf{x}_{NE}^m, \boldsymbol{\alpha}) \\ \vdots \\ g_{NP+NX*NE}(\mathbf{x}_1^m, \dots, \mathbf{x}_{NE}^m, \boldsymbol{\alpha}) \end{bmatrix} = \begin{bmatrix} \frac{\partial F_{Obj}}{\partial \alpha_1} \\ \vdots \\ \frac{\partial F_{Obj}}{\partial \alpha_{NP}} \\ \frac{\partial F_{Obj}}{\partial x_{1,1}^m} \\ \vdots \\ \frac{\partial F_{Obj}}{\partial x_{NX*NE}^m} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}, \quad \boldsymbol{\beta} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{NP} \\ x_{1,1}^m \\ \vdots \\ x_{NX,NE}^m \end{bmatrix} \quad (5.4)$$

que totalizam $NP + NE.NX$ equações algébricas, cuja solução fornece a solução do problema de estimação (a despeito do fato de esta ser apenas uma condição necessária, mas não suficiente, para a caracterização de um ponto mínimo). Considerando que as variáveis independentes são conhecidas com grande precisão, apenas os parâmetros do modelo precisam ser encontrados. Assim, o sistema de equações que precisa ser resolvido é

$$\mathbf{g}(\boldsymbol{\beta}) = \begin{bmatrix} g_1(\boldsymbol{\alpha}) \\ \vdots \\ g_{NP}(\boldsymbol{\alpha}) \end{bmatrix} = \begin{bmatrix} \frac{\partial F_{Obj}}{\partial \alpha_1} \\ \vdots \\ \frac{\partial F_{Obj}}{\partial \alpha_{NP}} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}, \quad \boldsymbol{\beta} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{NP} \end{bmatrix} \quad (5.5)$$

O sistema de equações algébricas descrito nas Equações (5.4-5) pode ser resolvido com o auxílio de técnicas numéricas clássicas, como o método de Newton-Raphson. O método de Newton-Raphson (e suas muitas variantes) propõe a solução numérica de um conjunto de equações algébricas de forma iterativa, a partir da solução de problemas mais simples gerados pela linearização sucessiva do sistema de equações que se pretende resolver. O método de Newton-Raphson pode ser definido de forma genérica na forma

$$\mathbf{g}(\boldsymbol{\beta}) \approx \mathbf{g}(\boldsymbol{\beta}_0) + \left. \frac{\partial \mathbf{g}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}_0} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) = \mathbf{g}_0 + \mathbf{J}_0 (\boldsymbol{\beta} - \boldsymbol{\beta}_0) = 0 \quad (5.6a)$$

$$\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0 - \mathbf{J}_0^{-1} \mathbf{g}_0$$

onde se supõe que o sistema de equações a ser resolvido ($\mathbf{g}(\boldsymbol{\beta})$) pode ser aproximado por uma reta nas proximidades de um ponto ($\boldsymbol{\beta}_0$), truncando-se a expansão em série de Taylor no primeiro termo (primeira derivada da função). Quanto mais próxima a

solução procurada (β) estiver do ponto inicial fornecido (β_0), mais acurada será a aproximação obtida da solução (β_1). Dessa forma, espera-se que o usuário seja capaz de fornecer uma estimativa inicial da solução (β_0), seja capaz de calcular as funções algébricas que pretende resolver ($g_0 = g(\beta_0)$) e as respectivas derivadas ($\left. \frac{\partial g(\beta)}{\partial \beta} \right|_{\beta_0} = J_0$) e seja capaz de inverter J_0 . Obviamente, o procedimento pode ser repetido indefinidamente e de forma iterativa na forma:

$$\beta_{k+1} = \beta_k - J_k^{-1} g_k \quad (5.6b)$$

onde k denota a k -ésima iteração do procedimento numérico, $g_k = g(\beta_k)$ e $J_k = \left. \frac{\partial g(\beta)}{\partial \beta} \right|_{\beta_k}$ é a **matriz Jacobiana** do sistema de equações, definida como

$$J_k = \begin{bmatrix} \left. \frac{\partial g_1(\beta)}{\partial \beta_1} \right|_{\beta_k} & \dots & \left. \frac{\partial g_1(\beta)}{\partial \beta_{NG}} \right|_{\beta_k} \\ \vdots & \ddots & \vdots \\ \left. \frac{\partial g_{NG}(\beta)}{\partial \beta_1} \right|_{\beta_k} & \dots & \left. \frac{\partial g_{NG}(\beta)}{\partial \beta_{NG}} \right|_{\beta_k} \end{bmatrix} \quad (5.7)$$

onde NG denota o número total de equações e variáveis que constituem o problema. Nos casos particulares analisados nas Equações (5.4-5), NG é igual respectivamente a $NP+NX*NE$ e NP .

Exemplo 5.1 - Considerando-se um problema de estimação de parâmetros de um modelo não linear, com medidas experimentais sujeitas a flutuações normais, sendo que as variáveis independentes são medidas com grande precisão, a função objetivo definida na Equação (5.2) pode ser utilizada. Um procedimento recursivo pode ser desenvolvido a partir da aproximação linear do modelo em torno de uma estimativa inicial dos parâmetros α^0 , como mostra a equação seguir:

$$y_i = f(x_i, \alpha) \approx f_i^0 + B_i(\alpha - \alpha^0)$$

onde i indica a condição experimental, f_i^0 é um vetor com dimensão NY na forma

$$f_i^0 = f(x_i, \alpha^0)$$

e B_i é uma matriz de dimensão $NY \times NP$ que contém as derivadas de cada variável dependente em relação a cada um dos parâmetros do modelo, também chamada de matriz de sensibilidades e definida na Equação (4.78) do capítulo anterior na forma:

$$\mathbf{B}_i = \begin{bmatrix} \left. \frac{\partial f_1(\mathbf{x}_i, \boldsymbol{\alpha})}{\partial \alpha_1} \right|_{\boldsymbol{\alpha}^0} & \dots & \left. \frac{\partial f_1(\mathbf{x}_i, \boldsymbol{\alpha})}{\partial \alpha_{NP}} \right|_{\boldsymbol{\alpha}^0} \\ \vdots & \ddots & \vdots \\ \left. \frac{\partial f_{NY}(\mathbf{x}_i, \boldsymbol{\alpha})}{\partial \alpha_1} \right|_{\boldsymbol{\alpha}^0} & \dots & \left. \frac{\partial f_{NY}(\mathbf{x}_i, \boldsymbol{\alpha})}{\partial \alpha_{NP}} \right|_{\boldsymbol{\alpha}^0} \end{bmatrix}$$

Substituindo a aproximação proposta acima na função objetivo definida na Equação (5.2), chega-se a:

$$F_{Obj} = \sum_{i=1}^{NE} \left[\mathbf{y}_i^e - \mathbf{f}_i^0 - \mathbf{B}_i (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) \right]^T \mathbf{V}_{Yi}^{-1} \left[\mathbf{y}_i^e - \mathbf{f}_i^0 - \mathbf{B}_i (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) \right]$$

Derivando-se a função objetivo mostrada acima em relação aos parâmetros,

$$\frac{\partial F_{Obj}}{\partial \boldsymbol{\alpha}} = -2 \sum_{i=1}^{NE} \mathbf{B}_i^T \mathbf{V}_{Yi}^{-1} \left[\mathbf{y}_i^e - \mathbf{f}_i^0 - \mathbf{B}_i (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) \right] = \mathbf{0}$$

Agrupando de forma conveniente os termos da equação acima

$$\sum_{i=1}^{NE} \mathbf{B}_i^T \mathbf{V}_{Yi}^{-1} \left[\mathbf{y}_i^e - \mathbf{f}_i^0 \right] - \sum_{i=1}^{NE} \mathbf{B}_i^T \mathbf{V}_{Yi}^{-1} \mathbf{B}_i (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) = \mathbf{0}$$

e definindo

$$\mathbf{U} = \sum_{i=1}^{NE} \mathbf{B}_i^T \mathbf{V}_{Yi}^{-1} \left[\mathbf{y}_i^e - \mathbf{f}_i^0 \right]$$

$$\mathbf{T} = \sum_{i=1}^{NE} \mathbf{B}_i^T \mathbf{V}_{Yi}^{-1} \mathbf{B}_i$$

onde \mathbf{U} é um vetor com dimensão NP e \mathbf{T} é uma matriz com dimensão $NP \times NP$, chega-se finalmente a

$$\boldsymbol{\alpha}^1 = \boldsymbol{\alpha}^0 - \mathbf{T}^{-1} \mathbf{U}$$

ou na forma iterativa

$$\boldsymbol{\alpha}^{k+1} = \boldsymbol{\alpha}^k - \mathbf{T}_k^{-1} \mathbf{U}_k$$

Assim, a partir de uma estimativa inicial $\boldsymbol{\alpha}^0$ proposta pelo usuário e do cálculo das derivadas do modelo em relação aos parâmetros (presentes nas matrizes \mathbf{T} e \mathbf{U}), a solução do problema de estimação pode ser obtida de forma recursiva. É claro que a garantia de convergência do procedimento recursivo está intimamente ligada à qualidade da estimativa inicial fornecida e à forma matemática do modelo, presente através das derivadas. Observe ainda que apenas as derivadas primeiras do modelo foram necessárias para o desenvolvimento do procedimento numérico, em decorrência da aproximação linear proposta para o modelo. Como será visto a seguir, o mesmo

resultado pode ser obtido quando um método de Newton é utilizado para a minimização da função objetivo, quando se utiliza a aproximação de Gauss para o cálculo da matriz Hessiana.

Exemplo 5.2 – Considerando um modelo definido na forma

$$y = \exp(-k x)$$

onde k é um parâmetro que deve ser estimado a partir dos seguintes dados experimentais:

i	x	y
1	0.0	0.98
2	0.1	0.76
3	0.3	0.54
4	0.6	0.29
5	1.0	0.15

Considerando ainda que os erros experimentais são sujeitos a flutuações normais com variância constante, a função objetivo pode ser definida como:

$$F_{Obj} = \sum_{i=1}^{NE} (y_i^e - y_i^m)^2 = \sum_{i=1}^{NE} (y_i^e - \exp(-k x_i^e))^2$$

Derivando-se a função objetivo em relação ao parâmetro k :

$$g_1 = \frac{\partial F_{Obj}}{\partial k} = 2 \sum_{i=1}^{NE} (y_i^e - \exp(-k x_i^e)) (-x_i^e) \exp(-k x_i^e) = 0$$

$$g_1 = \sum_{i=1}^{NE} (y_i^e - \exp(-k x_i^e)) x_i^e \exp(-k x_i^e) = 0$$

Neste caso, em que apenas um parâmetro é estimado, a matriz Jacobiana tem dimensão 1×1 e é composta apenas pelo valor da derivada de g_1 em relação ao único parâmetro do problema $\alpha_1 = k$.

$$J = \frac{\partial g_1}{\partial k} = \sum_{i=1}^{NE} (x_i^e \exp(-k x_i^e))^2 - (y_i^e - \exp(-k x_i^e)) (x_i^e)^2 \exp(-k x_i^e)$$

Assim, a equação recursiva do método de Newton-Raphson pode ser escrita na forma:

$$k_{i+1} = k_i - \frac{g_1(k_i)}{J(k_i)}$$

Partindo-se então de uma estimativa inicial para $k_0=1.000000$ (o grande número de casas decimais é proposital, para ilustrar a convergência do procedimento), os valores encontrados durante o procedimento recursivo são:

Tabela 5.1 – Resultados do procedimento iterativo do método de Newton-Raphson para uma estimativa inicial igual a $k_0=1.000000$.

iteração	k_i	F_{Obj}	g_1	J	k_{i+1}
0	1.000000	0.176161	-0.223113	0.447333	1.498763
1	1.498763	0.039183	-0.072336	0.193539	1.872518
2	1.872518	0.007354	-0.018337	0.104642	2.047757
3	2.047757	0.003855	-0.002396	0.078540	2.078265
4	2.078265	0.003780	-0.000059	0.074710	2.079054
5	2.079054	0.003780	0.000000	0.074614	2.079054

Pode ser observado que, após cinco iterações, o valor do gradiente é nulo (dentro da precisão utilizada) e os valores da função objetivo e do parâmetro k convergem respectivamente para 0.003780 e 2.079054.

O mesmo procedimento é usado agora, partindo-se de uma estimativa inicial diferente $k_0=5.000000$. A Tabela 5.2 mostra os valores encontrados ao longo do procedimento iterativo.

Tabela 5.2 – Resultados do procedimento iterativo do método de Newton-Raphson para uma estimativa inicial igual a $k_0=5.000000$.

iteração	k_i	F_{Obj}	g_1	J	k_{i+1}
0	5.000000	0.202586	0.038660	-0.003467	16.149476
1	16.149476	0.704960	0.012427	-0.001098	27.465807
2	27.465807	0.882527	0.004506	-0.000418	38.245644
3	38.245644	0.943487	0.001613	-0.000157	48.527895
4	48.527895	0.964394	0.000587	-0.000058	58.630075
5	58.630075	0.971887	0.000215	-0.000021	68.667411

Neste caso, o procedimento numérico não converge, já que o valor de k continua aumentando indefinidamente, à medida que as iterações se sucedem. O que ocorre nesse caso é que o algoritmo está evoluindo na direção de um máximo (em que k é infinitamente grande), onde a derivada da função objetivo também é nula. Quando o valor de k tende para infinito, a resposta do modelo fica sempre igual a 0, de forma que a função objetivo alcança um valor máximo (observe que ao longo das iterações o gradiente g_1 esta convergindo para zero!). Como as equações recursivas do método de Newton-Raphson não são capazes de distinguir entre um ponto de mínimo e um ponto de máximo, uma boa estimativa inicial dos parâmetros é fundamental para que a convergência para um ponto de mínimo ocorra.

Como mostrado no Exemplo 5.2, a resolução de um problema de estimação de parâmetros através da aplicação direta da condição de gradiente nulo (e uso de procedimentos numéricos derivados da estratégia de Newton-Raphson) pode ser inviável do ponto de vista prático. Em primeiro lugar, a geração do sistema de equações algébricas descrito pela Equação (5.4) pode ser difícil ou até mesmo impossível, porque

requer a derivação da função objetivo. No caso particular em que o modelo não é explícito (ou seja, em que é preciso usar um método numérico para calcular \mathbf{y}_i^m na Equação (5.3)), a definição explícita das Equações (5.4-5) pode não ser possível. Uma dificuldade adicional diz respeito ao cálculo da matriz Jacobiana (que envolve as derivadas segundas da função objetivo) e à inversão da matriz Jacobiana a cada iteração (o que é computacionalmente custoso). Finalmente, como ilustra o Exemplo 5.2, não é possível garantir a convergência do procedimento numérico, mesmo quando as derivadas podem ser calculadas analiticamente em problemas de pequena dimensão. Em outras palavras, o procedimento numérico de Newton-Raphson pode ser extremamente sensível à estimativa inicial dos parâmetros fornecida pelo usuário. Por todas essas razões, a despeito do enorme sucesso que as estratégias numéricas derivadas do procedimento de Newton-Raphson alcançaram em vários campos da matemática, é necessário desenvolver procedimentos alternativos e mais robustos para resolver numericamente o problema de estimação de parâmetros.

Na prática, a otimização da função objetivo é feita com o auxílio de métodos numéricos desenvolvidos especificamente para procedimentos de otimização. Nesse caso, ao invés de buscarmos os valores dos parâmetros que tornam o vetor gradiente da função objetivo nulo, como descrito nas Equações (5.4-5), são utilizados métodos que manipulam diretamente os valores da função objetivo. A idéia fundamental por trás desses métodos é que, ao invés de procurarmos diretamente o ponto mais baixo de um vale, devemos simplesmente tratar de descer o vale. Obviamente que, ao longo da trajetória de descida, esperamos encontrar o ponto de mínimo em algum momento.

A princípio, qualquer método de otimização pode ser utilizado para prover a solução numérica do problema de estimação de parâmetros. Entretanto, como será discutido a seguir, cada método de otimização tem suas particularidades, vantagens e desvantagens, de maneira que o sucesso na resolução do problema de estimação de parâmetros depende em grande parte da escolha de um método apropriado para o problema que está sendo resolvido. Como em problemas de estimação de parâmetros sempre se procura minimizar uma certa função objetivo, os métodos de otimização serão referenciados genericamente como métodos de minimização, para facilitar a compreensão. Além disso, salvo alguma ressalva, os métodos apresentados aqui podem ser aplicados para minimização de qualquer função objetivo, e não somente das funções objetivos resultantes dos problemas de estimação de parâmetros.

5.2 Método de Newton

O método de Newton é um método de minimização baseado em uma aproximação quadrática da função objetivo. Admita que uma função escalar qualquer $f(\boldsymbol{\alpha})$, onde $\boldsymbol{\alpha}$ é um vetor de N variáveis que devem ser manipuladas para permitir a otimização (minimização) de $f(\boldsymbol{\alpha})$, pode ser aproximada por uma série de Taylor, truncada no termo quadrático, em torno de um ponto $\boldsymbol{\alpha}^0$. Nesse caso,

$$f(\boldsymbol{\alpha}) \cong f(\boldsymbol{\alpha}^0) + (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0)^T \nabla \mathbf{f}_{\boldsymbol{\alpha}^0} + \frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0)^T \mathbf{H}_{\boldsymbol{\alpha}^0} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) \quad (5.8)$$

onde $\nabla \mathbf{f}_{\alpha^0}$ é o vetor gradiente da função (que contém as derivadas primeiras da função objetivo e tem dimensão igual a N) e \mathbf{H}_{α^0} é a matriz Hessiana (que contém as derivadas segundas da função objetivo e tem dimensão $N \times N$), ambos avaliados no ponto α^0 , como definido abaixo:

$$\nabla \mathbf{f}_{\alpha^0} = \left[\frac{\partial f}{\partial \alpha_1} \quad \dots \quad \frac{\partial f}{\partial \alpha_N} \right]^T \quad (5.9)$$

$$\mathbf{H}_{\alpha^0} = \begin{bmatrix} \frac{\partial^2 f}{\partial \alpha_1^2} & \dots & \frac{\partial^2 f}{\partial \alpha_1 \partial \alpha_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial \alpha_N \partial \alpha_1} & \dots & \frac{\partial^2 f}{\partial \alpha_N^2} \end{bmatrix} \quad (5.10)$$

Como definido anteriormente, no ponto de mínimo de $f(\alpha)$ o vetor gradiente é nulo. Assim, derivando-se a Equação (5.8) em relação a α , obtém-se a seguinte equação:

$$\nabla f(\alpha) \cong \nabla \mathbf{f}_{\alpha^0} + \mathbf{H}_{\alpha^0} (\alpha - \alpha^0) = 0 \quad (5.11)$$

que pode ser resolvida para α na forma:

$$\alpha = \alpha^0 - \mathbf{H}_{\alpha^0}^{-1} \nabla \mathbf{f}_{\alpha^0} \quad (5.12)$$

Assim, se a função que se está sendo minimizada é uma função quadrática, a aproximação definida na Equação (5.8) é exata e a Equação (5.12) fornece o valor exato de α que minimiza esta função. (Para o caso específico do problema de estimação de parâmetros, isto só ocorre quando o modelo é linear nos parâmetros e a função objetivo tem a forma quadrática resultante da hipótese de flutuações normais nas variáveis dependentes e ausência de erros nas variáveis independentes. Nesse caso particular, como amplamente discutido no Capítulo 4, o problema de estimação de parâmetros admite solução analítica.) Quando o modelo não é linear e/ou a função objetivo não tem a forma quadrática resultante da hipótese de flutuações normais, a Equação (5.12) fornece apenas uma solução aproximada para o problema de estimação, cuja acurácia depende da qualidade da aproximação quadrática. Se a aproximação quadrática for boa o suficiente, a Equação (5.12) pode ser aplicada de forma recursiva até que se atinja o valor mínimo da função objetivo, como mostra a equação recursiva abaixo:

$$\alpha^{k+1} = \alpha^k - \mathbf{H}_{\alpha^k}^{-1} \nabla \mathbf{f}_{\alpha^k} \quad (5.13a)$$

ou simplesmente

$$\alpha^{k+1} = \alpha^k - \mathbf{H}_k^{-1} (\nabla \mathbf{f})_k \quad (5.13b)$$

Assim, fornecida uma estimativa inicial α^0 , a Equação (5.13) pode ser aplicada recursivamente até que algum critério de convergência seja satisfeito, como por exemplo:

$$\|\alpha^{k+1} - \alpha^k\| < \varepsilon \quad (5.14a)$$

$$\frac{\|\alpha^{k+1} - \alpha^k\|}{\|\alpha^k\|} < \varepsilon \quad (5.14b)$$

e/ou

$$|f(\alpha^{k+1}) - f(\alpha^k)| < \varepsilon \quad (5.14c)$$

$$\left| \frac{f(\alpha^{k+1}) - f(\alpha^k)}{f(\alpha^{k+1})} \right| < \varepsilon \quad (5.14d)$$

onde ε é a precisão desejada (um valor suficientemente pequeno, como 10^{-5}). A Equação (5.14a) é uma medida da norma Euclidiana da distância entre duas aproximações sucessivas. A necessidade de satisfazer a Equação (5.14a) está associada ao fato de que é preciso identificar com precisão os parâmetros que levam a função objetivo ao valor mínimo. A Equação (5.14b) é uma medida relativa da variação da distância entre duas aproximações sucessivas. A necessidade de satisfazer a Equação (5.14b) está associada ao fato de que é preciso identificar com precisão a ordem de grandeza dos parâmetros que levam a função objetivo ao valor mínimo. A Equação (5.14c) é uma medida da variação absoluta da função objetivo em duas aproximações sucessivas. A necessidade de satisfazer a Equação (5.14c) está associada ao fato de que é preciso identificar com precisão o valor mínimo da função objetivo. A Equação (5.14d) é uma medida relativa da variação da função objetivo. A necessidade de satisfazer a Equação (5.14d) está associada ao fato de que é preciso identificar com precisão a ordem de grandeza do valor mínimo da função objetivo. Todas as Equações (5.14a-d) podem ser impostas (ou não) simultaneamente como critérios de convergência, a depender da expectativa particular do usuário em relação à qualidade do resultado final obtido. Por exemplo, se os parâmetros estimados têm ordens de grandeza muito diferentes, pode ser conveniente usar o critério de convergência definido pela Equação (5.14b). É importante ressaltar que, durante o processo iterativo, um critério particular de convergência pode ser satisfeito sem que os demais o sejam. Por isso, é importante que o usuário defina de forma apropriada o melhor critério (ou os melhores critérios) de convergência para cada problema especificado.

A grande vantagem dos métodos de Newton é a rápida convergência do procedimento recursivo, já que o método apresenta convergência quadrática. Além disso, no caso particular das funções quadráticas, a solução do problema numérico é alcançada com apenas uma iteração. O Exemplo 5.3 mostra a convergência quadrática do método de Newton.

Exemplo 5.3 – Admita que uma seqüência $\{\alpha_k\}_{k=0}^{\infty}$ converge para α^* . Admita ainda que $E_k = \alpha_k - \alpha^*$ define uma seqüência tal que

$$\lim_{k \rightarrow \infty} \frac{|E_{k+1}|}{|E_k|^p} = C$$

com $p \geq 1$ e $C \neq 0$. Diz-se então que p é a ordem de convergência da seqüência.

Considere a seqüência gerada pela Equação (5.13), que caracteriza o método de Newton. Por simplicidade, mas sem perda de generalidade, considere a Equação (5.13) escrita para uma única variável na forma

$$\alpha_{k+1} = \alpha_k - \frac{f'(\alpha_k)}{f''(\alpha_k)}$$

Assim, a função que gera os números da seqüência pode ser escrita como

$$\phi(\alpha) = \alpha - \frac{f'(\alpha)}{f''(\alpha)}$$

Expandindo a função $\phi(\alpha)$ na forma de uma série de Taylor em torno do ponto α^* e fazendo $\alpha = \alpha_k$ é possível escrever:

$$\phi(\alpha_k) = \alpha^* + (\alpha_k - \alpha^*)\phi'(\alpha^*) + (\alpha_k - \alpha^*)^2 \frac{\phi''(\alpha^*)}{2}$$

A derivada primeira de $\phi(\alpha)$ no ponto α^* é nula, já que $f'(\alpha^*) = 0$, conforme vemos abaixo:

$$\phi'(\alpha^*) = 1 - \frac{f''(\alpha^*)f''(\alpha^*) - f'''(\alpha^*)f'(\alpha^*)}{[f''(\alpha^*)]^2} = \frac{f'''(\alpha^*)f'(\alpha^*)}{[f''(\alpha^*)]^2} = 0$$

Como $\phi(\alpha_k) = \alpha_{k+1}$, a seguinte equação pode ser escrita:

$$|\alpha_{k+1} - \alpha^*| = |\alpha_k - \alpha^*|^2 \frac{\phi''(\alpha^*)}{2}$$

e ainda

$$\frac{|E_{k+1}|}{|E_k|^2} = \frac{\phi''(\alpha^*)}{2}$$

Como $\phi''(\alpha^*)$ é diferente de zero, a pode ser ainda escrita da seguinte forma:

$$\lim_{k \rightarrow \infty} \frac{|E_{k+1}|}{|E_k|^2} = \frac{\phi''(\alpha^*)}{2} = C \neq 0$$

mostrando a convergência quadrática do método de Newton.

O método de Newton apresenta uma alta eficiência, devido à convergência quadrática do procedimento recursivo. Assim, se a distância existente entre a estimativa numérica atual da solução e a solução real é da ordem de 10^{-4} , na próxima iteração será de 10^{-8} . Porém, é preciso que se observe que a convergência do método é quadrática somente nas proximidades da solução, quando a aproximação quadrática é de boa qualidade. Como a aproximação quadrática só é exata nos problemas de estimação de parâmetros quando os modelos são lineares nos parâmetros e a função objetivo é quadrática, na prática a convergência quadrática só é alcançada quando a solução do problema já está praticamente determinada. Assim, a eficiência do método é na verdade muito dependente de uma boa estimativa inicial do ponto ótimo.

A sensibilidade do desempenho do método de Newton em relação à qualidade da estimativa inicial da solução não chega a ser surpreendente. O leitor mais atento provavelmente já percebeu que as Equações (5.6b) e (5.13) são absolutamente análogas; ou seja, o método de Newton resulta em procedimento iterativo análogo ao do método de Newton-Raphson. Portanto, ambos os métodos resultam em desempenhos numéricos semelhantes. Assim, qual é exatamente a vantagem introduzida com a proposição do método de Newton, como escrito na Equação (5.13)? Além de não ser necessário derivar a função objetivo para a construção do sistema de equações, uma outra vantagem só pode ser percebida se a Equação (5.13) é modificada na forma

$$\boldsymbol{\alpha}^{k+1} = \boldsymbol{\alpha}^k - \lambda_k \mathbf{H}_k^{-1} (\nabla \mathbf{f})_k \quad (5.15)$$

onde λ_k é um parâmetro usado para **controle do passo**. Em outras palavras, a técnica de Newton-Raphson (e a Equação (5.6b)) procura o ponto onde um conjunto de equações se iguala a zero. Nesse caso, não é possível saber objetivamente ao longo do processo iterativo se a busca vai ser bem sucedida ou não. Contudo, na forma proposta pela técnica de Newton, procura-se caminhar na direção em que a função objetivo diminui. Logo, é possível saber de forma bem objetiva se a iteração foi bem sucedida ou não. Para isso, basta checar se a função objetivo diminuiu. Se a função objetivo aumentou ao longo de uma iteração, dois fatos podem ter ocorrido. O primeiro fato está relacionado ao tamanho do passo. Pode ser que o avanço tenha sido excessivo, em decorrência da aproximação quadrática não ser boa. Nesse caso, basta dar um passo um pouco menor e mais conservativo, dado que a aproximação proposta pela Equação (5.8) é sempre válida numa vizinhança suficientemente pequena da estimativa disponível. O segundo fato diz respeito à forma da função objetivo, que pode ser não convexa na região analisada; ou seja, não ter a curvatura que caracteriza a existência de um mínimo. Nesse caso, é necessário garantir que o procedimento numérico não vai caminhar para um ponto de máximo, ao invés de um ponto de mínimo. Esses dois aspectos são considerados no algoritmo apresentado a seguir.

Algoritmo 5.1 – Método de Newton com controle do passo.

Dados uma função objetivo e um conjunto de parâmetros dos quais a função objetivo depende na forma

$$F_{Obj} = F_{Obj}(\mathbf{a})$$

Dada ainda uma estimativa inicial \mathbf{a}^0 da solução procurada e um critério de convergência, é possível propor:

1- Inicializar o processo iterativo ($k=0, \lambda_0 = 1$);

2- Calcular $F_k = F_{Obj}(\mathbf{a}_k)$;

3- Calcular $\nabla F_k = \begin{bmatrix} \frac{\partial F_{Obj}(\mathbf{a})}{\partial \alpha_1} \Big|_{\mathbf{a}^k} \\ \vdots \\ \frac{\partial F_{Obj}(\mathbf{a})}{\partial \alpha_{NP}} \Big|_{\mathbf{a}^k} \end{bmatrix}$;

4- Calcular $\mathbf{H}_k = \begin{bmatrix} \frac{\partial^2 F_{Obj}(\mathbf{a})}{\partial \alpha_1 \partial \alpha_1} \Big|_{\mathbf{a}^k} & \dots & \frac{\partial^2 F_{Obj}(\mathbf{a})}{\partial \alpha_1 \partial \alpha_{NP}} \Big|_{\mathbf{a}^k} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F_{Obj}(\mathbf{a})}{\partial \alpha_1 \partial \alpha_{NP}} \Big|_{\mathbf{a}^k} & \dots & \frac{\partial^2 F_{Obj}(\mathbf{a})}{\partial \alpha_{NP} \partial \alpha_{NP}} \Big|_{\mathbf{a}^k} \end{bmatrix}$;

5- Inverter \mathbf{H}_k ;

6- Calcular $\mathbf{a}^{k+1} = \mathbf{a}^k - \lambda_k \mathbf{H}_k^{-1} \nabla F_k$;

7- Calcular $G_k = F_{Obj}(\mathbf{a}_{k+1})$;

8- Se $G_k < F_k$, a iteração foi bem sucedida;

8a- Verifica-se o critério de convergência. Se houve convergência, pare.

8b- Se não houve convergência, atualiza-se o procedimento:

$$k = k + 1;$$

$$\lambda_k = 1;$$

$$F_k = G_k;$$

8c- Retorna-se ao passo 3;

9- Se $G_k > F_k$, a iteração foi mal sucedida;

9a- Verifica-se a curvatura da aproximação quadrática, segundo a Equação (5.8):

$$\Delta F_{Obj}^{lin} = (\mathbf{a}^{k+1} - \mathbf{a}^k)^T \nabla F_k + \frac{(\mathbf{a}^{k+1} - \mathbf{a}^k)^T \mathbf{H}_k (\mathbf{a}^{k+1} - \mathbf{a}^k)}{2}$$

9b- Se $\Delta F_{Obj}^{lin} > 0$, inverte-se a direção de busca:

$$\lambda_k = -\lambda_k \text{ e volta-se ao passo 6;}$$

9c- Se $\Delta F_{Obj}^{lin} < 0$, reduz-se o tamanho do passo:

$$\lambda_k = r \lambda_k, \quad 0 < r < 1, \text{ e volta-se ao passo 6.}$$

Na forma proposta no Algoritmo 5.1, o algoritmo de Newton converge sempre. Isso não quer dizer, no entanto, que a solução obtida é o mínimo global do problema, já que a função objetivo pode ter muitos mínimos distintos. Além disso, dependendo da forma da função objetivo, o número de iterações pode ser excessivamente grande. Por

isso, é comum que os programas comerciais estabeleçam um número máximo de iterações que podem ser realizadas. Caso esse número máximo (e proibitivo) de iterações seja atingido, o usuário deve reanalisar o conjunto de estimativas iniciais utilizadas para atingir a solução do problema. Em ambos os casos, o usuário deve estar preparado para prover diferentes estimativas iniciais para o problema e nunca deve se contentar com a primeira solução obtida.

A dependência do método de Newton em relação a uma boa estimativa inicial não limita somente a eficiência do método, mas pode trazer outros transtornos para a sua aplicação. Por exemplo:

- a) Como já comentado, a função objetivo pode apresentar vários mínimos locais, presentes por causa das não-linearidades dos modelos. Assim, dependendo da estimativa inicial, a solução encontrada pode não ser a melhor solução possível, já que a minimização encontra apenas um mínimo local e não o mínimo global. Para superar esta dificuldade devem ser realizadas diversas minimizações com estimativas iniciais diferentes, para que se aumente a probabilidade de que o mínimo global seja encontrado.
- b) O segundo ponto está ligado ao cômputo das derivadas e à necessidade da inversão da matriz Hessiana. Uma estimativa inicial ruim pode gerar uma matriz Hessiana mal-condicionada, cuja inversão seja numericamente impossível, fazendo com que o procedimento iterativo seja interrompido. Esta dificuldade pode ser superada com a utilização de um método alternativo, como o método de Levenberg-Marquardt (que será discutido mais adiante), para garantir que a matriz Hessiana seja positiva definida. Isso garante a inversão da matriz Hessiana e permite que a busca prossiga na direção do valor mínimo.

Exemplo 5.4 - Considere o modelo cinético que representa uma reação de isomerização descrito por uma cinética de primeira ordem, como mostra a equação abaixo:

$$y_i = \exp\left[-(k_0 10^{17})t_i \exp\left(-\frac{E}{T_i}\right)\right]$$

onde i indica o experimento, t_i é o tempo, T_i é a temperatura, y_i é a fração que resta do reagente e k_0 e E são os parâmetros que devem ser estimados a partir dos dados experimentais, apresentados na Tabela 5.3. (Observe que de acordo com a Equação de Arrhenius, $E = \Delta E/R$) Como o valor de k_0 é da ordem de 10^{17} , foi inserida uma constante na equação do modelo, para que o parâmetro k_0 ficasse com um valor próximo de 1.

Tabela 5.3 - Dados experimentais¹ usados para a estimação de parâmetros.

Tempo [s]	Temperatura [K]	Fração de reagente	Tempo [s]	Temperatura [K]	Fração de reagente
120.0	600.0	0.900	60.0	620.0	0.802
60.0	600.0	0.949	60.0	620.0	0.802
60.0	612.0	0.886	60.0	620.0	0.804
120.0	612.0	0.785	60.0	620.0	0.794

¹ Fonte: Srinivasan e Levi (1963), J. Amer. Chem. Soc, 85, 3363-3365.

120.0	612.0	0.791	60.0	620.0	0.804
60.0	612.0	0.890	60.0	620.0	0.799
60.0	620.0	0.787	30.0	631.0	0.764
30.0	620.0	0.877	45.1	631.0	0.688
15.0	620.0	0.938	30.0	631.0	0.717
60.0	620.0	0.782	30.0	631.0	0.802
45.1	620.0	0.827	45.0	631.0	0.695
90.0	620.0	0.696	15.0	639.0	0.808
150.0	620.0	0.582	30.0	639.0	0.655
60.0	620.0	0.795	90.0	639.0	0.309
60.0	620.0	0.800	25.0	639.0	0.689
60.0	620.0	0.790	60.1	639.0	0.437
30.0	620.0	0.883	60.0	639.0	0.425
90.0	620.0	0.712	30.0	639.0	0.638
150.0	620.0	0.576	30.0	639.0	0.659
90.4	620.0	0.715	60.0	639.0	0.449
120.0	620.0	0.673			

O vetor gradiente (Equação 5.9) e a matriz Hessiana (Equação 5.10) foram calculadas analiticamente. Partindo-se de estimativas iniciais iguais a 0.6000 e 25000 para os parâmetros k_0 e E , respectivamente, a utilização da Equação (5.15) gera a seqüência de valores apresentados na Tabela 5.4.

Tabela 5.4 - Seqüência obtida pelo método de Newton ao longo das iterações.

iteração	k_0 [s^{-1}]	E [K]	F_{Obj} ($\times 10^{-2}$)	λ
0	0.6000	25000.0	2276.56	----
1	0.6006	25263.8	2224.46	-1.00
2	0.6035	25595.0	2063.93	-1.00
3	0.6084	25994.4	1608.54	-1.00
4	0.6151	26875.3	231.708	-1.00
5	0.6173	27310.4	9.54981	+1.00
6	0.6176	27418.5	1.14302	+1.00
7	0.6155	27430.6	1.03352	+1.00
8	0.6569	27472.9	1.03217	-0.06
9	0.6779	27491.8	1.03081	+1.00
10	0.7273	27537.2	1.03065	+1.00
11	0.7379	27544.8	1.02912	+1.00
12	0.7703	27572.4	1.02875	+0.44
13	0.7869	27585.3	1.02835	+1.00
14	0.8222	27613.4	1.02823	+1.00
15	0.8299	27618.6	1.02800	+1.00
16	0.8561	27638.4	1.02798	+1.00
17	0.8581	27639.6	1.02793	+1.00
18	0.8622	27642.6	1.02793	+1.00
19	0.8623	27642.7	1.02793	+1.00

A coluna λ indica o valor do passo em cada iteração. Pode ser observado que nas primeiras iterações o passo teve que ser invertido, já que a estimativa inicial estava em

uma região côncava e quase plana, como pode ser observado nas Figuras 5.1 e 5.2. Nas iterações 8 e 14 o passo teve que ser diminuído, para evitar que a função objetivo aumentasse de uma iteração para outra por causa do tamanho excessivo do passo. Nas Figuras 5.1 e 5.2 também pode ser observado que o mínimo se localiza em um vale estreito cercado por duas regiões aproximadamente planas.

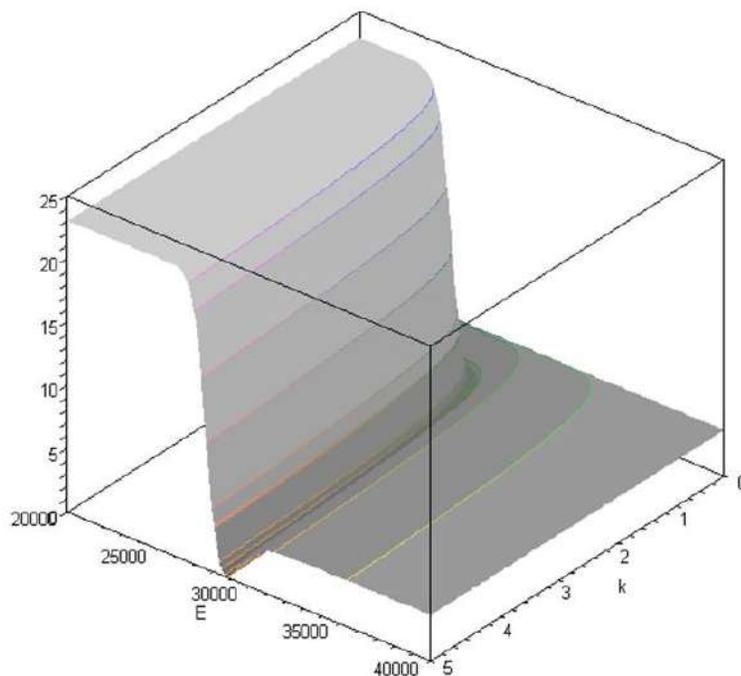


Figura 5.1 - Gráfico da função objetivo como função dos parâmetros.

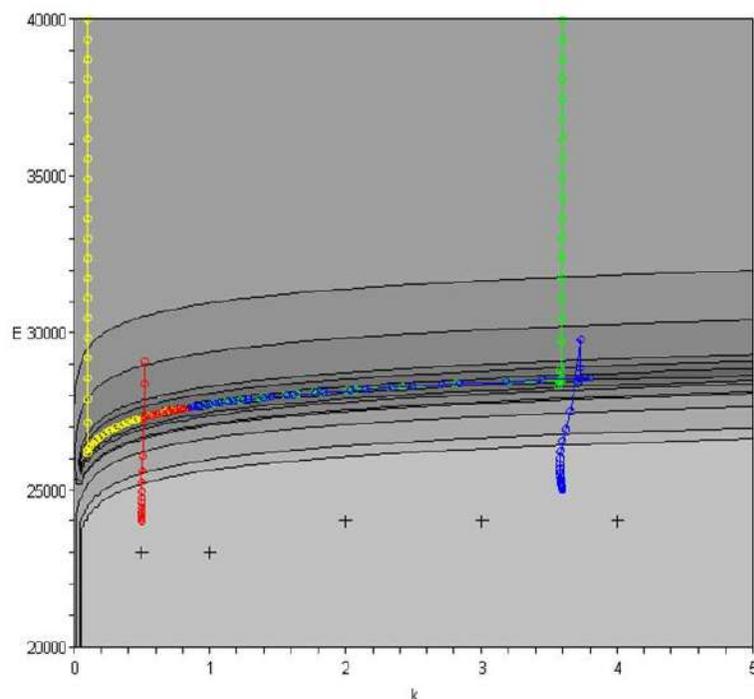


Figura 5.2 - Gráfico de contorno da função objetivo como função dos parâmetros, mostrando os caminhos percorridos pelo método de Newton para diferentes estimativas iniciais: amarelo = $[0.1; 40000]$; verde = $[3.6; 40000]$; vermelho = $[0.5; 24000]$; azul = $[3.6; 25000]$; os símbolos + indicam as estimativas iniciais que levaram a matrizes Hessianas mal condicionadas (numericamente não inversíveis).

Na Figura 5.2 são apresentados os caminhos percorridos por diversas buscas iniciadas em pontos diferentes. Observe que em todos os casos a busca inicialmente altera apenas o parâmetro E , até que o vale seja encontrado. A busca então prossegue pelo vale até encontrar o ponto de mínimo. Observe que os pontos marcados com + são estimativas iniciais que não convergem para o mínimo. Em todos esses casos, as estimativas iniciais levam a regiões onde a função objetivo é (praticamente) constante, o que faz com que a matriz Hessiana não admita inversa (a função não é quadrática), impedindo que a busca evolua.

Além da dependência com a estimativa inicial, uma dificuldade adicional para a aplicação do método de Newton é o cálculo das derivadas da função objetivo, necessárias para compor o vetor gradiente e a matriz Hessiana. O desenvolvimento das expressões analíticas para cômputo das derivadas de primeira e de segunda ordem pode ser muito trabalhoso, além de ter que ser feito a cada vez que um novo modelo é usado. Por isso, essas derivadas são frequentemente calculadas numericamente, fazendo-se pequenas perturbações nos parâmetros, como mostrado nos algoritmos descritos a seguir.

Algoritmo 5.2 – Cômputo numérico do vetor gradiente da função objetivo.

Dados uma função objetivo e um conjunto de parâmetros dos quais a função objetivo depende na forma

$$F_{Obj} = F_{Obj}(\mathbf{a})$$

Dada ainda uma estimativa \mathbf{a}^k da solução procurada e uma perturbação suficientemente pequena dos parâmetros (tipicamente da ordem de $\delta=10^{-4}$ em valor absoluto ou relativo, caso os parâmetros sejam muito grandes):

1- Para cada um dos parâmetros ($i=1 \dots NP$);

1a- Perturbar o i-ésimo parâmetro para a frente:

$$\alpha_i^k = \alpha_i^k + \delta_i;$$

1b- Calcular a função objetivo:

$$F_1 = F_{Obj}(\mathbf{a}_k)$$

1c- Perturbar o i-ésimo parâmetro para trás:

$$\alpha_i^k = \alpha_i^k - 2 \delta_i;$$

1d- Calcular a função objetivo:

$$F_2 = F_{Obj}(\mathbf{a}_k)$$

1e- Remover a perturbação do i-ésimo parâmetro:

$$\alpha_i^k = \alpha_i^k + \delta_i;$$

1f- Calcular o i-ésimo componente do vetor gradiente:

$$\frac{\partial F_{Obj}}{\partial \alpha_i} \approx \frac{F_1 - F_2}{2 \delta_i}$$

Algoritmo 5.3 – Cômputo numérico da matriz Hessiana da função objetivo.

Dados uma função objetivo e um conjunto de parâmetros dos quais a função objetivo depende na forma

$$F_{Obj} = F_{Obj}(\mathbf{a})$$

Dada ainda uma estimativa \mathbf{a}^k da solução procurada e uma perturbação suficientemente pequena dos parâmetros (tipicamente da ordem de $\delta=10^{-4}$ em valor absoluto ou relativo, caso os parâmetros sejam muito grandes):

1- Para cada um dos parâmetros ($i=1 \dots NP$);

2- Para cada um dos parâmetros ($j=1 \dots NP$);

Se ($i \neq j$) são parâmetros diferentes:

2a- Perturbar o i-ésimo parâmetro para a frente:

$$\alpha_i^k = \alpha_i^k + \delta_i;$$

2b- Perturbar o j-ésimo parâmetro para a frente:

$$\alpha_j^k = \alpha_j^k + \delta_j;$$

2c- Calcular a função objetivo:

$$F_1 = F_{Obj}(\mathbf{a}_k)$$

2d- Perturbar o j-ésimo parâmetro para trás:

$$\alpha_j^k = \alpha_j^k - 2 \delta_j;$$

2e- Calcular a função objetivo:

$$F_2 = F_{Obj}(\alpha_k)$$

2f- Perturbar o i-ésimo parâmetro para trás:

$$\alpha_i^k = \alpha_i^k - 2 \delta_i;$$

2g- Perturbar o j-ésimo parâmetro para a frente:

$$\alpha_j^k = \alpha_j^k + 2 \delta_j;$$

2h- Calcular a função objetivo:

$$F_3 = F_{Obj}(\alpha_k)$$

2i- Perturbar o j-ésimo parâmetro para trás:

$$\alpha_j^k = \alpha_j^k - 2 \delta_j;$$

2j- Calcular a função objetivo:

$$F_4 = F_{Obj}(\alpha_k)$$

2k- Remover as perturbações dos parâmetros:

$$\alpha_i^k = \alpha_i^k + \delta_i;$$

$$\alpha_j^k = \alpha_j^k + \delta_j;$$

2l- Calcular o componente i,j da matriz Hessiana:

$$\frac{\partial^2 F_{Obj}}{\partial \alpha_i \partial \alpha_j} \approx \frac{\left(\frac{F_1 - F_2}{2 \delta_j} \right) - \left(\frac{F_3 - F_4}{2 \delta_j} \right)}{2 \delta_i}$$

Se ($i = j$) é o mesmo parâmetro:

2m- Perturbar o i-ésimo parâmetro para a frente:

$$\alpha_i^k = \alpha_i^k + \delta_i;$$

2n- Calcular a função objetivo:

$$F_1 = F_{Obj}(\alpha_k)$$

2o- Perturbar o i-ésimo parâmetro para trás:

$$\alpha_i^k = \alpha_i^k - 2 \delta_i;$$

2p- Calcular a função objetivo:

$$F_2 = F_{Obj}(\alpha_k)$$

2q- Remover a perturbação do parâmetro:

$$\alpha_i^k = \alpha_i^k + \delta_i;$$

2r- Calcular a função objetivo:

$$F_3 = F_{Obj}(\alpha_k)$$

2s- Calcular o componente i,i da matriz Hessiana:

$$\frac{\partial^2 F_{Obj}}{\partial \alpha_i \partial \alpha_i} \approx \frac{\left(\frac{F_1 - F_3}{\delta_i} \right) - \left(\frac{F_3 - F_2}{\delta_i} \right)}{\delta_i}$$

Embora o cálculo numérico das derivadas da função objetivo possa ser feito facilmente no computador, como mostram os Algoritmos 5.2 e 5.3 apresentados acima, deve ficar claro que o cálculo das derivadas pode ser muito dispendioso. Por exemplo, o cálculo das derivadas primeiras exige que a função objetivo seja calculada $2*NP$ vezes, o que pode ser proibitivo em problemas de grande dimensão. Da mesma forma, o cálculo das derivadas segundas exige que a função objetivo seja chamada $4*NP*(NP-1) + 3*NP$ vezes, o que pode tornar o processo proibitivo. (Para o cálculo acima, já se levou em conta que a matriz Hessiana é sempre simétrica, de maneira que parte dos cálculos pode ser evitado, já que $H_{ij} = H_{ji}$.) Além disso, quando o modelo da Equação (5.3) é calculado numericamente, o cálculo numérico das derivadas impõe que o modelo seja calculado com precisão ainda maior, para que os erros de aproximação não sejam acumulados excessivamente e prejudiquem o andamento do procedimento numérico. Isso também pode tornar o cálculo computacional demasiadamente dispendioso. Portanto, parece haver motivação para propor modificações do método de Newton, visando à redução da complexidade algorítmica induzida pela necessidade de calcular derivadas durante o procedimento numérico.

5.3 Método de Gauss-Newton

É comum encontrar na literatura de estimação de parâmetros referências ao método de Gauss-Newton, que nada mais é que do que o método de Newton, quando se utiliza a aproximação de Gauss para a matriz Hessiana. Esta aproximação foi discutida com detalhes no Capítulo 4 (ver Equação (4.77)) e será lembrada aqui por causa da sua importância numérica.

Considere a função objetivo definida na Equação (5.2) e reescrita abaixo:

$$F_{Obj} = \sum_{i=1}^{NE} (\mathbf{y}_i^e - \mathbf{y}_i)^T \mathbf{V}_{\mathbf{y}_i}^{-1} (\mathbf{y}_i^e - \mathbf{y}_i) \quad (5.2)$$

onde, por simplicidade, não foram considerados os desvios nas variáveis independentes. O vetor gradiente desta função em relação aos parâmetros pode ser dada na forma:

$$\nabla_{\mathbf{a}} F_{Obj} = -2 \sum_{i=1}^{NE} \left(\frac{\partial \mathbf{y}_i}{\partial \mathbf{a}} \right)^T \mathbf{V}_{\mathbf{y}_i}^{-1} (\mathbf{y}_i^e - \mathbf{y}_i) \quad (5.16)$$

sendo a matriz Hessiana definida como

$$H_{jk} = 2 \sum_{i=1}^{NE} \left\{ \left(\frac{\partial \mathbf{y}_i}{\partial \alpha_j} \right)^T \mathbf{V}_{\mathbf{y}_i}^{-1} \left(\frac{\partial \mathbf{y}_i}{\partial \alpha_k} \right) - \left(\frac{\partial^2 \mathbf{y}_i}{\partial \alpha_j \partial \alpha_k} \right)^T \mathbf{V}_{\mathbf{y}_i}^{-1} (\mathbf{y}_i^e - \mathbf{y}_i) \right\} \quad (5.17)$$

A aproximação de Gauss consiste em admitir que o segundo termo do somatório da Equação (5.17) é desprezível e se aproxima de zero, em conformidade com as hipóteses de que o experimento é bem feito e de que o modelo é perfeito, de maneira que os desvios entre as predições do modelo e os dados experimentais são pequenos. Assim, a matriz Hessiana pode ser calculada da seguinte forma simplificada:

$$\mathbf{H}_a = 2 \sum_{i=1}^{NE} \left(\frac{\partial \mathbf{y}_i}{\partial \boldsymbol{\alpha}} \right)^T \mathbf{V}_{\mathbf{y}_i}^{-1} \left(\frac{\partial \mathbf{y}_i}{\partial \boldsymbol{\alpha}} \right) \quad (5.18)$$

A grande vantagem desta aproximação está associada ao fato de que a matriz das derivadas segundas da função objetivo pode ser obtida apenas como função das derivadas primeiras do modelo em relação aos parâmetros. Portanto, essa aproximação permite grande economia computacional. É muito importante observar, contudo, que a aproximação de Gauss só é válida para procedimentos de estimação de parâmetros baseados na função objetivo definida na Equação (5.2), não podendo ser aplicada em outros procedimentos genéricos de otimização.

Quando o modelo é linear nos parâmetros, a Equação (5.18) é exata, já que as derivadas segundas do modelo com relação aos parâmetros são nulas. Portanto, quanto melhor é a aproximação linear de um modelo não linear, melhor é a aproximação de Gauss. De forma similar, quanto mais próximas as predições do modelo estão dos dados experimentais, melhor é a aproximação de Gauss. Portanto, a aproximação de Gauss é boa para bons modelos, independentemente da natureza não linear dos procedimentos numéricos.

Observe que, no início da minimização, as diferenças entre as predições do modelo e os dados experimentais podem ser grandes. À medida que a minimização se aproxima do ponto ótimo, as diferenças entre as predições do modelo e os dados experimentais diminuem, melhorando a qualidade da aproximação de Gauss. Este fato aumenta a dependência do método de Gauss-Newton em relação a uma boa estimativa inicial. Se a estimativa inicial é ruim, o segundo termo do lado direito da Equação (5.18) pode não ser desprezível. Portanto, uma boa estimativa inicial para os parâmetros é fundamental para o sucesso da estimação, como pode ser observado no Exemplo 5.5.

Exemplo 5.5 - O problema de estimação do Exemplo 5.4 foi resolvido com o método de Gauss-Newton e os novos resultados obtidos são comparados com aqueles obtidos anteriormente com o método de Newton.

Para começar, é importante dizer que todas as estimativas iniciais usadas no Exemplo 5.4 (apresentadas na Tabela 5.4 e na Figura 5.2) não permitiram a estimação bem sucedida dos parâmetros com o método de Gauss-Newton. Apenas estimativas iniciais mais próximas da região do mínimo possibilitaram a estimação dos parâmetros com o método de Gauss-Newton, mostrando a maior dependência desse método em relação à disponibilidade de uma boa estimativa inicial.

Na Tabela 5.5 é feita uma comparação entre os resultados obtidos com os métodos de Newton e de Gauss-Newton, quando ambos partem de uma mesma estimativa inicial. Observa-se que os procedimentos evoluem de forma distinta, já que as matrizes Hessianas não são calculadas exatamente da mesma forma. Os dados apresentados na Tabela 5.5 também são apresentados na Figura 5.3, onde se vê a seqüência obtida com o método de Newton na cor verde e a seqüência obtida com o método de Gauss-Newton na cor amarela. Deve ser observado que a busca realizada com o método de Gauss-Newton se afasta do ponto de mínimo na primeira iteração, mostrando que a direção foi mal calculada, em função da pior qualidade da matriz

Hessiana. Isso explica também porque o método de Gauss-Newton realiza maior número de iterações (11) que o método de Newton (8).

Utilizando uma estimativa inicial diferente (cor azul para o método de Newton e cor vermelha para o método de Gauss-Newton), o método de Gauss-Newton precisa agora de 14 iterações para chegar ao ótimo, enquanto o método de Newton precisa de 31 iterações para encontrar os parâmetros do modelo. Observe que o método de Newton segue na direção do vale, para depois seguir na direção do mínimo (cor azul). Já o método de Gauss-Newton dá um grande passo, indo parar fortuitamente dentro do vale, em uma posição bem próxima do mínimo. Esse caráter fortuito de desempenho dos métodos numéricos é creditado freqüentemente ao melhor desempenho de uma ou outra técnica de estimação, o que nem sempre é verdade. Deve ficar claro que, em geral, o método de Newton permite que se atinja a solução procurada de forma mais eficiente e em menor número de iterações, embora as iterações possam demorar mais por causa da necessidade de calcular as derivadas segundas da função objetivo, que compõem a matriz Hessiana.

Tabela 5.5 - Comparação das seqüências obtidas pelos métodos de Newton e de Gauss-Newton ao longo das iterações.

iteração	Método de Newton			Método de Gauss-Newton		
	k_0 [s ⁻¹]	E [K]	F_{Obj} (x10 ²)	k_0 [s ⁻¹]	E [K]	F_{Obj} (x10 ²)
0	0.8000	27000.0	270.759	0.8000	27000.0	270.759
1	0.8033	27473.8	9.80078	0.1323	26578.4	7.25781
2	0.8039	27583.7	1.14325	0.1636	26711.6	7.19324
3	0.8034	27597.9	1.02821	0.1974	26825.6	6.78987
4	0.8412	27627.4	1.02800	0.2507	26973.0	6.47109
5	0.8482	27632.3	1.02794	0.3334	27147.6	6.00706
6	0.8603	27641.2	1.02793	0.5145	27418.5	5.81509
7	0.8617	27642.2	1.02793	0.6929	27532.5	1.40441
8	0.8623	27642.7	1.02793	0.8288	27628.4	1.08588
9	---	---	---	0.8554	27638.0	1.02800
10	---	---	---	0.8621	27642.5	1.02793
11	---	---	---	0.8623	27642.7	1.02793

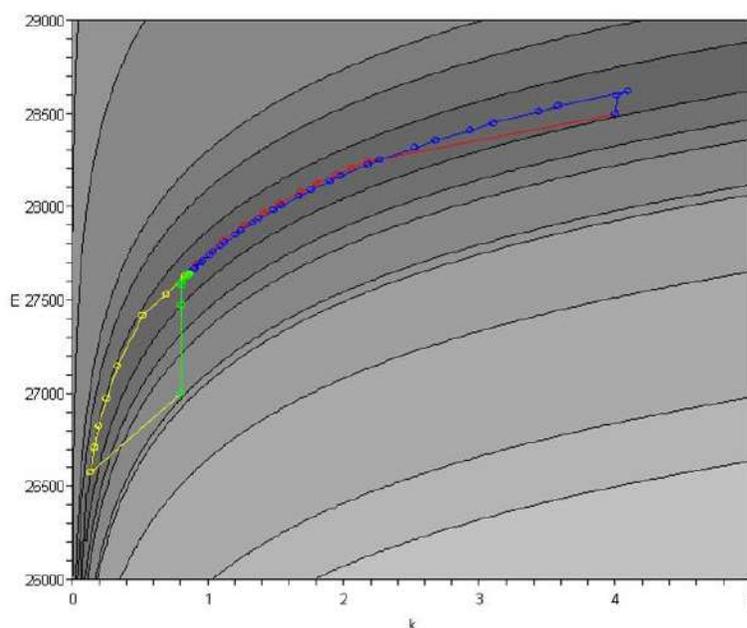


Figura 5.3 - Gráfico de contorno da função objetivo como função dos parâmetros, mostrando os caminhos percorridos pelo método de Gauss-Newton (amarelo e vermelho) e pelo método de Newton (verde e azul) para duas estimativas iniciais diferentes: [0.8; 27000] e [4.0; 28500].

5.4 Outros Métodos com Derivadas

Como discutido acima, os métodos de Newton são caracterizados pelo uso das derivadas primeira e segunda da função objetivo durante o procedimento iterativo. Esses métodos requerem a inversão da matriz Hessiana ao longo das iterações, o que pode ser computacionalmente ineficiente em problemas de grande porte. Outra classe de métodos que utilizam derivadas são os métodos de gradiente, os quais usam apenas a informação disponível na derivada primeira da função objetivo (isto é, o vetor gradiente que dá nome ao método). Como o vetor gradiente indica a direção e o sentido em que a função aumenta de forma mais íngreme, a busca numérica deve seguir o sentido contrário a apontado pelo vetor gradiente, para que a função seja minimizada. Isso permite que se escreva o seguinte procedimento iterativo:

$$\mathbf{a}^{k+1} = \mathbf{a}^k - \lambda_k \nabla \mathbf{f}_{\mathbf{a}^k} \quad (5.19)$$

onde λ é um escalar que define o tamanho passo que será dado ao longo da direção determinada pelo vetor gradiente. Comparando-se a Equação (5.19) com a Equação (5.13), conclui-se que a técnica de gradiente pode ser interpretada como uma técnica de Newton em que se aproxima a matriz Hessiana (de forma grosseira) pela matriz identidade. Contudo, diferentemente do método de Newton, que propõe que uma boa aproximação para o valor de λ é o valor 1, no método do gradiente o valor do passo

deve ser definido de forma heurística a cada iteração. De forma geral, o método do gradiente pode ser escrito na forma apresentada no Algoritmo 5.4.

Algoritmo 5.4 – Método do gradiente com controle do passo.

Dados uma função objetivo e um conjunto de parâmetros dos quais a função objetivo depende na forma

$$F_{Obj} = F_{Obj}(\mathbf{a})$$

Dada ainda uma estimativa inicial \mathbf{a}^0 da solução procurada e um critério de convergência, é possível propor:

1- Inicializar o processo iterativo ($k=0, \lambda_0 = 1$);

2- Calcular $F_k = F_{Obj}(\mathbf{a}_k)$;

3- Calcular $\nabla F_k = \begin{bmatrix} \left. \frac{\partial F_{Obj}(\mathbf{a})}{\partial \alpha_1} \right|_{\mathbf{a}^k} \\ \vdots \\ \left. \frac{\partial F_{Obj}(\mathbf{a})}{\partial \alpha_{NP}} \right|_{\mathbf{a}^k} \end{bmatrix}$;

4- Calcular $\mathbf{a}^{k+1} = \mathbf{a}^k - \lambda_k \nabla F_k$;

5- Calcular $G_k = F_{Obj}(\mathbf{a}_{k+1})$;

6- Se $G_k < F_k$, a iteração foi bem sucedida;

6a- Verifica-se o critério de convergência. Se houve convergência, pare.

6b- Se não houve convergência, atualiza-se o procedimento:

$$k = k + 1;$$

$$\lambda_k = 1/r, \quad 0 < r < 1, \text{ para acelerar a busca;}$$

$$F_k = G_k;$$

6c- Retorna-se ao passo 3;

7- Se $G_k > F_k$, a iteração foi mal sucedida;

7a- Reduz-se o tamanho do passo:

$$\lambda_k = r \lambda_k, \quad 0 < r < 1, \text{ e volta-se ao passo 4.}$$

Na forma proposta no Algoritmo 5.4, o método do gradiente converge sempre, embora isso não signifique que seja possível obter a solução em tempo hábil. Por isso, como discutido anteriormente, é comum que os pacotes comerciais imponham um número máximo admissível de iterações. Caso o número máximo de iterações seja atingido, o usuário deve refletir sobre a qualidade das estimativas iniciais fornecidas e sobre a forma da função objetivo proposta.

O método do gradiente tem a vantagem de evitar o cômputo da matriz Hessiana e a sua inversão, o que torna as iterações mais rápidas. Entretanto, como os passos são dados com um conjunto menor de informações sobre a forma da função objetivo, o

número de iterações necessárias para que o mínimo seja alcançado é usualmente maior do que nos métodos de Newton e Gauss-Newton. Além disso, a eficiência do método é dependente do controle do valor de λ . Como no caso do método de Newton, a disponibilidade de uma boa estimativa inicial para os parâmetros é um ponto importante para o método do gradiente.

Exemplo 5.6 - Como no Exemplo 5.3, a convergência do método do gradiente é analisada a seguir. Inicialmente, considera-se a Equação (5.19) escrita para uma única variável na forma:

$$\alpha_{k+1} = \alpha_k - \lambda \cdot f'(\alpha_k)$$

Assim, a função que gera os números da seqüência pode ser escrita como

$$\phi(\alpha) = \alpha - \lambda \cdot f'(\alpha)$$

Expandindo a função $\phi(\alpha)$ em série de Taylor em torno do ponto α^* e fazendo $\alpha = \alpha_k$ encontramos:

$$\phi(\alpha_k) = \alpha^* + (\alpha_k - \alpha^*) \cdot \phi'(\alpha^*)$$

A derivada primeira de $\phi(\alpha)$ no ponto α^* é igual a:

$$\phi'(\alpha^*) = 1 - \lambda \cdot f''(\alpha^*)$$

Como $\phi(\alpha_k) = \alpha_{k+1}$, pode-se escrever:

$$|\alpha_{k+1} - \alpha^*| = |\alpha_k - \alpha^*| \cdot \phi'(\alpha^*)$$

e ainda

$$\frac{|E_{k+1}|}{|E_k|} = \phi'(\alpha^*)$$

Assim, pode-se escrever

$$\lim_{k \rightarrow \infty} \frac{|E_{k+1}|}{|E_k|} = \phi'(\alpha^*) = 1 - \lambda \cdot f''(\alpha^*) \neq 0$$

mostrando a convergência linear do método do gradiente, mais lenta que a convergência quadrática característica do método de Newton.

Apesar da convergência linear do método do gradiente (mais lenta que a convergência quadrática do método de Newton), o método do gradiente é mais robusto que o método de Newton, já que não utiliza as derivadas segundas e, dessa forma, não

apresenta problemas com a inversão da matriz Hessiana. Procurando aliar a robustez do método do gradiente com a eficiência do método de Newton, pode-se propor uma ponderação entre as duas matrizes que caracterizam os dois métodos (a matriz Hessiana e a matriz identidade). Dessa forma, é possível iniciar a busca com método do gradiente (maior robustez e menor sensibilidade à estimativa inicial) e mudar a busca progressivamente para o método de Newton (maior velocidade e eficiência) ao longo das iterações, acelerando a convergência. Este método pode ser escrito na forma:

$$\mathbf{\alpha}^{k+1} = \mathbf{\alpha}^k - \lambda_k [(1-h)\mathbf{I} + h \mathbf{H}_k]^{-1} (\nabla \mathbf{f})_k \quad (5.20a)$$

onde \mathbf{I} é a matriz identidade e h é um fator de ponderação, que muda de 1 para 0 ao longo das iterações, fazendo o procedimento de busca mudar de um método de gradiente para um método de Newton. Por exemplo, h pode mudar de forma heurística como

$$h = q^k, \quad 0 < q < 1 \quad (5.20b)$$

onde q é usualmente chamado de fator de esquecimento e controla a velocidade com que a mudança do método de gradiente para o método de Newton ocorre. (Obviamente, o Algoritmo 5.1 pode ser usado igualmente para implementar o método proposto pela Equação (5.20a-b)).

Um importante teorema da matemática, que não será provado aqui por falta de espaço, diz que a Equação (5.20a) pode ser escrita na forma genérica

$$\mathbf{\alpha}^{k+1} = \mathbf{\alpha}^k - \lambda_k \mathbf{A}_k^{-1} (\nabla \mathbf{f})_k \quad (5.20c)$$

onde \mathbf{A}_k é uma matriz positiva definida qualquer. (Isto quer dizer que todos os valores característicos da matriz são positivos \mathbf{A}_k . Uma discussão sobre valores característicos e matrizes positivas definidas é apresentada no final do Capítulo 3.) Portanto, basta que o vetor gradiente seja multiplicado por uma matriz positiva definida para que se garanta a convergência do procedimento numérico da Equação (5.20) para um mínimo. Por isso, vários métodos propostos exploram as consequências desse teorema para fins de proposição de algoritmos de minimização.

O **método de Levenberg-Marquardt** utiliza uma aproximação do tipo apresentado na Equação (5.20a), substituindo a matriz identidade por uma matriz positiva definida \mathbf{A} genérica que aproxima de alguma forma a matriz Hessiana. A matriz \mathbf{A} pode ser obtida ao longo do procedimento iterativo, à medida que a forma da função objetivo (e do seu gradiente) vai sendo revelada pelo processo de busca. Várias soluções já foram propostas para obtenção eficiente da matriz \mathbf{A} , incluindo formas que evitam a necessidade de inverter a matriz Hessiana aproximada. Esses métodos são chamados genericamente de **métodos de Quasi-Newton**, por usarem uma formulação aproximada da matriz Hessiana do método de Newton. O leitor interessado deve consultar a lista final de leituras sugeridas para obter maiores detalhes a respeito dessas técnicas.

5.5. Métodos de Busca Direta

O termo Busca Direta é utilizado para identificar os métodos que não utilizam derivadas para a minimização de uma função objetivo genérica. A minimização nesses casos é feita através da análise dos valores da função objetivo numa região de busca. A busca em geral pode ser dividida em duas etapas:

- a) Etapa de exploração: nesta etapa é determinada a direção de busca do mínimo, a partir de um ponto inicial;
- b) Etapa de progressão: nesta etapa faz-se a progressão na direção de busca, enquanto for possível observar a redução do valor da função objetivo.

Vários métodos de busca direta já foram propostos e descritos na literatura. A diferença principal entre esses métodos diz respeito quase sempre à forma com que se faz exploração da região de busca. O método de Hooke e Jeeves faz a busca ao longo dos eixos de coordenadas, enquanto o método de Rosenbrock faz a busca ao longo de direções ortogonais sugeridas pelos vetores característicos de uma aproximação da matriz Hessiana. Os métodos chamados de Simplex fazem a busca com a ajuda de figuras geométricas que exploram de forma mais eficiente o espaço (por exemplo, triângulos no plano, pirâmides no espaço tri-dimensional, etc.).

A principal vantagem destes métodos é evitar o cômputo das derivadas da função objetivo. Dessa forma, reduz-se a complexidade algorítmica, reduz-se o custo computacional associado ao cômputo de derivadas numéricas e evitam-se problemas com a inversão de matrizes. Entretanto, diversos estudos mostram que os métodos de Newton (e até mesmo o método de gradiente) com aproximação das derivadas por diferenças finitas têm eficiência muito superior à eficiência característica dos métodos de busca direta, tanto no que diz respeito à precisão, quanto no que diz respeito à velocidade de convergência. Além disso, o desempenho numérico dos métodos de busca direta também é fortemente dependente da disponibilidade de uma boa estimativa inicial para os parâmetros, não sendo capaz de lidar com a existência de mínimos locais de outra forma que não seja a realização de diferentes minimizações para diferentes estimativas iniciais dos parâmetros. O Algoritmo 5.5 apresentado abaixo ilustra um procedimento de busca direta para minimização da função objetivo.

Algoritmo 5.5 – Método de Hooke e Jeeves para minimização.

Dados uma função objetivo e um conjunto de parâmetros dos quais a função objetivo depende na forma

$$F_{Obj} = F_{Obj}(\mathbf{a})$$

Dada ainda uma estimativa inicial \mathbf{a}^0 da solução procurada, um critério de convergência e um conjunto de perturbações paramétricas δ^0 para realizar a busca, é possível propor:

- 1- Inicializar o processo iterativo ($k=0$, $\lambda_0 = 1$);
- 2- Calcular $F_k = F_{Obj}(\mathbf{a}_k)$;
- 3- Calcular $\mathbf{\beta}_k = \mathbf{a}_k$, $F_k^{ini} = F_{Obj}(\mathbf{a}_k)$;

4- Fase de exploração. Para cada um dos parâmetros do problema ($i=1 \dots NP$);

4a- Perturbar o i -ésimo parâmetro para a frente:

$$\alpha_i^k = \alpha_i^k + \delta_i^k$$

4b- Calcular $G_k = F_{Obj}(\mathbf{a}_k)$;

4c- Se $G_k < F_k$, a busca foi bem sucedida:

$$F_k = G_k$$

$$\boldsymbol{\beta}_k = \mathbf{a}_k$$

4d- Perturbar o i -ésimo parâmetro para trás:

$$\alpha_i^k = \alpha_i^k - 2 \delta_i^k$$

4e- Calcular $G_k = F_{Obj}(\mathbf{a}_k)$;

4f- Se $G_k < F_k$, a busca foi bem sucedida:

$$F_k = G_k$$

$$\boldsymbol{\beta}_k = \mathbf{a}_k$$

4g- Remover a perturbação do i -ésimo parâmetro:

$$\alpha_i^k = \alpha_i^k + \delta_i^k$$

5- Se $F_k < F_k^{ini}$, a exploração foi bem sucedida:

5a- Calcular a direção de busca:

$$\mathbf{v}_k = \boldsymbol{\beta}_k - \mathbf{a}_k$$

5b- Avançar na direção de busca:

$$\boldsymbol{\beta}_k = \boldsymbol{\beta}_k + \lambda_k \mathbf{v}_k$$

5c- Calcular $G_k = F_{Obj}(\boldsymbol{\beta}_k)$;

5d- Se $G_k < F_k$, a busca foi bem sucedida:

$$F_k = G_k$$

$$\lambda_k = \lambda_k / r, \quad r < 1, \text{ para acelerar a busca;}$$

Retorna ao passo 5b, para continuar avançando;

5e- Se $G_k > F_k$, a busca foi mal sucedida:

$$\mathbf{a}_{k+1} = \boldsymbol{\beta}_k - \lambda_k \mathbf{r}_k$$

$$\lambda_{k+1} = 1$$

$$k = k + 1$$

Retorna ao passo 2, para nova exploração;

6- Se $F_k > F_k^{ini}$, a exploração foi mal sucedida:

6a- Verificam-se os critérios de convergência. Se convergiu, pare;

6b- Explora-se uma região mais próxima do ponto central:

$$\boldsymbol{\delta}_k = r \boldsymbol{\delta}_k, \quad r < 1$$

Retorna-se ao passo 4 para nova exploração.

5.6. Métodos Heurísticos

Os algoritmos heurísticos de otimização são caracterizados pela realização de um grande número de avaliações da função objetivo em toda a região de busca, de

forma a aumentar a probabilidade de encontrar o ótimo global da função objetivo. Além disso, estes métodos não precisam de uma estimativa inicial muito precisa da solução e não utilizam as derivadas para chegar ao ponto ótimo, evitando assim muitas das dificuldades associadas aos métodos tradicionais. São portanto algoritmos adequados para lidar com funções objetivos fortemente não lineares e com problemas onde não estão disponíveis boas estimativas iniciais para os parâmetros. Dentre os métodos heurísticos, destacam-se o método de Monte Carlo, o Algoritmo Genético (*Genetic Algorithm*), o Algoritmo de Recozimento Simulado (*Simulated Annealing*) e o Enxame de Partículas (*Particle Swarm Optimization*). A seguir, cada um destes métodos é apresentado com detalhes. O leitor é convidado a consultar o Capítulo 2 e o Exemplo 2.10, para entender como números aleatórios podem ser facilmente gerados no computador.

5.6.1. Monte Carlo

O método de Monte Carlo é um procedimento de busca totalmente aleatório, que pode ser usado para encontrar o ponto ótimo de uma função qualquer. A busca consiste em sortear aleatoriamente um número significativamente grande de pontos numa região de busca e considerar o melhor ponto encontrado como o melhor candidato a ponto ótimo. O procedimento pode ser repetido um número arbitrariamente grande de vezes, até que se satisfaçam os critérios de convergência. É um método extremamente simples e robusto, pois é facilmente aplicado e sempre funciona, independente da estimativa inicial dos parâmetros, da existência ou não de derivadas para a função objetivo e da dimensão do sistema. É claro que a grande limitação deste método é a necessidade de um número muito grande de avaliações da função objetivo para que se possa ter uma alta probabilidade de que o ponto ótimo tenha sido encontrado com uma precisão razoável.

A geração dos pontos onde a função objetivo deve ser avaliada pode ser feita com o auxílio da seguinte equação:

$$\alpha_d = \alpha_d^L + r(\alpha_d^H - \alpha_d^L) \quad (5.21)$$

onde d indica a direção de busca, α^L e α^H são os limites inferior e superior da região de busca e r é um número aleatório com distribuição uniforme no intervalo $[0, 1]$. Uma forma de aumentar a eficiência deste método consiste em reduzir a região de busca ao longo das iterações, ao redor do melhor ponto encontrado até o momento. Esta redução deve ser lenta, para evitar que mínimos localizados próximo aos limites da região de busca sejam excluídos do procedimento de minimização. As equações utilizadas para a redução da região de busca são escritas abaixo

$$\alpha_d^H = (\alpha_d^{H_0} - \alpha_d^{oim}) \cdot (1 - TR)^k + \alpha_d^{oim} \quad (5.22a)$$

$$\alpha_d^L = (\alpha_d^{L_0} - \alpha_d^{oim}) \cdot (1 - TR)^k + \alpha_d^{oim} \quad (5.22b)$$

onde α^{L_0} e α^{H_0} são os limites da região no início da busca, α^{oim} é o ponto ótimo encontrado até a iteração k e TR controla a taxa de redução da região de busca. A inclusão do parâmetro TR reduz a robustez do método, já que o sucesso da minimização

fica dependente da escolha adequada do valor de TR , que deve ser da ordem de 1% (podendo variar de acordo com o problema que está sendo resolvido). Os outros parâmetros de busca do método de Monte Carlo são o número de iterações ($Niter$) e o número de pontos (Npt) avaliados em cada iteração. Um fluxograma ilustrativo do algoritmo do método de Monte Carlo é apresentado na Figura 5.4.

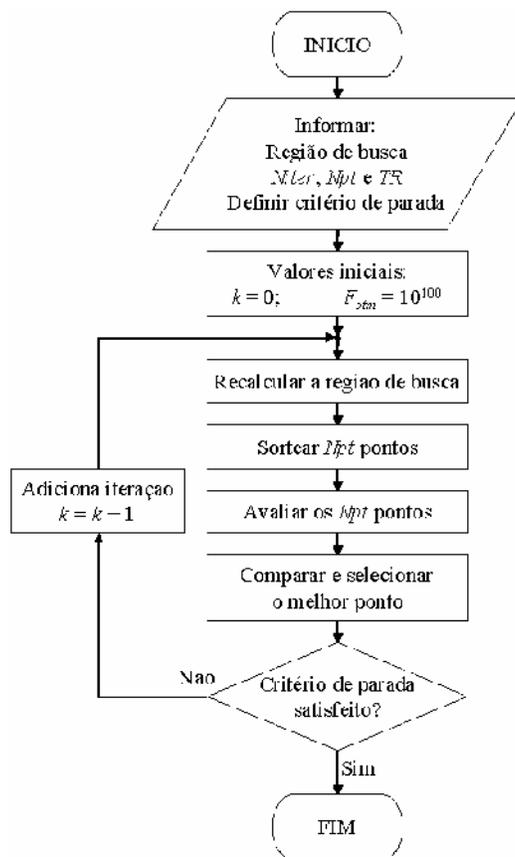


Figura 5.4 - Fluxograma esquemático do método de Monte Carlo.

Como pode ser visto na Figura 5.4, existe a necessidade da definição de um critério de parada para o algoritmo. Este critério pode ser um número máximo de iterações, um valor mínimo a ser atingido pela a função objetivo, o número de iterações sucessivas sem encontrar um novo ponto ótimo, dentre outros. Um critério de parada interessante é definir o número de iterações a partir da definição da taxa de redução. A diferença entre as Equações (5.22a) e (5.22b), quando k é igual a $Niter$, fornece a seguinte equação:

$$\Delta\alpha_f = \Delta\alpha_0 \cdot (1 - TR)^{Niter} \tag{5.23}$$

onde $\Delta\alpha_0$ e $\Delta\alpha_f$ correspondem ao tamanho inicial e final da região de busca. Assim, o número máximo de iterações $Niter$ pode ser calculado como

$$Niter = \frac{\ln(\Delta\alpha_f / \Delta\alpha_0)}{\ln(1 - TR)} \tag{5.24}$$

Portanto, é possível definir o número de iterações necessárias para que, dada uma TR , a região de busca final seja uma fração $\Delta\alpha_f/\Delta\alpha_0$ da região de busca inicial, o que dá uma idéia da precisão que se deseja obter. Algumas vezes é recomendável definir *a priori* o valor do número de iterações $Niter$, para que seja possível controlar o tempo de CPU do procedimento numérico. Dessa forma, torna-se possível calcular valor adequado da taxa de redução através da seguinte equação

$$TR = 1 - (\Delta\alpha_f / \Delta\alpha_0)^{1/Niter} \tag{5.25}$$

De qualquer forma, recomenda-se que a definição do valor de TR seja feita de forma cuidadosa. Isso é necessário porque, quando o ponto ótimo se encontra próximo aos limites da região de busca, a possibilidade deste ponto ser excluído da região de busca durante o procedimento iterativo é grande.

Exemplo 5.7 - O problema de estimação do Exemplo 5.4 foi resolvido com o método de Monte Carlo. Foram utilizadas 500 iterações e 50 avaliações aleatórias por iteração, com uma taxa de redução da região de busca de 1% por iteração, sendo a região de busca inicial delimitada no intervalo [0, 10] para o parâmetro k_0 e [0, 50000] para o parâmetro E . Os valores finais obtidos para três tentativas são apresentados na Tabela 5.6.

Tabela 5.6 - Resultados obtidos em três minimizações com o método de Monte Carlo.

Tentativa	k_0 [s ⁻¹]	E [K]	F_{Obj} (x 10 ⁻²)
1	3.0341	28433.4	1.10474
2	2.1133	28205.6	1.06702
3	1.0272	27752.6	1.02942

Do Exemplo 5.4, depreende-se que os valores ótimos dos parâmetros são iguais a 0.8623 s⁻¹ e 27642.7 K respectivamente para os parâmetros k_0 e E . O valor mínimo da função objetivo é igual a 1.02793x10⁻². Os valores obtidos em cada tentativa encontram-se próximos daqueles obtidos pelo método de Newton, mas não alcançaram a mesma precisão. Como o método de Monte Carlo é um método totalmente aleatório, um melhor desempenho pode ser alcançado com um número muito maior de avaliações da função objetivo.

Nas Figuras 5.5 e 5.6 são apresentadas as evoluções da função objetivo e dos parâmetros ao longo das iterações. Deve ser observado que a queda dos valores da função objetivo é muito significativa nas duzentas iterações iniciais. A partir daí, a queda dos valores da função objetivo e a variação dos valores dos parâmetros são muito pequenas, o que mostra que o método de Monte Carlo é eficiente para identificar a região onde o mínimo se encontra, mas tem dificuldades para obter a solução com uma boa precisão. Aumentando-se o número de avaliações por iteração de 50 para 500, foi possível encontrar a mesma solução que o método de Newton encontrou (sem problemas de definição da estimativa inicial dos parâmetros, mas com custo computacional muito mais elevado).

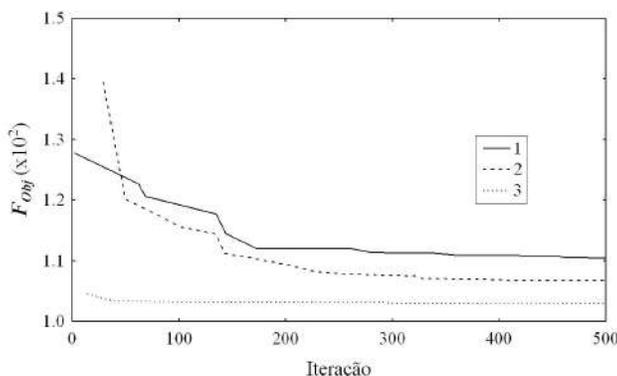


Figura 5.5 - Variação da função objetivo ao longo das iterações com o método de Monte Carlo.

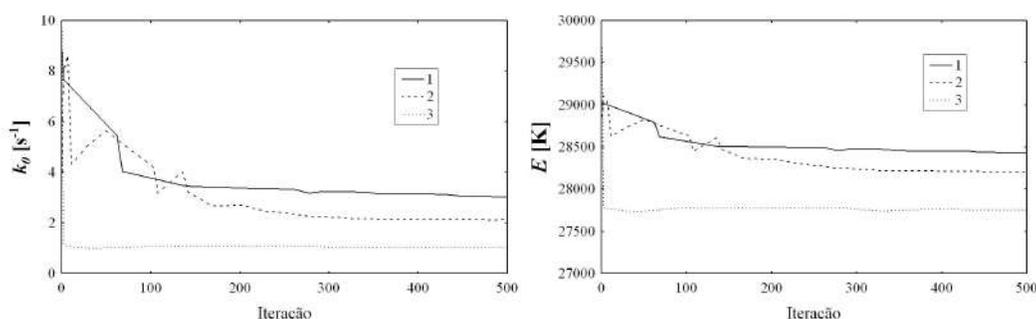


Figura 5.6 - Variação dos melhores parâmetros obtidos ao longo das iterações com o método de Monte Carlo.

5.6.2. Algoritmo Genético

O Algoritmo Genético é um método de otimização cuja proposição é inspirada na evolução dos seres vivos, onde os indivíduos mais aptos tendem a sobreviver (teoria da seleção natural de Darwin) e as características de um indivíduo são passadas para seus descendentes através do cruzamento dos indivíduos (princípios da herança genética de Mendel). A analogia com um procedimento de otimização é baseada na avaliação da aptidão de um indivíduo, feita através de uma função objetivo. Cada indivíduo (estimativa) comporta um conjunto de valores das variáveis de otimização (parâmetros) que é considerado como a carga genética deste indivíduo. Durante o cruzamento entre indivíduos pré-selecionados (estimativas diferentes), ocorre a troca de informações; isto é, a carga genética dos indivíduos é recombinada (os valores dos parâmetros são combinados). Assim, novos indivíduos (estimativas) são gerados a partir do cruzamento entre os indivíduos selecionados do grupo, de acordo com o valor da função objetivo que cada estimativa fornece. Dessa forma, indivíduos mais aptos (que levam a menores valores da função objetivo) têm maior chance de serem selecionados e realizarem o cruzamento. Por outro lado, ocasionalmente alguns indivíduos sofrem mutações (mudança aleatória); ou seja, sua carga genética (valores dos parâmetros) é alterada aleatoriamente, gerando um novo indivíduo com características diferentes (para aumentar a chance de encontrar o ótimo global). A mutação ocorre com igual probabilidade em qualquer indivíduo do grupo, independentemente de sua aptidão, possibilitando uma maior diversificação no processo de busca (e facilitando a busca do ótimo global). Tradicionalmente, a evolução da busca era creditada ao cruzamento,

enquanto a mutação evitava a convergência prematura. Porém, estudos recentes defendem a mutação com sendo o verdadeiro responsável pela evolução de todo grupo, já que promove transformações mais profundas no grupo ao gerar indivíduos completamente diferentes, que trazem novas características genéticas ao grupo de indivíduos.

A Figura 5.7 apresenta um fluxograma com as seqüências características de operações realizadas pelo algoritmo genético durante o procedimento de otimização.

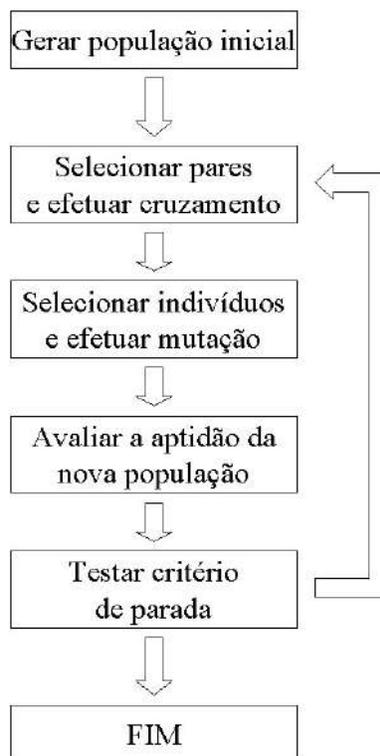


Figura 5.7 - Fluxograma das operações realizadas pelo Algoritmo Genético.

A geração da população inicial pode ser feita de diversas formas. Por exemplo, pode-se utilizar algum critério para espalhar os indivíduos uniformemente na região de busca. Entretanto, a forma mais usual consiste em gerar os indivíduos aleatoriamente na região de busca, usando para isso a Equação (5.21).

O cruzamento dos indivíduos consiste na seleção e combinação dos indivíduos para geração de novos indivíduos. A seleção é feita de forma a privilegiar os indivíduos mais aptos; isto é, aqueles que têm um menor valor da função objetivo (no caso de uma minimização). As formas mais usadas de seleção são a seleção por roleta e a seleção por torneio. A seleção por roleta consiste em atribuir uma probabilidade para cada indivíduo do conjunto, de forma que a soma das probabilidades seja igual a 1 e que a probabilidade de seleção de um indivíduo seja maior, à medida que aumenta a sua aptidão. Na Figura 5.8 pode ser observada uma forma de se obter as faixas para cada indivíduo em um problema de minimização. A seleção consiste em sortear um número aleatório com distribuição uniforme no intervalo $[0, 1]$ e selecionar o indivíduo de acordo com a faixa que contém o número sorteado. A seleção por roleta não é indicada em problemas onde um indivíduo domina a roleta (tem desempenho muito melhor que

os demais) e o sorteio acaba escolhendo sempre este mesmo indivíduo, levando o grupo todo a uma convergência prematura.

Ind	F_{Obj}	$1/F_{Obj}$	Probabilidade	Faixa
1	5.6	0.18	0.19	0 – 0.19
2	12	0.08	0.09	0.19 – 0.28
3	8	0.13	0.13	0.28 – 0.41
4	2.4	0.42	0.44	0.41 – 0.85
5	7	0.14	0.15	0.85 – 1.00
Σ	---	0.95	1.00	

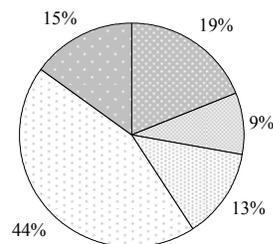


Figura 5.8 - Exemplo de seleção por roleta no Algoritmo Genético.

A seleção por torneio é uma forma extremamente simples de seleção aleatória que garante maior diversidade na seleção, independentemente da existência de indivíduos dominantes. Esta seleção consiste em escolher ao acaso dois indivíduos quaisquer do grupo e selecionar o indivíduo que tiver a melhor aptidão. Esta forma é extremamente simples e de fácil implementação computacional. Deve-se tomar certo cuidado, no entanto, uma vez que indivíduos com uma boa aptidão podem ser perdidos, já que a probabilidade de não serem escolhidos para o torneio é relativamente alta.

Uma vez selecionados os indivíduos, o cruzamento entre dois destes indivíduos é realizado, de maneira que um ou dois novos indivíduos são gerados. Entretanto, para que o cruzamento realmente ocorra, uma certa probabilidade de cruzamento deve ser satisfeita. Em termos algorítmicos, uma probabilidade de cruzamento é definida e um número aleatório é sorteado. Se o número sorteado é menor que a probabilidade de cruzamento, o cruzamento é realizado e novos indivíduos são gerados. Se o número sorteado é maior que a probabilidade de cruzamento, o cruzamento não é realizado e os indivíduos selecionados passam para a nova população sem sofrer alteração alguma.

A forma com que o cruzamento é realizado depende da codificação que está sendo utilizada para os indivíduos. Originalmente, o Algoritmo Genético utilizava a codificação binária para representar as características dos indivíduos do grupo. (Por exemplo, o número decimal 10 pode ser representado na forma binária como 1010, que significa $0 \times 2^0 + 1 \times 2^1 + 0 \times 2^2 + 1 \times 2^3$.) Assim dois indivíduos quaisquer (números) podem ser representados pelas seguintes seqüências:

P1: 1 0 1 1 0 0 1 1 1 0
 P2: 1 1 0 1 0 1 0 1 1 0

O cruzamento entre os dois indivíduos acima é realizado através da troca de parte do conjunto de informações de cada indivíduo. Uma das formas de cruzamento consiste em sortear um ou mais pontos de quebra e cruzar as informações, conforme exemplificado abaixo:

P1: 0 1 1 1 0 | 0 1 1 1 0
 P2: 1 1 0 1 0 | 1 0 1 1 0
 F1: 0 1 1 1 0 | 1 0 1 1 0
 F2: 1 1 0 1 0 | 0 1 1 1 0

Da mesma forma que o cruzamento, a mutação de um determinado indivíduo só ocorre se uma certa probabilidade de mutação for satisfeita. Assim, um número aleatório com distribuição uniforme no intervalo $[0, 1]$ é sorteado. Se este número for menor que a probabilidade de mutação, o indivíduo sofre a mutação; caso contrário, nada acontece e o indivíduo passa para a população seguinte. No cruzamento, os indivíduos mais aptos têm maior probabilidade de serem selecionados para o cruzamento. Já na mutação, todos os indivíduos são selecionados para mutação, que ocorre ou não de acordo com a probabilidade de mutação. Uma forma de realizar a mutação consiste em sortear um ou mais pontos do conjunto de informações de um indivíduo e inverter o valor deste ponto, conforme exemplificado abaixo:

F1: 0 1 1 1 0 1 0 1 1 0
M1: 0 1 0 1 0 1 1 1 1 0

Entretanto, para problemas de otimização em que as variáveis são contínuas, a utilização da codificação binária não é recomendada, já que o algoritmo de representação binária pode apresentar complexidade significativa. Um destes problemas ocorre porque a mudança de apenas um ponto do indivíduo pode levar a grandes modificações no indivíduo, o que atrapalha a convergência do algoritmo, conforme é mostrado abaixo:

11111	31	11111	31	11111	31	11111	31	11111	31
11110	30	11101	29	11011	27	10111	23	01111	15

Outro problema associado é a necessidade de discretizar as variáveis contínuas, o que pode levar a aumento considerável da dimensão do problema (quanto maior a precisão, maior o número de bits necessários para representar um indivíduo). Por fim, existe ainda a necessidade de converter os indivíduos da codificação real para codificação binária e vice-versa, o que aumenta o custo computacional do algoritmo. Portanto, a utilização da codificação decimal usual para representar o conjunto de informações de um indivíduo é mais adequada quando são utilizadas variáveis contínuas no procedimento de otimização. Dessa forma, o indivíduo (estimativa) é representado por um vetor dos valores das variáveis de otimização (parâmetros).

P1: $[\alpha_{11} \alpha_{21} \alpha_{31} \dots \alpha_{n1}]$
P2: $[\alpha_{12} \alpha_{22} \alpha_{32} \dots \alpha_{n2}]$

No caso de utilização da codificação real, o cruzamento pode ser realizado através da combinação linear dos valores de cada indivíduo, na forma:

$$\alpha_{i,Novo} = \alpha_{i,P1} + r(\alpha_{i,P2} - \alpha_{i,P1}) \tag{5.26}$$

onde i corresponde ao i -ésimo elemento dos vetores que representam os indivíduos, $\alpha_{i,Novo}$ é o novo ponto gerado do cruzamento entre $\alpha_{i,P1}$ e $\alpha_{i,P2}$, e r é um número aleatório com distribuição uniforme no intervalo $[0, 1]$.

A mutação pode ser feita através da geração de um indivíduo totalmente novo dentro da região de busca, gerado pela Equação (5.21), a mesma equação que é usada para geração dos pontos pelo método de Monte Carlo. Esta forma de mutação impede

que ocorra a convergência dos indivíduos, já que sempre são inseridos novos indivíduos na população, independentemente do estágio em que a busca se encontra. Alternativas para diminuir a probabilidade de mutação ao longo das iterações, ou realizar a mutação de apenas um dos componentes do vetor de cada indivíduo, podem ou não melhorar a eficiência do algoritmo.

Um ponto importante que deve ser levado em consideração é o balanceamento entre as frequências de cruzamento e de mutação. O cruzamento, conforme descrito pela Equação (5.26), sempre gera novos indivíduos na sub-região delimitada pelos pontos selecionados para o cruzamento. Dessa maneira, restringe-se a região onde os novos indivíduos são gerados, o que promove a aceleração da convergência dos indivíduos. Assim, é importante que a mutação seja responsável pela inserção de indivíduos novos, que não estejam “presos” a uma pequena parte da região de busca, aumentando a aleatoriedade da busca e permitindo que o algoritmo “escape” de mínimos locais e encontre o mínimo global.

Um outro ponto importante neste algoritmo é a aplicação de um operador chamado de elitismo. Este operador consiste em passar para a nova população o melhor ponto já encontrado (ou os melhores pontos já encontrados), mantendo-se a qualidade da população. Se isso não for feito, devido à aleatoriedade intrínseca do método, o melhor ponto encontrado, apesar de apresentar a maior aptidão, pode ser perdido ao longo das iterações.

Exemplo 5.8 - O problema de estimação do Exemplo 5.4 foi resolvido com o método do Algoritmo Genético. Foram utilizadas 500 iterações e 50 indivíduos, com uma probabilidade de cruzamento de 70% e uma probabilidade de mutação de 10%. A região de busca é delimitada no intervalo $[0, 10]$ para o parâmetro k_0 e $[0, 50000]$ para o parâmetro E . Os valores finais obtidos para três tentativas são apresentados na Tabela 5.7.

Tabela 5.7 - Resultados obtidos em três minimizações com o método do Algoritmo Genético.

Tentativa	k_0 [s^{-1}]	E [K]	F_{Obj} ($\times 10^{-2}$)
1	0.8773	27653.7	1.02796
2	5.4801	28805.0	1.19328
3	1.5712	28020.4	1.04576

Do Exemplo 5.4, depreende-se que os valores ótimos dos parâmetros são iguais a $0.8623 s^{-1}$ e $27642.7 K$ respectivamente para os parâmetros k_0 e E . O valor mínimo da função objetivo é igual a 1.02793×10^{-2} . Como observado com o método de Monte Carlo (Exemplo 5.7), os valores obtidos para os parâmetros após cada tentativa são próximos daqueles obtidos pelo método de Newton, mas não alcançam a mesma precisão. Apesar disso, os parâmetros de busca podem ser alterados para melhorar a eficiência da minimização. Entretanto, os valores ótimos dos parâmetros de busca, como as probabilidades de cruzamento e de mutação, dependem das características do problema que está sendo resolvido, o que torna difícil a proposição de valores absolutos que possam ser utilizados em diferentes problemas.

Nas Figuras 5.9 e 5.10 são apresentados os melhores valores dos parâmetros encontrados ao longo das iterações. Mais uma vez, observa-se no início da busca uma melhora significativa da função objetivo, sendo que a partir da centésima iteração praticamente não ocorrem alterações nem da função objetivo e nem dos parâmetros. Como no método de Monte Carlo, a região próxima ao mínimo é encontrada, mas não existe uma busca local com qualidade para que o mínimo seja definido com boa precisão.

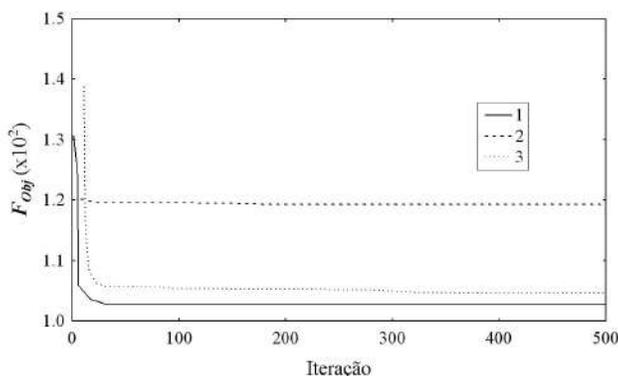


Figura 5.9 - Variação da função objetivo ao longo das iterações com o método do Algoritmo Genético.

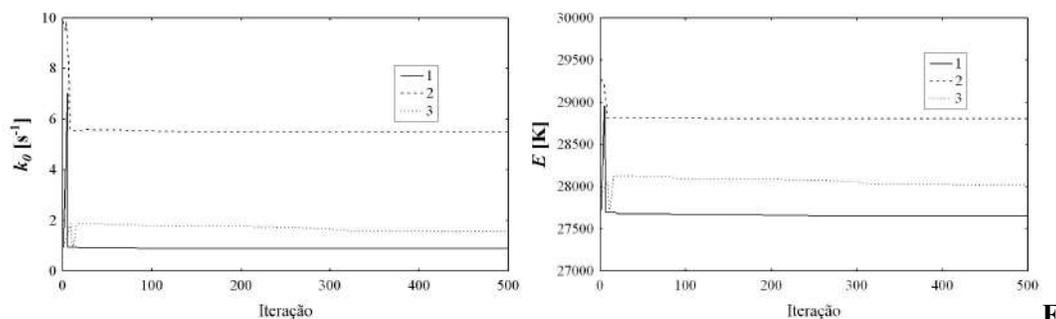


figura 5.10 - Variação dos melhores parâmetros obtidos ao longo das iterações com o método do Algoritmo Genético.

5.6.3. Recozimento Simulado

O método do Recozimento Simulado foi proposto inicialmente para problemas de otimização combinatorial, tendo sido inspirado no processo de minimização da energia contida na estrutura cristalina de um corpo metálico durante o processo de recozimento. Este processo consiste no aquecimento de um corpo metálico até uma temperatura suficientemente alta, a partir da qual ocorre o lento resfriamento do corpo metálico, possibilitando a reorganização dos átomos em uma estrutura com menor energia.

O procedimento de otimização proposto pelo algoritmo de Recozimento Simulado consiste em explorar a região de busca através de transições sucessivas (análogas às mudanças da estrutura cristalina). As transições consistem em pequenas mudanças das variáveis de otimização (parâmetros) de diferentes indivíduos

(estimativas diferentes), de acordo com o valor da função objetivo que apresentam e do valor da “temperatura”, que regula a possibilidade ou não da transição. A transição pode ser ou não ser aceita: se a transição para o novo ponto levar a um menor valor da função objetivo a mudança é aceita; caso contrário, se o valor da função objetivo no novo ponto for maior que no anterior, a transição pode ser aceita ou não, de acordo com o cálculo de uma probabilidade de transição. Esta probabilidade deve aumentar, à medida que diminui a diferença entre os valores da função objetivo nos dois pontos considerados. A forma tradicional do cálculo da probabilidade de transição é feita de acordo com a estatística de Boltzmann, que corresponde à seguinte equação:

$$p = \exp\left(-\frac{F^{k+1} - F^k}{T}\right) \quad (5.27)$$

onde p é a probabilidade de transição, F é o valor da função objetivo, k indica o valor atual, $k+1$ indica o novo valor e T é o parâmetro “temperatura”. O valor calculado de p é comparado com um número aleatório sorteado a partir de uma distribuição uniforme no intervalo $[0, 1]$. Se p é maior que este número aleatório, a transição é aceita; caso contrário, a transição não ocorre.

O valor de T deve ser alto no início, aumentando a probabilidade de ocorrerem transições para pontos onde a função objetivo aumenta, possibilitando que ocorra uma boa exploração do espaço de busca e levando o algoritmo a encontrar o mínimo global. Na forma tradicional do algoritmo, o valor de T é diminuído exponencialmente ao longo da busca, de acordo com a seguinte equação:

$$T_{n+1} = \beta T_n \quad (5.28)$$

sendo β um valor positivo menor que 1.

O primeiro ponto importante deste algoritmo consiste em definir uma estrutura de vizinhança; isto é, uma região em torno do ponto atual, para onde a transição possa ocorrer. Uma forma simples de fazer isto é definir um valor $\Delta\alpha$ tal que o novo ponto possa ser definido pela equação abaixo

$$\alpha_d^{novo} = \alpha_d^{atual} + \Delta\alpha_d (2r - 1) \quad (5.29)$$

onde d denota uma das direções de busca e r é um número aleatório com distribuição uniforme no intervalo $[0, 1]$.

Outro ponto importante deste algoritmo é a definição do valor inicial da temperatura, pois a ordem de grandeza da função objetivo pode variar muito em problemas distintos. Uma definição adequada da temperatura inicial garante uma boa exploração da região de busca no início do procedimento de minimização. Uma forma de atingir uma eficiente exploração da região de busca é definir a temperatura inicial de forma a garantir que a maior parte das transições iniciais sejam aceitas. Por exemplo, a temperatura inicial pode ser definida como:

$$T_0 = -\frac{\Delta F^*}{\ln(0.95)} \quad (5.30)$$

onde ΔF^* é a diferença entre o maior e o menor valor da função objetivo em um pequeno número de avaliações aleatórias iniciais da função objetivo. 0,95 é a probabilidade das transições ocorrerem (ou seja, no início 95% das transições serão aceitas).

A Figura 5.11 é apresentada um fluxograma onde as etapas do algoritmo de Recozimento Simulado são apresentadas de forma detalhada. É importante definir aqui que N_{pt} corresponde ao número de tentativas de transições em uma dada temperatura e NT é o número de temperaturas avaliadas. Como critério de parada, pode ser usado um número fixo de temperaturas, encerrando-se a busca assim que uma dada temperatura não leve à aceitação de qualquer transição proposta. Pode-se ainda fixar um valor mínimo a ser atingido pela função objetivo ou definir um outro critério de parada qualquer, baseado na variação dos melhores valores obtidos para os parâmetros e/ou função objetivo.

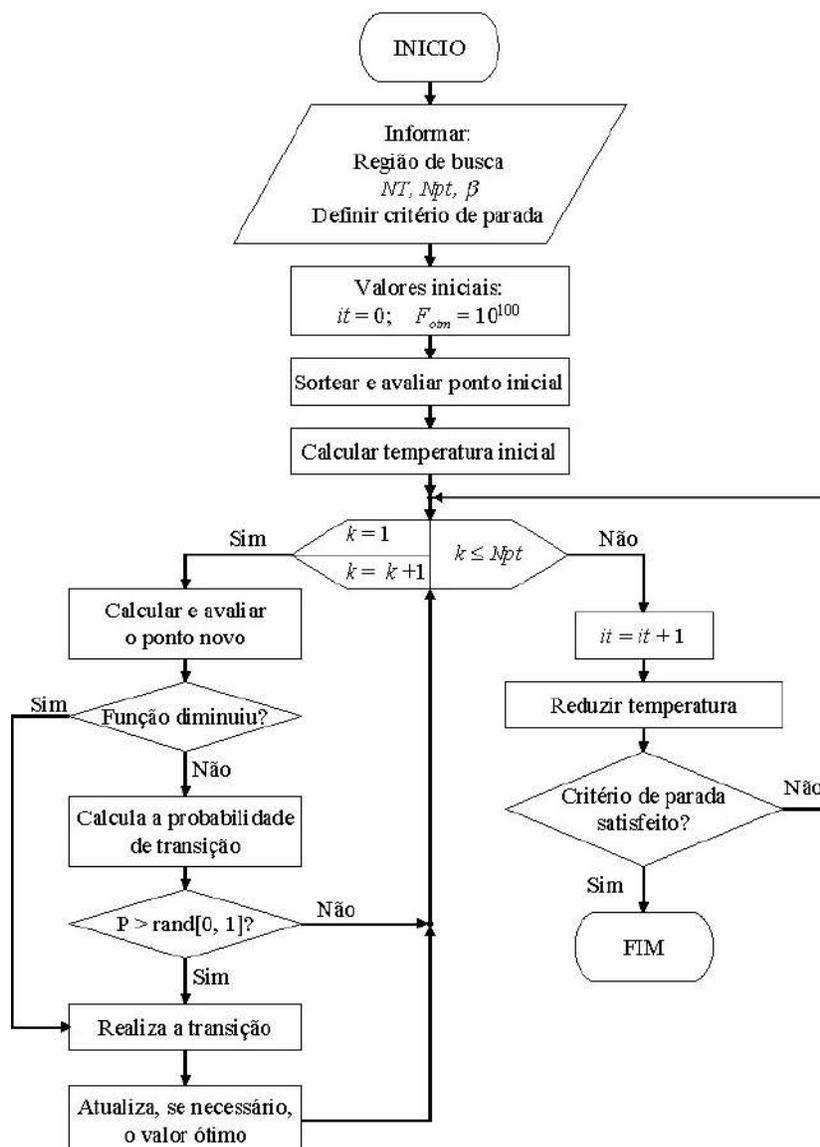


Figura 5.11 - Fluxograma do método do Recozimento Simulado.

Exemplo 5.9 - O problema de estimação do Exemplo 5.4 foi resolvido com o método do Recozimento Simulado. Foram utilizadas 500 reduções de temperatura e 50 avaliações por temperatura, com um fator de redução de temperatura de 0,90 (Equação (5.28)) e uma perturbação de 1% para o cálculo de novos pontos (Equação (5.29)). A região de busca foi delimitada no intervalo [0, 10] para o parâmetro k_0 e [0, 50000] para o parâmetro E . Os valores finais obtidos para três tentativas são apresentados na Tabela 5.8.

Tabela 5.8 - Resultados obtidos em três minimizações com o método do Recozimento Simulado.

Tentativa	k_0 [s ⁻¹]	E [K]	F_{Obj} (x 10 ⁻²)
1	0.8843	27658.3	1.0280
2	0.8383	27624.9	1.0280
3	7.7508	29023.5	1.2609

Do Exemplo 5.4, depreende-se que os valores ótimos dos parâmetros são iguais a $0,8623 \text{ s}^{-1}$ e $27642,7 \text{ K}$ respectivamente para os parâmetros k_0 e E . O valor mínimo da função objetivo é igual a $1,02793 \cdot 10^{-2}$. Nas duas primeiras tentativas o método de Recozimento Simulado foi capaz de obter o mínimo com grande precisão, o que mostra que este método tem uma maior capacidade de busca local que os dois métodos heurísticos apresentados anteriormente. Entretanto, a terceira tentativa resultou em uma solução muito ruim, mostrando também que este método tem maior dificuldade em encontrar rapidamente a região próxima do mínimo. Isto também pode ser verificado através dos resultados apresentados nas Figuras 5.12 e 5.13. Observa-se nessas figuras que somente após trezentas iterações o método se aproxima do mínimo.

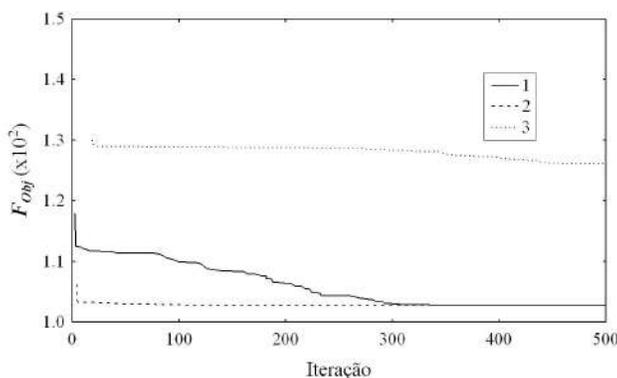


Figura 5.12 - Variação da função objetivo ao longo das iterações com o método do Recozimento Simulado.

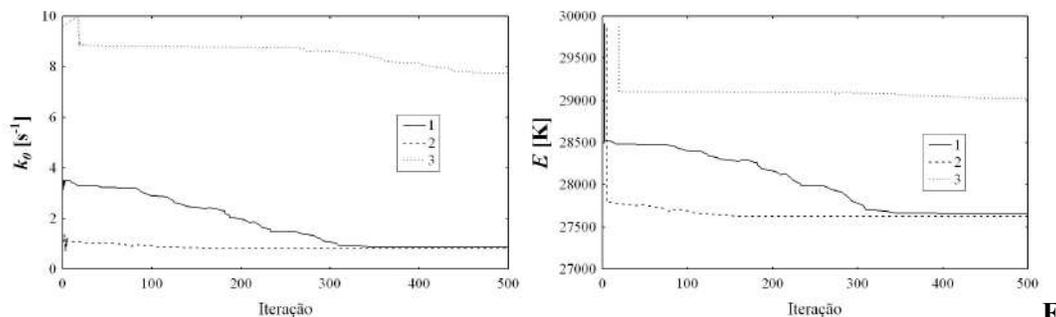


Figura 5.13 - Variação dos melhores parâmetros obtidos ao longo das iterações com o método do Recozimento Simulado.

5.6.4. Enxame de Partículas

O método do Enxame de Partículas é um algoritmo de otimização heurística inspirado no comportamento gregário de animais (peixes, pássaros, etc.). Este método consiste na otimização de uma função objetivo através da troca de informações entre os elementos ou partículas de um grupo (estimativas dos parâmetros), resultando em um algoritmo eficiente, robusto e de simples implementação computacional. Admite-se que os indivíduos (estimativas) movem-se com velocidade variável no espaço de busca (parâmetros), como aves que voam nos céus. O movimento de cada partícula em cada iteração corresponde à soma de três termos distintos: o primeiro é um termo relativo à inércia da partícula, que traduz o modo com que a que a partícula vem se movendo; o

segundo é um termo relativo à atração da partícula ao melhor ponto que ela mesma já encontrou; e o terceiro termo é relativo à atração da partícula ao melhor ponto que todo o grupo (ou uma parte do grupo) já encontrou.

O termo Enxame foi utilizado pela primeira vez em um trabalho de Milonas, que desenvolveu modelos para aplicações em vida artificial e articulou cinco princípios básicos da Inteligência de Enxames, sobre os quais o método do Enxame de Partículas é fundamentado. Os princípios são os seguintes:

- a) *Proximidade*: o enxame deve ser capaz de realizar cálculos simples de tempo e espaço. O método do Enxame de Partículas realiza uma série de cálculos no espaço n -dimensional em muitos intervalos de tempo.
- b) *Qualidade*: o enxame deve ser capaz de responder a fatores de qualidade do ambiente. O método responde a fatores de qualidade definidos pela melhor posição encontrada pelo enxame e por cada melhor posição encontrada por cada partícula.
- c) *Respostas Diversas*: o enxame não deve submeter sua atividade em meios excessivamente limitados. A definição do movimento de cada partícula do enxame em relação às melhores posições individuais e à melhor posição do enxame garante a diversidade das respostas.
- d) *Estabilidade*: o enxame não deve mudar seu comportamento a todo momento que o ambiente se altera. O comportamento do procedimento só é alterado quando ocorrem mudanças nos melhores valores encontrados.
- e) *Adaptabilidade*: o enxame deve ser capaz de alterar seu comportamento, quando o custo computacional não for proibitivo. Assim que os melhores valores são alterados, o comportamento do enxame se adapta imediatamente aos novos valores.

É interessante observar que, embora os dois últimos princípios sejam opostos, o método do Enxame assegura que ambos sejam satisfeitos simultaneamente, já que o melhor valor encontrado não é necessariamente alterado a todo instante. Quando este é alterado, o enxame se adapta a este novo valor, de forma a assegurar a estabilidade e a adaptabilidade do método.

As equações que determinam a velocidade e a posição das partículas ao longo das iterações podem ser escritas como:

$$v_{i,d}^{k+1} = wv_{i,d}^k + c_1r_1(p_{i,d}^k - \alpha_{i,d}^k) + c_2r_2(p_{global,d}^k - \alpha_{i,d}^k) \quad (5.31)$$

$$\alpha_{i,d}^{k+1} = \alpha_{i,d}^k + v_{i,d}^{k+1} \quad (5.32)$$

onde os índices k , i e d denotam, respectivamente, a iteração, a partícula e a direção de busca; v é a velocidade e α é a posição no espaço de busca (estimativas dos parâmetros); w é um peso de inércia; c_1 e c_2 são duas constantes positivas, chamadas respectivamente de parâmetro cognitivo e social; r_1 e r_2 são dois números aleatórios com distribuição uniforme no intervalo $[0, 1]$; p_i é o melhor ponto encontrado pela partícula e p_{global} é o melhor valor encontrado por todo enxame (ou parte dele).

O parâmetro w , chamado de peso de inércia (*inertia weight*) ou fator de inércia, tem o papel de balancear o caráter global e local da busca. Quanto maior o valor de w , maior é o número de iterações para que as partículas converjam. Porém, quando w é maior que 1, a convergência nunca é alcançada (ou melhor, as partículas divergem).

Assim, um valor de w próximo a 1 corresponde a uma alta inércia do movimento das partículas e mantém as partículas em movimento por um maior número de iterações, aumentando a exploração da região de busca. Um valor de w próximo a 0 diminui rapidamente a velocidade das partículas, levando a uma rápida convergência e diminuindo a exploração da região de busca.

Para tirar proveito dessa mudança de comportamento do enxame com o valor do fator de inércia, uma forma interessante de realizar a busca consiste em iniciar a busca com um valor de w próximo a 1, possibilitando uma boa exploração da região de busca, e diminuir o valor de w ao longo das iterações, forçando a convergência das partículas e aumentando a precisão do melhor ponto obtido pelo enxame. A diminuição do valor de w ao longo das iterações pode ser feita por uma infinidade de formas. A forma mais usual consiste em uma redução linear, segundo a equação abaixo:

$$w = w_i + \frac{k}{Niter}(w_f - w_i) \quad (5.33)$$

onde w_o é o valor inicial e w_f é o valor final de w e $Niter$ é o número total de iterações.

Os parâmetros c_1 e c_2 ponderam, respectivamente, a contribuição individual e a contribuição do grupo na velocidade de cada partícula. A contribuição individual corresponde a um passo na direção do melhor ponto já encontrado pela partícula em questão. Já a contribuição global consiste em um passo na direção do melhor ponto encontrado por todo o grupo ou por parte do grupo. A definição destes parâmetros permite controlar o caráter social da busca, que pode então variar de uma busca altamente individual até uma busca completamente social. Quanto maior o caráter individual da busca, maior é a exploração da região de busca, já que um maior número de iterações é necessário para que ocorra a convergência. Algum cuidado deve ser tomado com a definição dos valores de c_1 e c_2 , já que valores muito altos podem fazer com que o enxame divirja, devido aos valores muito altos atingidos pelas velocidades. Por este motivo, torna-se interessante limitar de alguma forma a velocidade e/ou a posição de cada partícula.

Pode-se observar que tanto o parâmetro w como os parâmetros c_1 e c_2 influenciam o comportamento do enxame, aumentando ou diminuindo a velocidade com que as partículas convergem. Apesar de diversos estudos já terem sido realizados em relação à definição apropriada destes parâmetros, não existe nenhum conjunto de parâmetros que seja ótimo para qualquer problema. Além disso, o caráter aleatório da busca exige que um problema seja resolvido diversas vezes (mudando ou não os valores dos parâmetros de busca), para garantir que um valor adequado seja definido como o mínimo da função objetivo em questão.

As principais vantagens deste método são a grande simplicidade das equações recursivas (permitindo a implementação de um algoritmo simples) e a robustez do procedimento de busca (devido ao caráter aleatório e global da busca). Na Figura 5.14 é apresentado um fluxograma do algoritmo do método do Enxame de Partículas.

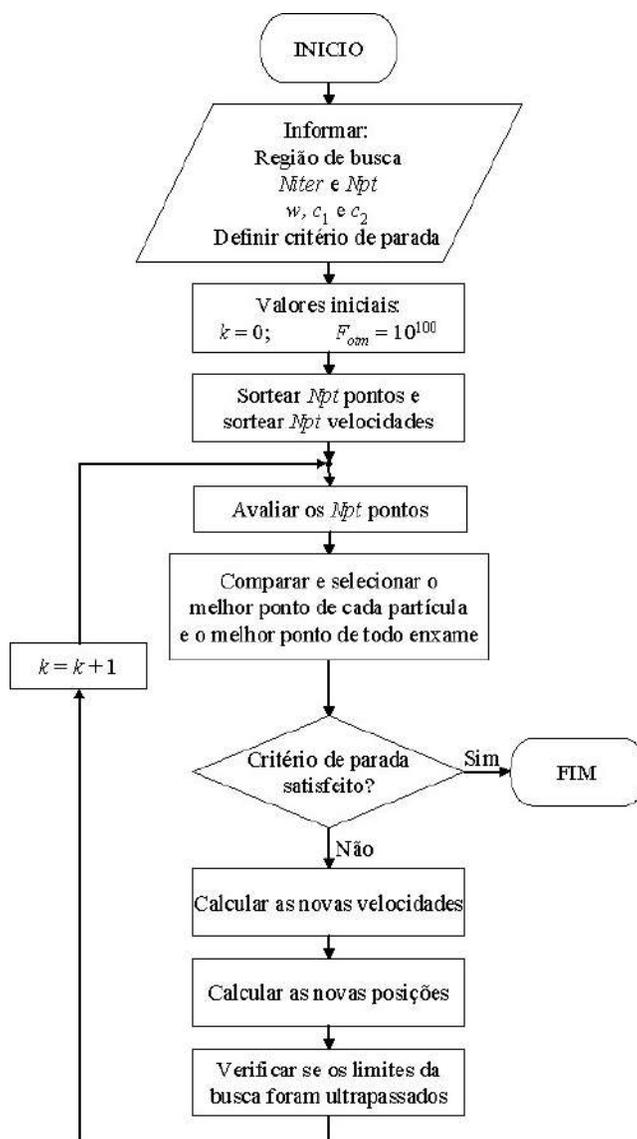


Figura 5.14 - Fluxograma do método do Enxame de Partículas.

Exemplo 5.10 - O problema de estimação do Exemplo 5.4 foi resolvido com o método do Enxame de Partículas. Foram utilizadas 500 iterações e 50 partículas, com um fator de inércia w igual a 0,75 e c_1 e c_2 iguais a 1,5. A região de busca foi delimitada no intervalo $[0, 10]$ para o parâmetro k_0 e $[0, 50000]$ para o parâmetro E . Os valores finais obtidos para três tentativas são mostrados na Tabela 5.9.

Tabela 5.9 - Resultados obtidos em três minimizações com o método do Enxame de Partículas.

Tentativa	k_0 [s ⁻¹]	E [K]	F_{Obj} ($\times 10^2$)
1	0.8591	27640.3	1.02793
2	0.8623	27642.7	1.02793
3	0.8623	27642.6	1.02793

Do Exemplo 5.4, depreende-se que os valores ótimos dos parâmetros são iguais a 0.8623 s⁻¹ e 27642.7 K respectivamente para os parâmetros k_0 e E . O valor mínimo da

função objetivo é igual a 1.02793×10^{-2} . Observe que o método do Enxame de Partículas obteve valores finais praticamente iguais aos valores calculados com o método de Newton. Como discutido acima, este método tem a capacidade de realizar uma busca com característica global no início, sendo que a característica da busca passa a ser local ao longo das iterações, o que possibilita que o resultado seja encontrado com grande precisão.

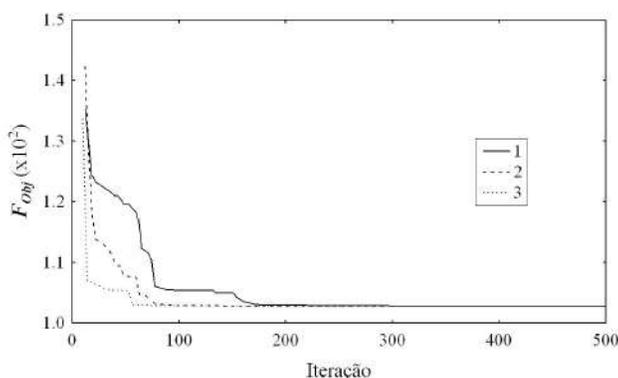


Figura 5.15 - Variação da função objetivo ao longo das iterações com o método do Enxame de Partículas.

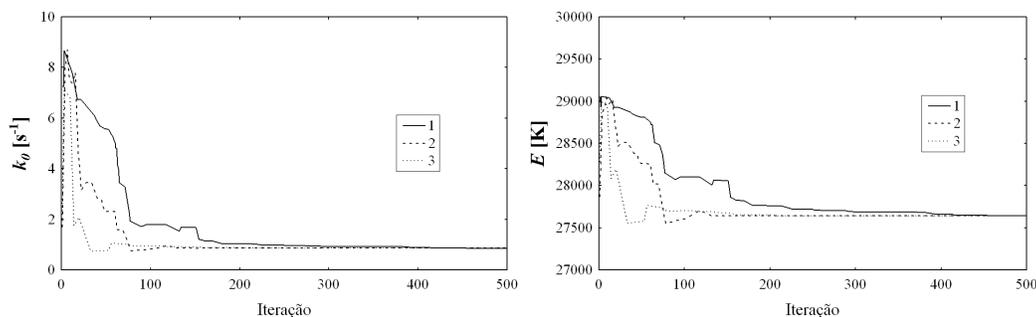


Figura 5.16 - Variação dos melhores parâmetros obtidos ao longo das iterações com o método do Enxame de Partículas.

A Figura 5.16 mostra que os valores finais dos parâmetros e da função objetivo foram muito próximos entre si nas três tentativas, o que não ocorreu com os demais métodos heurísticos.

5.6.5. Implementações Híbridas

Além dos métodos de otimização apresentados nas seções anteriores, muitos outros métodos podem ser ainda encontrados em livros especializados no assunto. Por isso, o leitor interessado é estimulado a consultar a literatura adicional sugerida ao final do capítulo. Também é interessante avaliar a possibilidade de construir algoritmos híbridos, onde a busca se inicia com um método heurístico (que realiza uma busca global e obtém uma boa estimativa inicial para os parâmetros) e termina com um método determinístico, como o método de Newton (que acelera a convergência final e permite atingir precisão elevada das estimativas). Entretanto, os métodos heurísticos, e em particular o método do Enxame de Partículas, vêm demonstrando enorme eficiência

e utilidade, quando se pretende desenvolver uma avaliação estatística rigorosa dos resultados finais, como será visto a seguir.

5.7. Cálculo Numérico da Região de Confiança dos Parâmetros

O procedimento de estimação de parâmetros não termina após a minimização da função objetivo. Uma etapa fundamental é a interpretação estatística dos resultados obtidos, como discutido no Capítulo 4. Portanto, é necessário interpretar os valores finais da função objetivo e dos parâmetros. A avaliação do valor mínimo da função objetivo consiste em comparar o valor obtido com os valores mínimos e máximos esperados, a partir da distribuição dos erros experimentais usados para gerar a função objetivo. No caso de flutuações normais, os valores mínimo e máximo são definidos pela distribuição χ^2 (chi-quadrado), dados o número de graus de liberdade e o nível de confiança desejado pelo usuário. Assim, se o valor da função objetivo no ponto de mínimo é maior que o limite superior da distribuição χ^2 , o modelo não é adequado. Por outro lado, se o valor final da função objetivo é menor que o limite inferior da distribuição χ^2 , o modelo é bom demais, indicando uma provável super parametrização do modelo, como discutido no Capítulo 4.

Já os parâmetros do modelo são avaliados em relação à região de confiança e ao nível de confiança. O procedimento clássico consiste em obter a aproximação quadrática da função objetivo, que leva à seguinte definição da matriz de covariância dos parâmetros (ver Seção 4.6.2 do Capítulo 4):

$$\mathbf{V}_a = 2\mathbf{H}_a^{-1} = [\mathbf{B}^T \mathbf{V}_y^{-1} \mathbf{B}]^{-1} \quad (5.34)$$

conforme as Equações (4.81) e (4.82) do Capítulo 4. A expressão normalmente usada para determinar da região (e o nível) de confiança dos parâmetros passa pela aproximação quadrática da função objetivo, como escrito a seguir

$$F_{Obj}(\mathbf{a}) = F_{Obj}(\hat{\mathbf{a}}) + (\mathbf{a} - \hat{\mathbf{a}})^T \nabla F_{Obj\hat{\mathbf{a}}} + \frac{1}{2}(\mathbf{a} - \hat{\mathbf{a}})^T \mathbf{H}_a (\mathbf{a} - \hat{\mathbf{a}}) \quad (5.35)$$

onde o sinal $\hat{}$ indica o valor estimado; isto é, o valor para o qual a função objetivo atinge o valor mínimo. Lembrando que no ponto de mínimo o gradiente da função objetivo é nulo, a partir da Equação (5.34) é possível escrever:

$$F_{Obj}(\mathbf{a}) - F_{Obj}(\hat{\mathbf{a}}) = (\mathbf{a} - \hat{\mathbf{a}})^T \mathbf{V}_a^{-1} (\mathbf{a} - \hat{\mathbf{a}}) \quad (5.36)$$

Admitindo que os desvios existentes entre os valores preditos pelo modelo e os valores experimentais são pequenos e seguem uma distribuição normal e considerando ainda que o modelo é linear (ou que pode ser linearizado nas proximidades do ponto de mínimo), conclui-se que as incertezas paramétricas também seguem a distribuição normal. Portanto, o lado direito da Equação (5.36), onde aparece a soma dos desvios quadrados de variáveis que seguem a distribuição normal, indica que ambos os lados da Equação (5.36) seguem a distribuição χ^2 (chi-quadrado) com NP graus de liberdade. Assim, utilizando-se o limite superior da distribuição χ^2 , é possível definir a região de confiança dos parâmetros que satisfazem a seguinte inequação:

$$(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})^T \mathbf{V}_a^{-1} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}) \leq \chi_{NP}^\alpha \quad (5.37)$$

onde α indica o nível de confiança. Assim, todos os conjuntos de valores dos parâmetros $\boldsymbol{\alpha}$ que satisfazem a Equação (5.37) pertencem à região de confiança. A igualdade da Equação (5.37) gera uma elipse em duas dimensões, como observado no Capítulo 3 (item 3.5.1), onde a geometria da região de confiança da curva normal foi discutida.

Na maioria das vezes, apenas uma aproximação da matriz de covariâncias dos desvios experimentais (\mathbf{V}_y) está disponível. Assim, é conveniente dividir ambos os lados da Equação (5.36) pelo valor da função objetivo no ponto de mínimo, como na equação seguinte:

$$\frac{F_{Obj}(\boldsymbol{\alpha}) - F_{Obj}(\hat{\boldsymbol{\alpha}})}{F_{Obj}(\hat{\boldsymbol{\alpha}})} = \frac{(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})^T \mathbf{V}_a^{-1} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})}{F_{Obj}(\hat{\boldsymbol{\alpha}})} \quad (5.38)$$

Como a função objetivo também pode ser vista como a soma de desvios quadrados de variáveis que seguem a distribuição normal, a função objetivo também tem distribuição χ^2 , mas com $NE.NY - NP = N - NP$ graus de liberdade. Como a Equação (5.38) apresenta a razão entre duas variáveis, cada qual com uma distribuição χ^2 , esta razão segue uma distribuição F de Fisher. Usando o limite superior da distribuição F para o lado direito da Equação (5.38), pode-se determinar a seguinte expressão para região de confiança:

$$(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})^T \mathbf{V}_a^{-1} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}) \leq F_{Obj}(\hat{\boldsymbol{\alpha}}) \frac{NP}{N - NP} F_{NP, N - NP}^\alpha \quad (5.39)$$

onde a razão entre NP e $NE.NY - NP = N - NP$ aparece para normalizar as duas distribuições que têm graus de liberdade diferentes.

A região de confiança definida pela Equação (5.39) também tem a forma elíptica, como a definida pela Equação (5.37), mas tende a ser maior que a primeira. Isso ocorre porque o grau de incerteza aumenta quando se usa a distribuição F ao invés da distribuição χ^2 (são duas as fontes de incerteza). Porém, a utilização da Equação (5.39) deve ser preferida, pois, como comentado acima, na maioria das vezes não se conhecem profundamente os erros experimentais.

Devido à não linearidade dos modelos usados no dia-a-dia de cientistas e engenheiros, as aproximações feitas anteriormente podem levar a regiões de confiança mal determinadas, que não representam acuradamente a realidade estatística do problema. Isso ocorre porque a região de confiança só é elíptica quando os parâmetros têm distribuição normal, o que só é rigorosamente verdadeiro quando os dados experimentais seguem a distribuição normal e o modelo é linear. Uma forma mais adequada para definir a região de confiança é utilizar o lado esquerdo da Equação (5.38) diretamente. Apesar da Equação (5.38) estar fundamentada nas mesmas hipóteses fundamentais que geraram as demais equações, ela não restringe a forma da região de

confiança, como será discutido adiante. Assim, usando o lado esquerdo da Equação (5.38), é possível escrever:

$$F_{Obj}(\boldsymbol{\alpha}) \leq F_{Obj}(\hat{\boldsymbol{\alpha}}) \left(1 + \frac{NP}{N-NP} F_{NP, N-NP}^{\alpha} \right) \quad (5.40)$$

A região de confiança pode ser determinada genericamente como todos os conjuntos de parâmetros que levam a valores da função objetivo que satisfazem a Equação (5.40). Para um modelo linear, a região de confiança determinada pelas Equações (5.39) e (5.40) são iguais; contudo, podem aparecer diferenças significativas para modelos não lineares. Apesar da Equação (5.40) estar fundamentada na aproximação quadrática da função objetivo (ou aproximação linear do modelo), a forma da região de confiança não é restringida à forma elíptica. Assim, regiões de confiança muito próximas da região de confiança estatisticamente real podem ser obtidas a partir da Equação (5.40). As regiões de confiança determinadas a partir desta equação são usualmente chamadas de Regiões de Verossimilhança.

A obtenção da região de confiança a partir da Equação (5.40) pode constituir uma tarefa numérica muito difícil, pois é virtualmente impossível determinar o contorno definido pela igualdade da Equação (5.40) com modelos não lineares. Felizmente, a aplicação dos métodos heurísticos para minimização da função objetivo gera um número muito grande de pontos, boa parte deles avaliada nas proximidades do ponto de mínimo. A utilização destes pontos para determinar a região de confiança dos parâmetros é muito simples, já que basta comparar o valor da função objetivo obtida com o limite definido na Equação (5.40) para saber se o conjunto de parâmetros deve ser incluído ou não na região de confiança dos parâmetros. Assim, além de facilitar a obtenção de uma região de confiança muito próxima da real, a maior desvantagem dos métodos heurísticos, que é o grande número de avaliações da função objetivo, é transformada em uma grande vantagem numérica em relação aos métodos determinísticos.

Nos Exemplos a seguir são resolvidos problemas de estimação de parâmetros com o método do Enxame de Partículas. Os resultados obtidos são também usados para determinação da região de confiança. Os resultados obtidos com o uso das Equações (5.39) e (5.40) são comparados. Também é analisado o efeito da reparametrização do modelo sobre a evolução do procedimento de estimação de parâmetros, em particular sobre a forma da região de confiança.

Exemplo 5.11 - O problema de estimação do Exemplo 5.4 foi resolvido com o método do Enxame de Partículas, como mostrado no Exemplo 5.10. Agora, será construída a região de confiança, utilizando-se a aproximação elíptica (Equação (5.39)) e a região de verossimilhança (Equação (5.40)).

Para a determinação da região elíptica e da região de verossimilhança foi utilizado um algoritmo híbrido. A busca é iniciada com o método do Enxame de Partículas (cujas avaliações são usadas para a determinação da região de verossimilhança) e o melhor valor encontrado é usado como estimativa inicial do estimador Máxima² (que consiste basicamente no método de Gauss-Newton, mas

² Noronha et al.(1993), Relatório Técnico PEQ/COPPE, UFRJ, Rio de Janeiro, RJ, Brasil

implementado com controles de convergência que aumentam a robustez do método). Com o Máxima, obtêm-se a solução refinada do problema e a matriz de covariâncias dos parâmetros, como definida na Equação (5.34), que é então usada para a determinação da região de confiança elíptica.

Deve-se lembrar de que como o valor de k_0 é da ordem de 10^{17} , foi inserida uma constante na equação do modelo, para que o parâmetro k_0 ficasse com um valor próximo de 1, como mostra a equação abaixo.

$$y_i = \exp \left[-k_0 10^{17} t_i \exp \left(-\frac{E}{T_i} \right) \right]$$

O ponto mínimo é encontrado com um valor da função objetivo igual a 1.02793×10^{-2} , quando os parâmetros assumem os valores de 0.862314 e 27642.7 para k_0 e E , respectivamente (lembre que k_0 já é multiplicado por 10^{17} na equação do modelo). A matriz de covariância dos parâmetros obtida é igual a:

$$\mathbf{V}_a = \begin{bmatrix} 1.51917 \times 10^3 & 1.10690 \times 10^6 \\ 1.10690 \times 10^6 & 8.06697 \times 10^8 \end{bmatrix}$$

de onde é possível observar que a correlação entre os parâmetros é igual a 0.99981, mostrando que os parâmetros estão altamente correlacionados. Como discutido na Seção 4.7 do capítulo 4, isso é muito ruim e indica problemas de formulação do modelo e/ou dos dados experimentais, não sendo possível avaliar os parâmetros de forma independente.

Substituindo a matriz de covariância dos parâmetros, os parâmetros estimados, o valor mínimo da função objetivo, NP , $NY.NE-NP=N-NP$ e o valor máximo admissível para a variável F (igual a 3.238096 nesse caso, para um nível de confiança de 95%) na Equação (5.39), obtêm-se a seguinte expressão

$$\left(\begin{bmatrix} k_0 \\ E \end{bmatrix} - \begin{bmatrix} 0.862314 \\ 27642.7 \end{bmatrix} \right)^T \begin{bmatrix} 1.51917 \times 10^3 & 1.10690 \times 10^6 \\ 1.10690 \times 10^6 & 8.06697 \times 10^8 \end{bmatrix}^{-1} \left(\begin{bmatrix} k_0 \\ E \end{bmatrix} - \begin{bmatrix} 0.862314 \\ 27642.7 \end{bmatrix} \right) \leq 1.7069 \times 10^{-3}$$

que define uma elipse no espaço dos parâmetros.

A região de confiança de verossimilhança é definida pela Equação (5.40). A substituição do valor mínimo da função objetivo, NP , $NE.NY-NP=N-NP$ e do valor máximo admissível para a variável F (igual a 3.238096 nesse caso, para um nível de confiança de 95%) leva à seguinte equação:

$$F_{Obj}(\mathbf{a}) \leq 1.1986 \times 10^{-2}$$

Assim, dos pontos previamente avaliados pelo método do Enxame de Partículas para a minimização, aqueles que tiverem um valor da função objetivo menor que $1,1986 \cdot 10^{-2}$ estão dentro da região de confiança.

As Figuras 5.17 e 5.18 apresentam as regiões de confiança calculadas pelas Equações (5.39) e (5.40). Também é apresentada a região de verossimilhança exata, para verificar a qualidade da aproximação obtida pelo método do Enxame de Partículas.

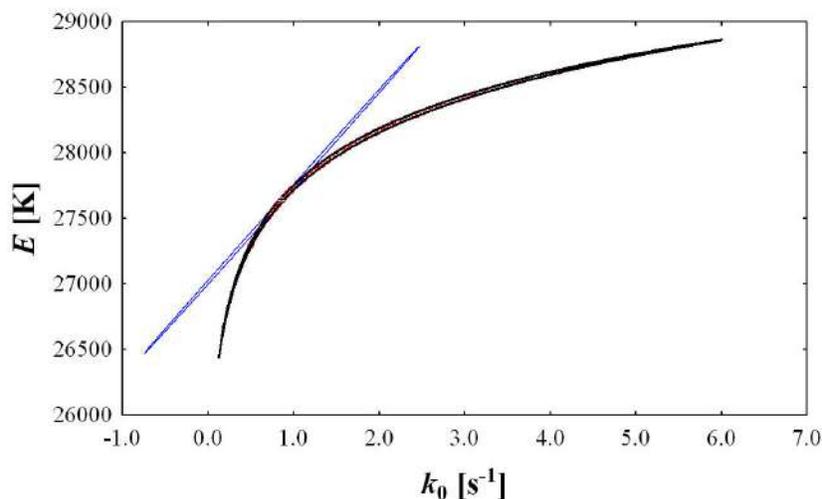


Figura 5.17 - Região de confiança elíptica (linha azul), de verossimilhança obtida pelo Enxame de Partículas (pontos vermelhos) e de verossimilhança exata (linha preta).

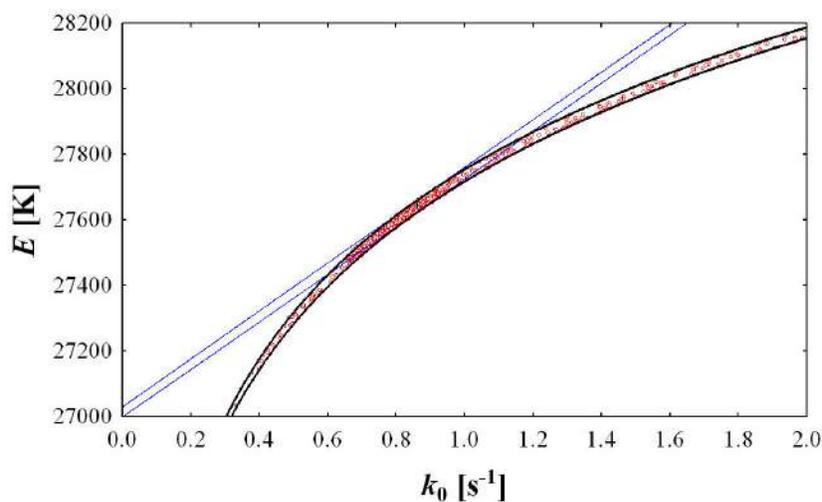


Figura 5.18 - Ampliação da Figura 5.17 na região próxima ao mínimo.

Na Figura 5.17 observa-se que a região de confiança elíptica prevê que o parâmetro k_0 pode assumir valores negativos, o que seria indicação de que este parâmetro não é significativo. Mas o que acontece na realidade é que a curvatura da região de confiança real nas proximidades do ponto de mínimo faz com que a aproximação quadrática seja muito ruim, levando a uma região de confiança elíptica totalmente equivocada. Além disso, a região de confiança real mostra-se como uma faixa estreita e comprida. Isso dificulta a estimação dos parâmetros, por ser difícil encontrar esta pequena região onde o mínimo se encontra. Essa característica também faz com que a correlação entre os parâmetros seja muito alta, dificultando a identificação dos parâmetros.

Na Figura 5.18 pode-se observar que a região elíptica e a região de verossimilhança se confundem nas proximidades do ponto de mínimo. Contudo, à medida que se afastam do ponto de mínimo, as duas regiões tornam-se muito diferentes.

Devido à estreita faixa que forma a região de confiança, fica difícil avaliar o desempenho do método do Exame de Partículas para determinar a região de confiança de verossimilhança na Figura 5.17. Contudo, a Figura 5.18 mostra que os pontos avaliados pelo Exame representam adequadamente o comportamento da região de verossimilhança; isto é, mostram a curvatura da região de confiança e se afastam da aproximação elíptica.

Exemplo 5.12 - O mesmo problema de estimação do Exemplo 5.11 foi resolvido para duas formas de reparametrização da equação do modelo. A primeira consiste em apenas colocar o parâmetro k_0 dentro da segunda exponencial (juntamente com a constante 10^{17}), onde agora aparece o parâmetro A que é um parâmetro adimensional numericamente equivalente a $\ln(10^{17} \cdot k_0)$.

$$y_i = \exp \left[-t_i \exp \left(A - \frac{E}{T_i} \right) \right]$$

O ponto mínimo é encontrado para um valor final da função objetivo igual a 1.02793×10^{-2} , sendo os valores finais dos parâmetros iguais a 38.9958 e 27642.7 para A e E , respectivamente. A matriz de covariâncias dos parâmetros é calculada como:

$$\mathbf{V}_a = \begin{bmatrix} 2.04304 \times 10^3 & 1.28364 \times 10^6 \\ 1.28364 \times 10^6 & 8.06698 \times 10^8 \end{bmatrix}$$

de onde pode-se calcular a correlação entre os parâmetros, que é igual a 0.99981. Uma vez mais, os parâmetros estão altamente correlacionados, o que é muito ruim para a estimação e a formulação do modelo. Observe que a variância do parâmetro E e a correlação entre os parâmetros não sofreu modificação alguma. Apenas os valores do primeiro parâmetro e da sua variância sofreram alterações, como já podia ser esperado, já que este parâmetro foi definido de uma outra maneira.

Usando Equação (5.39), como feito no Exemplo 5.11 chega-se a:

$$\left(\begin{bmatrix} A \\ E \end{bmatrix} - \begin{bmatrix} 38.9958 \\ 27642.7 \end{bmatrix} \right)^T \begin{bmatrix} 1.51917 \times 10^3 & 1.10690 \times 10^6 \\ 1.10690 \times 10^6 & 8.06697 \times 10^8 \end{bmatrix}^{-1} \left(\begin{bmatrix} A \\ E \end{bmatrix} - \begin{bmatrix} 38.9958 \\ 27642.7 \end{bmatrix} \right) \leq 1.7069 \times 10^{-3}$$

que define uma elipse no espaço dos parâmetros.

A região de confiança de verossimilhança é definida pela Equação (5.40). A substituição do valor mínimo da função objetivo, NP , $NE.NY-NP=N-NP$ e do valor máximo admissível para a variável F (igual a 3.238096 nesse caso, para um nível de confiança de 95%) leva à seguinte equação:

$$F_{Obj}(\mathbf{a}) \leq 1.1986 \times 10^{-2}$$

Observe que não há modificação no valor limite para a definição da região de verossimilhança, porque foi modificada apenas a definição do parâmetro, não o significado estatístico do problema. Nas Figuras 5.19 e 5.20 são apresentadas as regiões de confiança para a primeira forma de reparametrizar o modelo.

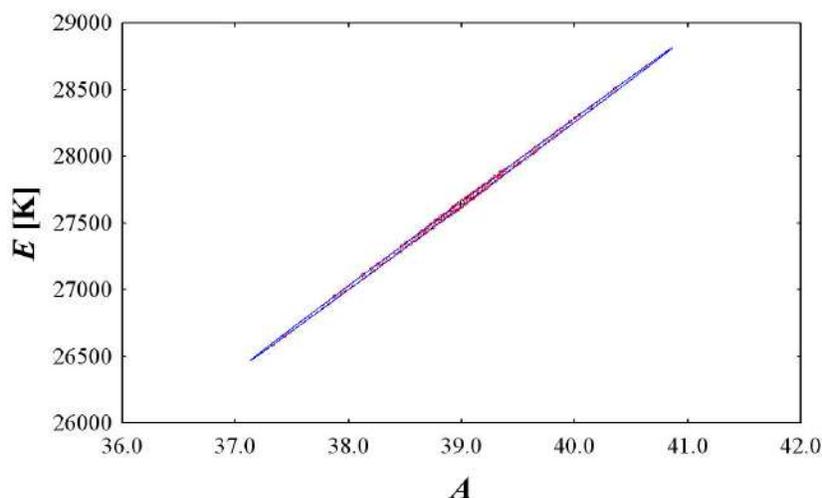


Figura 5.19 - Região de confiança elíptica (linha azul) e de verossimilhança obtidas pelo Enxame de Partículas (pontos vermelhos) com a primeira reparametrização do modelo.

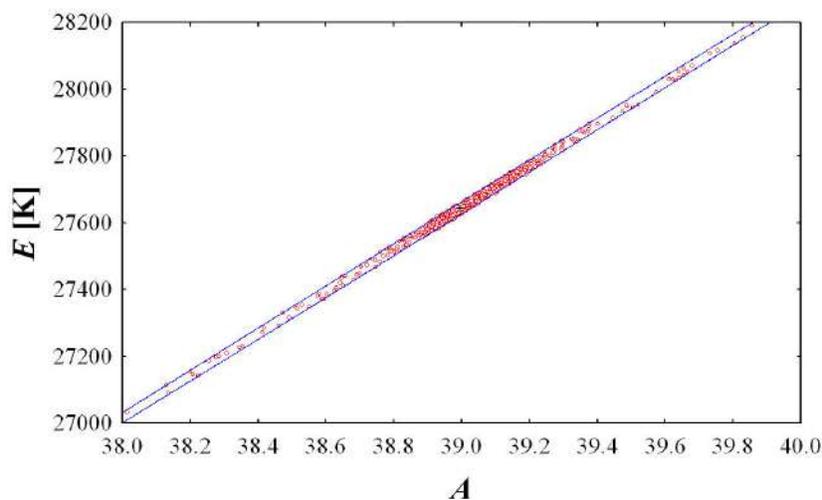


Figura 5.20 - Ampliação da Figura 5.19 na região próxima ao mínimo.

Apesar da correlação entre os parâmetros continuar altíssima, a aproximação elíptica da região de confiança apresenta boa qualidade para esta forma de reparametrização do modelo. Mais uma vez, a região obtida pelo método do Enxame de Partículas apresenta uma boa qualidade, mesmo em um problema que apresenta uma região de confiança muito estreita e de difícil determinação. É importante observar que a forma da região de confiança muda com a forma de parametrização do modelo. Portanto, é possível mudar a forma da região de confiança através da definição

apropriada dos parâmetros. Observe que o usuário pode definir os parâmetros da maneira que quiser, uma vez que os parâmetros não são efetivamente medidos, mas inferidos com o auxílio do modelo.

A segunda forma de reparametrização consiste em utilizar uma temperatura de referencia T_0 , definida para este problema como 630 K, e reescrever o modelo da seguinte maneira:

$$y_i = \exp \left[-k_{T_0} t_i \exp \left(-E \left(\frac{1}{T_i} - \frac{1}{T_0} \right) \right) \right]$$

onde os parâmetros que devem ser estimados são k_{T_0} e E . (Observe que aqui a constante 10^{17} não é usada, já que k_{T_0} é a constante cinética na temperatura T_0 e seu valor é da faixa de 10^{-3} e por isso não foi usada nenhuma constante).

O ponto de mínimo é encontrado com um valor final da função objetivo igual a 1.02793×10^{-2} , sendo os valores finais dos parâmetros iguais a $7.58624 \times 10^{-3} \text{ s}^{-1}$ e 27642.7 K para k_{T_0} e E , respectivamente. A matriz de covariância dos parâmetros pode ser calculada como:

$$\mathbf{V}_a = \begin{bmatrix} 2.87894 \times 10^{-5} & 2.40003 \times 10^1 \\ 2.40003 \times 10^1 & 8.06698 \times 10^8 \end{bmatrix}$$

de onde se pode calcular que a correlação entre os parâmetros é igual a 0.157487. Isso mostra que esta forma de reparametrização praticamente elimina a correlação entre os parâmetros. Essa observação é fantástica, pois mostra que o usuário tem liberdade para influir sobre a qualidade e significância dos resultados obtidos.

Usando Equação (5.39), como no Exemplo 5.11, chega-se a:

$$\left(\begin{bmatrix} k_{T_0} \\ E \end{bmatrix} - \begin{bmatrix} 7.58624 \times 10^{-3} \\ 27642.7 \end{bmatrix} \right)^T \begin{bmatrix} 2.87894 \times 10^{-5} & 24.0003 \\ 24.0003 & 8.06698 \times 10^8 \end{bmatrix}^{-1} \left(\begin{bmatrix} k_{T_0} \\ E \end{bmatrix} - \begin{bmatrix} 7.58624 \times 10^{-3} \\ 27642.7 \end{bmatrix} \right) \leq 1.7069 \times 10^{-3}$$

A expressão usada para construir a região de confiança de verossimilhança é a mesma já obtida para as demais formas de parametrização do modelo. Na Figura 5.21 são apresentadas as regiões de confiança obtidas com a segunda forma de reparametrizar o modelo.

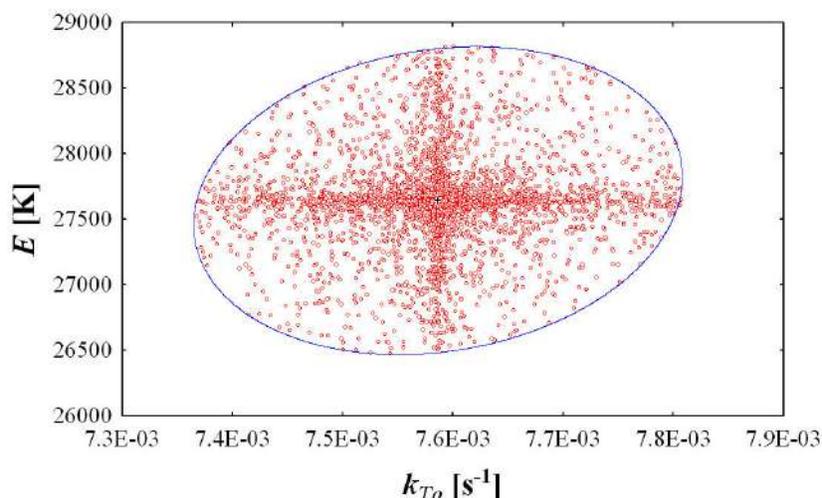


Figura 5.21 - Região de confiança elíptica (linha azul) e de verossimilhança obtidas pelo Enxame de Partículas (pontos vermelhos) com a segunda reparametrização do modelo.

Observe que a aproximação elíptica da região de confiança tem excelente qualidade e que o Enxame de Partículas consegue descrever esta região também com excelente qualidade. Observe que, como os parâmetros têm uma correlação próxima a zero, os eixos da elipse são praticamente ortogonais aos eixos cartesianos na Figura 5.21, permitindo identificação independente dos diferentes parâmetros do modelo. Conclui-se, portanto, que o usuário deve investigar diferentes formas de apresentação do modelo, para tornar possível a identificação independente dos diferentes efeitos considerados.

5.8. A Forma de Apresentação dos Dados Experimentais

O Exemplo 5.12 parece incentivar o usuário a manipular a apresentação do modelo, de modo a contribuir com a qualidade da representação estatística dos resultados finais. Isso é possível porque o usuário é livre para apresentar o modelo na forma que achar mais conveniente. Portanto, uma pergunta que parece bastante pertinente, como já comentado no Exemplo 4.17 do Capítulo 4, diz respeito à possibilidade de implementar modificações nos dados experimentais, para facilitar a representação matemática do modelo. Por exemplo, os modelos apresentados na Equação (5.41)

$$y^m = k_0 \exp\left(-\frac{\Delta E}{R x^m}\right) \text{ e } y^m = \frac{K_1 x^m}{1 + K_2 x^m} \tag{5.41}$$

podem ser escritos na forma linear da Equação (5.42)

$$z^m = \ln(y^m) = \ln(k_0) - \frac{\Delta E}{R x^m} = \alpha_1 + \alpha_2 w^m \tag{5.42}$$

e

$$z^m = \frac{1}{y^m} = \frac{K_2}{K_1} + \frac{1}{K_1 x^m} = \alpha_1 + \alpha_2 w^m$$

As representações das Equações (5.41) e (5.42) (representação de y como função de x ou, alternativamente, de z como função de w) são absolutamente equivalentes, do ponto de vista do modelo. Apesar disso, essas transformações jamais devem ser efetuadas, porque elas **não são equivalentes** do ponto de vista dos dados experimentais. O usuário deve perceber que as propriedades estatísticas de y e de z (e, analogamente, de x e de w) não são as mesmas. Portanto, uma função objetivo desenvolvida para a variável medida y (e, por extensão, x) não pode ser usada para a variável transformada z (e, por extensão, w). Por exemplo, suponha que os desvios experimentais a que a medida y está sujeita sejam pequenos (o que não é necessariamente verdade). Nesse caso, combinando a Equação (5.42) com a Equação (4.65), é possível escrever:

$$dz = \frac{dy}{y} \rightarrow \sigma_z^2 = \frac{\sigma_y^2}{y^2} \quad (5.43)$$

e

$$dz = -\frac{dy}{y^2} \rightarrow \sigma_z^2 = \frac{\sigma_y^2}{y^4}$$

A Equação (5.43) mostra de forma bastante clara que as variâncias de medida das variáveis transformadas podem ser funções complexas das condições de medição. Nesse caso, mesmo que faça sentido usar a técnica de mínimos quadrados para a representação proposta na Equação (5.41), a partir da informação de que os erros de medida da variável y são constantes na região de experimentação, o uso da técnica de mínimos quadrados para a representação da Equação (5.42) é completamente descabido, como mostra a Equação (5.43). Ainda mais sério, ao usar a representação da Equação (5.42) para estimar os parâmetros do modelo, o usuário estará implicitamente admitindo que os erros de medida da variável y crescem com uma potência elevada do valor medido, o que é quase sempre um absurdo! Ou seja, a transformação das medidas experimentais pode provocar deformações profundas no procedimento de estimação de parâmetros e no conjunto final de resultados obtidos. Por isso, **dados experimentais nunca devem ser transformados.**

O usuário deve perceber que a aplicação das técnicas numéricas apresentadas nesse capítulo permite realizar os procedimentos de estimação de parâmetros de forma adequada, quaisquer que sejam as naturezas dos dados experimentais e dos modelos utilizados. Por isso, o usuário deve resistir à tentação de representar os dados experimentais numa falsa representação linear, que permita a obtenção de soluções analíticas para o problema de estimação. Apenas uma situação pode justificar a transformação das medidas experimentais – a necessidade de obter um conjunto de estimativas iniciais confiáveis para o conjunto de parâmetros do modelo, para simplificar a busca realizada pelo algoritmo numérico. Nesse caso, a estimação deve ser realizada ao menos duas vezes: uma na forma transformada da medida, como na Equação (5.42), para que se obtenha uma estimativa inicial confiável dos parâmetros; outra na forma natural da medida, como na Equação (5.41), para que se obtenham as estimativas reais compatíveis com as propriedades estatísticas das medidas efetuadas.

Exemplo 5.13 - O problema de estimação apresentado no Exemplo 5.4 é resolvido agora através da representação linear da equação do modelo (lembrando que nos exemplos anteriores sempre foi utilizada a forma original não linear do modelo). A equação original do modelo é definida como:

$$y_i = \exp \left[-k_0 t_i \exp \left(-\frac{E}{T_i} \right) \right]$$

Aplicando-se o logaritmo em ambos os lados da equação, chega-se a:

$$\ln(y_i) = -k_0 t_i \exp \left(-\frac{E}{T_i} \right)$$

Multiplicando-se os dois lados da equação por -1 e aplicando-se uma vez mais a operação de logaritmo em ambos os lados da equação, obtém-se:

$$\ln[-\ln(y_i)] = \ln(k_0) + \ln(t_i) - \frac{E}{T_i}$$

Portanto, é possível escrever o modelo na forma:

$$\eta = \alpha_1 + x_1 - \alpha_2 x_1$$

onde as novas variáveis são:

$$\eta = \ln[-\ln(y_i)]$$

$$x_1 = \ln(t_i)$$

$$x_2 = 1/T_i$$

e os parâmetros são:

$$\alpha_1 = \ln(k_0)$$

$$\alpha_2 = E$$

Os parâmetros são agora estimados de forma a minimizar a função de mínimos quadrados:

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

Aqui aparece um primeiro ponto importante. Nos exemplos apresentados ao longo deste capítulo sempre foi considerado que os desvios na variável y seguem a distribuição normal, com erro constante na região de experimentação. Mas, agora, qual é a função de densidade de probabilidades para a nova variável considerada η ? Se as

variâncias da variável y fossem diferentes em cada condição experimental, como seria o comportamento das variâncias da nova variável η ? Admitindo-se que os erros de medida de y sejam pequenos, pode-se escrever

$$d\eta = -\frac{d \ln(y)}{\ln(y)} = -\frac{dy}{y \ln(y)} \rightarrow \sigma_{\eta}^2 = \frac{\sigma_y^2}{[y \ln(y)]^2}$$

que mostra a complexidade da transformação de erros proposta. (Observe que a transformação tende a infinito quando y se aproxima dos valores 0 e 1.)

Além destas questões relacionadas à caracterização dos erros, é necessário perceber que o ponto de mínimo da função de mínimos quadrados das diferenças entre y experimental e y calculado:

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

não é igual ao ponto de mínimo da função de mínimos quadrados das diferenças entre η experimental e η calculado:

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

Fazendo então a minimização da função objetivo das diferenças da nova variável η , os parâmetros obtidos são $\alpha_1 = 39.1228$ e $\alpha_2 = 27710.6$ K. Este resultado pode ser comparado ao obtido no Exemplo 5.12, já que os parâmetros são definidos de forma análoga. Os valores dos parâmetros obtidos no Exemplo 5.12 são iguais a 38.9958 e 217642.7 K. Observe que o resultado obtido não é o mesmo, mostrando que o ponto ótimo da função objetivo muda com a mudança das variáveis. Como o objetivo real do modelo é fazer previsões da variável y , e não da variável η , a representação linear do modelo leva a valores dos parâmetros que não devem ser considerados como valores ótimos.

5.9. Conclusões

Foram apresentadas nesse capítulo diferentes técnicas numéricas que permitem a obtenção dos parâmetros que minimizam as funções objetivos propostas para redução da distância existente entre os dados experimentais e as previsões fornecidas pelo modelo. As diferentes técnicas apresentam vantagens e desvantagens características, de maneira que o usuário é levado a interagir com o problema para definir apropriadamente as técnicas que deve usar. Por exemplo, técnicas numéricas determinísticas derivadas do método de Newton são eficientes (convergem rapidamente com relativamente poucos pontos), mas são excessivamente dependentes da disponibilidade de boas estimativas iniciais para os parâmetros (são pouco robustas). Por outro lado, técnicas heurísticas são muito robustas (dependem pouco da disponibilidade de boas estimativas iniciais), mas

são computacionalmente custosas (dependem de grande número de avaliações da função objetivo). Entre um extremo e outro, é possível encontrar técnicas numéricas que apresentam virtualmente todos os tipos de características intermediárias. O usuário deve considerar com carinho a possibilidade de combinar os métodos heurísticos e determinísticos, para ganhar o que os dois grupos de métodos oferecem de melhor: baixa sensibilidade às condições iniciais e convergência rápida nas proximidades do ponto de mínimo.

Foram também apresentadas técnicas numéricas que permitem a avaliação da região de confiança dos parâmetros. Curiosamente, observou-se que a forma de parametrização do modelo influi decisivamente sobre a qualidade dos resultados obtidos. Portanto, o usuário deve investigar diferentes formas de apresentação do modelo, para tornar possível a identificação independente dos diferentes efeitos paramétricos considerados. Finalmente, foi mostrado que, ao contrário dos parâmetros do modelo, a forma de apresentação dos dados experimentais não deve ser jamais modificada, sob pena de descaracterizar completamente o significado estatístico da análise efetuada.

5.10. Leitura Adicional

A literatura dedicada à apresentação de métodos numéricos para a solução de problemas de otimização é muito vasta. Não cabe aqui, portanto, uma revisão exaustiva dessa área. O leitor interessado encontrará centenas de livros que abordam esses assuntos em qualquer biblioteca dedicada à Matemática e à Engenharia. Algumas dessas referências são apresentadas abaixo, para problemas específicos de estimação de parâmetros.

“Applied Nonlinear Programming”, D.M. Himmelblau, McGraw-Hill, New York, **1972**.

“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**.

Para aqueles leitores interessados em um universo mais amplo de procedimentos numéricos, voltados para a área genérica de otimização de processos, sugerem-se as seguintes referências adicionais:

“Optimization of Chemical Processes”, T.F. Edgar e D.M. Himmelblau, McGraw-Hill, New York, **1988**.

“Genetic algorithms in search, optimization and machine learning”, D.E. Goldberg. Addison Wesley Longman, Inc., Boston, 1989.

“Numerical Optimization”, J. Nocedal, S.J. Wright, Springer-Verlag Inc, New York, 1999.

“Swarm Intelligence”, J. Kennedy, R.C. Eberhart, Y. Shi, Morgan Kaufman, 2001.

“Practical Genetic Algorithms”, R.L. Haupt, S.E. Haupt, John Wiley & Sons, 2ª edição, New York, 2004.

Os leitores interessados na análise teórica e experimental do efeito dos erros experimentais sobre a estimação de parâmetros, sugerem-se as seguintes referências adicionais:

Santos, T.J., Pinto, J.C., 1998, “Taking variable correlation into consideration during parameter estimation”, *Brazilian Journal of Chemical Engineering*, 15, 1-20.

Larentis, A.L., Bentes Jr., A.M.P., Resende, N.S., Salim, V.M.M., Pinto, J.C., 2003, “Analysis of experimental errors in catalytic tests for production of synthesis gas”, *Applied Catalysis A*, 242, 365-379.

Os leitores interessados numa análise mais aprofundada do efeito da reparametrização do modelo sobre a evolução e os resultados obtidos nos procedimentos de estimação de parâmetros, sugere-se a seguinte referência adicional:

Schwaab, M., Pinto, J.C., 2007, “Optimum reference temperature for reparameterization of the Arrhenius equation. Part 1: Problems involving one kinetic constant.”, *Chemical Engineering Science*, 67, 2750-2764.

5.11. Exercícios Sugeridos

1- Considere o seguinte modelo não linear

$$y = \alpha_1 (1 - \exp(-\alpha_2 x))$$

onde α_1 e α_2 são os parâmetros a serem estimados a partir do seguinte conjunto de dados:

x	y	σ^2
0.50	7.92	25.00
1.00	18.51	25.00
1.50	20.09	9.00
2.00	18.97	9.00
3.00	26.67	1.00
4.00	29.45	1.00
5.00	32.58	0.25
7.00	34.54	0.25

10.00	34.62	0.01
-------	-------	------

Para a minimização da função objetivo, utilize algum dos métodos numéricos propostos ou faça uso de algum pacote comercial disponível. Considere que as variáveis independentes não estão sujeitas a erros e que os desvios experimentais nas variáveis dependentes, caracterizados pela variância fornecida para cada condição experimental, são independentes e seguem a distribuição normal.

2- Considere o seguinte conjunto de dados:

x^e	y^e
1.00	1.66
2.00	6.07
3.00	7.55
5.00	9.72
10.00	15.24
15.00	18.79
20.00	19.33
30.00	22.38
40.00	24.27
50.00	25.51

sendo a variância dos desvios experimentais constante e igual a 1.0 em toda a região de experimentação. São propostos dois modelos para descrever estes dados:

$$\text{M1: } y = \frac{\alpha_1 x}{1 + \alpha_2 x}$$

$$\text{M2: } y = \alpha_1 x^{\alpha_2}$$

onde α_1 e α_2 são os parâmetros a serem estimados para cada modelo. Faça a estimação e compare os resultados obtidos por cada modelo. Na sua opinião, qual é o melhor modelo ?