

IDENTIFICAÇÃO DE PARÂMETROS EM
MODELOS ECOLÓGICOS

RENATO BORGES GUERRA

Orientador:
Prof. Dr. José Mário Martínez.

Dissertação apresentada no Instituto de Matemática, Estatística e Ciência da Computação, como requisito parcial para obtenção do título de Mestre em Matemática Aplicada.

Junho/1982

UNICAMP
BIBLIOTECA CENTRAL

Classif	T
Autor	g 9372
V.	Ex.
Ex.	
Tombo BCI	4555

CM-00030248-1

A

Pergy.

AGRADECIMENTOS

A todos os professores do curso de Matemática Aplicada, aos colegas e amigos que de forma direta ou indireta me ajudaram a concluir este trabalho. Em particular, ao Prof. Dr. José Mário Martínez pela sugestão do problema e pela paciente orientação.

ÍNDICE

INTRODUÇÃO.....	5
CAPÍTULO I - FORMULAÇÃO DO PROBLEMA	
1.1) Introdução.....	7
1.2) A Equação Logística.....	7
1.3) A Curva Logística.....	9
1.4) Duas Espécies Competindo.....	11
1.5) Comportamento da Solução.....	12
1.6) O Modelo Discreto.....	16
1.7) O Problema Inverso.....	18
CAPÍTULO II - MÉTODO DOS QUADRADOS MÍNIMOS LINEAR	
2.1) Introdução.....	19
2.2) Método dos Quadrados Mínimos Linear.....	19
2.3) Interpretação Geométrica.....	21
2.4) Aplicação ao Problema.....	22
CAPÍTULO III - MÉTODO DE LEVENBERG-MARQUARDT	
3.1) Introdução.....	25
3.2) Método de Levenberg-Marquardt.....	25
3.3) Aplicação ao Problema.....	36
CAPÍTULO IV - MÉTODO DE BREMERMAN	
4.1) Introdução.....	39
4.2) Método de Bremermann.....	39
4.3) Aplicação ao Problema.....	42
CAPÍTULO V - EXPERIÊNCIAS NUMÉRICAS	
5.1) Introdução.....	43
5.2) Resultados das Experiências Numéricas.....	43
CONCLUSÃO.....	58
BIBLIOGRAFIA.....	59

INTRODUÇÃO

Nos modelos de crescimento populacional aparecem parâmetros que precisam ser estimados de modo a ajustar estes modelos a dados observados. Estes modelos são expressos por sistemas de equações diferenciais não lineares as quais, geralmente, não podem ser explicitamente resolvidas, tornando assim, difícil o processo de estimação de parâmetros. Desta forma, o objetivo deste trabalho é estimar parâmetros em modelos discretos, obtidos com fundamento nos modelos anteriormente citados, e escolher a melhor técnica para isto.

Sendo assim, no Capítulo I, com o propósito de motivação, é feito um estudo do modelo de crescimento populacional de duas espécies, competindo pelos mesmos recursos limitados, para mostrar a importância dos parâmetros. Em seguida o problema é formulado.

No Capítulo II, devido às características próprias do modelo, o problema é abordado como um problema de estimação de parâmetros lineares. Por este motivo é feita uma descrição do método dos quadrados mínimos linear.

No Capítulo III é mostrado que o problema considerado é um problema de estimação de parâmetros não lineares e em seguida é apresentado o método de Levenberg-Marquardt o qual é específico para resolver problemas deste tipo. Devido à não simplicidade do cálculo das derivadas, exigidas por este método, é mostrado também como foram obtidas estas derivadas.

As dificuldades encontradas na aplicação de métodos

que usam derivadas para resolver problemas do tipo considerado são citadas no Capítulo IV. Assim é apresentado, neste capítulo, o método de Bremermann o qual dispensa o cálculo das derivadas e tem se mostrado, conforme o próprio autor, eficiente na resolução de problemas de estimação de parâmetros não lineares.

Finalmente, no Capítulo V, são mostrados os resultados obtidos com os sistemas testes considerados e é feita a comparação entre os métodos citados.

CAPÍTULO I

FORMULAÇÃO DO PROBLEMA

1.1) INTRODUÇÃO:

Neste capítulo, com o propósito de formular o nosso problema, faremos um estudo das equações que controlam o crescimento de duas espécies em competição pelos mesmos recursos. Neste estudo observaremos que estas equações não podem ser resolvidas explicitamente e por isso, apresentaremos um modelo discreto que será usado na formulação do problema.

1.2) A EQUAÇÃO LOGÍSTICA:

O modelo mais simples de crescimento populacional de uma espécie simples, considera a taxa de crescimento constante e diretamente proporcional à grandeza da população. Se denotarmos a população no tempo t por y_t e a taxa de crescimento por k , teremos

$$\frac{d y_t}{dt} = k y_t \quad (1.2.1)$$

onde a taxa de crescimento k é definida como a diferença entre a taxa de nascimento η e a taxa de morte μ , portanto, podendo ser negativa ou positiva. Entretanto, para populações naturais, k é geralmente positiva. Assim, se y_0 é a população inicial no instante $t = 0$, a solução da equação (1.2.1) é

$$y_t = y_0 e^{kt} \quad (1.2.2)$$

o que mostra um crescimento exponencial e ilimitado. Obviamente,

isto pode ser verdadeiro somente para uma população, de tal modo pequena, em que não exista interferência entre os seus membros. O crescimento de qualquer população, num ambiente limitado, precisa ser limitado por uma falta de recursos (por exemplo: alimentos) . Desta forma, um estágio é atingido quando a exigência feita pela população existente impede o crescimento adicional. Assim, o crescimento populacional é, eventualmente, limitado pela disponibilidade de recursos, o que acarreta que a taxa de crescimento por indivíduo, seja uma função da grandeza da população, ou seja, $k = f(y_t)$. Desta forma, a equação (1.2.1) pode ser escrita na forma:

$$\frac{d y_t}{dt} = y_t f(y_t) \quad (1.2.3)$$

Como os recursos disponíveis, por indivíduo, diminuem à medida que a população cresce, o que produz um efeito inibitório no crescimento populacional, a função $f(y_t)$ diminui à medida que y_t cresce, assim $df(y_t)/dt$ é negativa. A hipótese mais simples a ser feita sobre $f(y_t)$, é que ela seja uma função da forma $a-by_t$ com a e b positivos. Assim a equação (1.2.3) será escrita na forma:

$$\frac{d y_t}{dt} = y_t (a - by_t) \quad (1.2.4)$$

onde as constantes a e b são conhecidas como os "parâmetros logísticos" da espécie na condição ambiental particular. A equação (1.2.4) é conhecida como a "equação logística" de Verhulst-Pearl.

Existem outros argumentos que levam a equação (1.2.4), citados em [1], mas não os citaremos, por fugirem do nosso objetivo.

1.3) A CURVA LOGÍSTICA:

A solução da equação logística é da forma

$$y_t = \frac{a/b}{1 + e^{-at/cb}} \quad (1.3.1)$$

onde c é a constante de integração. Se y_0 é a população inicial no instante $t = 0$, temos

$$c = y_0 / (a - by_0)$$

Substituindo em (1.3.1), obtemos

$$y_t = \frac{a}{b} \left[1 + e^{-at} \left(\frac{a/b - y_0}{y_0} \right) \right] \quad (1.3.2)$$

A equação (1.3.2) é conhecida como "curva logística", a qual é mostrada na figura 1.

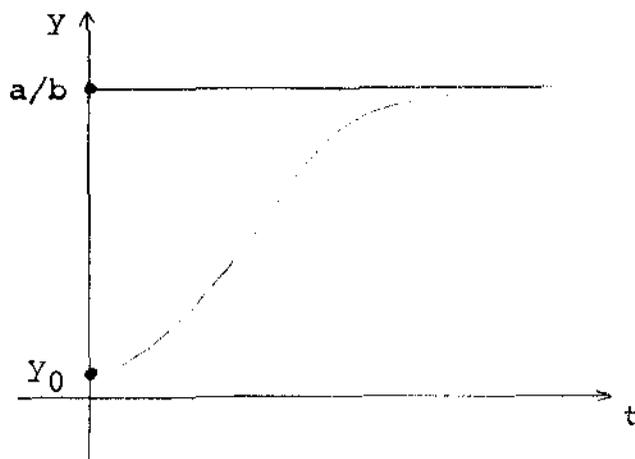


Figura 1

Da equação (1.3.2), observamos que quando $t \rightarrow \infty$, $y_t \rightarrow a/b$. Portanto, existe um valor assintótico de y_t que não pode ser excedido, devido a limitação do ambiente. Este valor assintótico $y = a/b$ é conhecido como "nível de saturação". Daremos, agora, uma idéia, mais clara, do nível de saturação.

A taxa de crescimento é a diferença entre a taxa de nascimento e taxa de morte, as quais são, geralmente, funções da grandeza da população. Se denotarmos a taxa de nascimento por $n(y)$ e a taxa de morte por $\mu(y)$, a taxa de crescimento será $f(y) = n(y) - \mu(y)$. Como $f(y)$ decresce quando y cresce, então $n(y)$ precisa decrescer e $\mu(y)$ crescer. Desta forma, podemos por $n(y) = a_1 - b_1 y$ e $\mu(y) = a_2 + b_2 y$ com $a_1, a_2 > 0$ e $b_1, b_2 > 0$. Assim a equação (1.2.4) pode ser escrita na forma

$$\frac{d y_t}{dt} = y_t [n(y_t) - \mu(y_t)]$$

ou

$$\frac{d y_t}{dt} = y_t [(a_1 - a_2) - (b_1 + b_2) y_t]$$

Deste modo, quando a falta de recursos impede o crescimento adicional, temos que a taxa de crescimento é nula, ou seja

$$n(y) - \mu(y) = 0$$

ou

$$a_1 - b_1 y = a_2 + b_2 y$$

ou ainda que

$$y = \frac{a_1 - a_2}{b_1 + b_2} = \frac{a}{b}$$

Assim, a "população de equilíbrio", onde a taxa de nascimento é igual a taxa de morte, é que chamamos de "nível de saturação".

1.4) DUAS ESPÉCIES COMPETINDO:

Na seção anterior, abordamos o caso em que somente uma espécie estava presente e que a população, desta espécie, cresceria tanto quanto os recursos disponíveis permitissem. Agora, consideramos duas espécies vivendo conjuntamente e competindo uma com a outra pelos mesmos recursos limitados. Assim, se y_1 é o número de indivíduos da espécie 1 e y_2 o da espécie 2, então, de acordo com a seção anterior, o crescimento da espécie 1, na ausência da espécie 2, é controlado pela equação

$$\frac{d y_1}{dt} = y_1 (a_1 - b_1) \quad (1.4.1)$$

e o crescimento da espécie 2, na ausência da espécie 1, por

$$\frac{d y_2}{dt} = y_2 (a_2 - b_2) \quad (1.4.2)$$

Como as duas espécies estão vivendo conjuntamente e competindo pelos mesmos recursos, então o crescimento de cada população está inibido não somente pelos membros de sua própria espécie, mas também pelos membros da outra. Desta forma, a taxa de crescimento de ambas as espécies é reduzida. A expressão mais sim

ples que representa esta situação é obtida substituindo os fatores $a_1 - b_1 y_1$, em (1.4.1) por $a_1 - b_1 y_1 - c_1 y_2$ e $a_2 - b_2 y_2$, em (1.4.2), por $a_2 - b_2 y_2 - c_2 y_1$, ou seja

$$\begin{aligned} \frac{d y_1}{d t} &= y_1 (a_1 - b_1 y_1 - c_1 y_2) \\ \frac{d y_2}{d t} &= y_2 (a_2 - b_2 y_2 - c_2 y_1) \end{aligned} \tag{1.4.3}$$

onde o parâmetro c_1 mede o grau com que a presença da espécie 2 afeta o crescimento da espécie 1 e o oposto para c_2 .

1.5) COMPORTAMENTO DA SOLUÇÃO:

De um modo geral, as equações simultâneas (1.4.3) não podem ser explicitamente resolvidas. Desta forma, o estudo do comportamento das soluções de (1.4.3) será feito graficamente. Nos gráficos a composição da combinação das populações das duas espécies, num instante t , será representada por um ponto, tendo como coordenadas os valores $y_1(t)$ e $y_2(t)$.

Consideremos, primeiramente, o local dos pontos para os quais $\frac{d y_1}{d t} = 0$. Este local é a reta que corta o eixo- y_1 em $y_1 = a_1/b_1$ e o eixo- y_2 em $y_2 = a_1/c_1$. Das equações (1.4.3), observamos que a espécie 1 cresce quando $\frac{d y_1}{d t} > 0$ e decresce em caso contrário. De modo análogo, $\frac{d y_2}{d t} = 0$ é a reta que corta o eixo- y_1 em $y_1 = a_2/c_2$ e o eixo- y_2 em $y_2 = a_2/b_2$. Do mesmo modo, a espécie 2 cresce quando $d y_2/d t > 0$ e decresce no caso oposto. Estas situações estão mostradas nas figuras 2 e 3,

respectivamente.

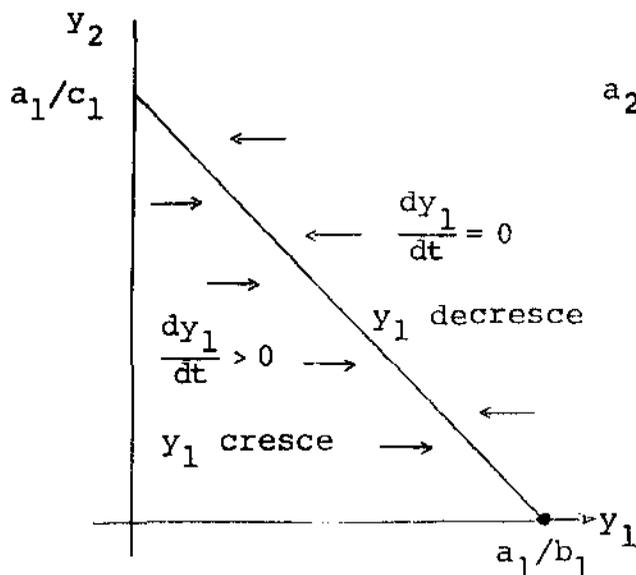


Figura 2

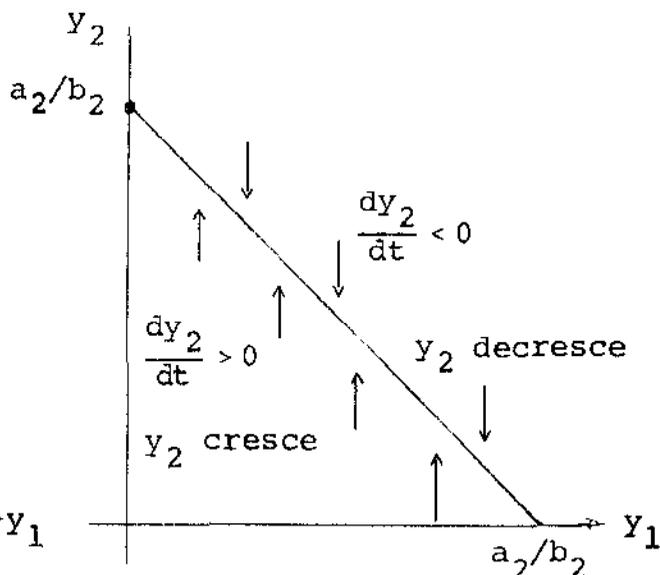


Figura 3

Para vermos o que acontece com as duas espécies simultaneamente, faremos a superposição dos gráficos da figura 2 e figura 3. Deste modo, observamos que, dependendo da posição relativa das retas $\frac{dy_1}{dt} = 0$ e $\frac{dy_2}{dt} = 0$, portanto dependendo dos parâmetros, existem quatro possibilidades. Analisaremos cada uma dessas possibilidades a seguir.

Consideremos, primeiramente, o caso em que a reta $\frac{dy_1}{dt} = 0$ está situada, inteiramente, acima da reta $\frac{dy_2}{dt} = 0$. Neste caso, temos que $a_1 b_2 > a_2 c_1$ e $a_1 c_2 > a_2 b_1$, e observamos que, para uma população combinada inicial abaixo destas retas, as espécies crescerão até atingirem um estágio em que somente a espécie 1 crescerá e a espécie 2 decrescerá. Na verdade, para qualquer população combinada inicial, a espécie 1 será a vencedora enquanto a espécie 2 será extinta. Isto está representado -

na figura 4 .

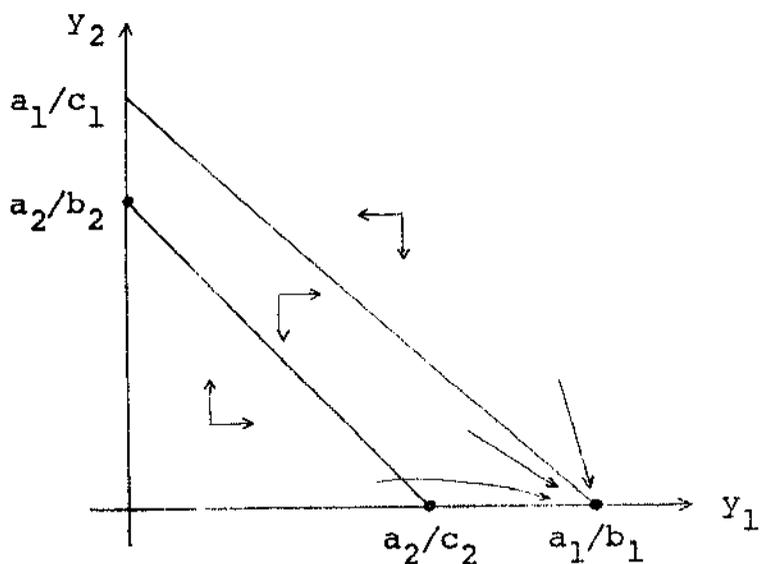


Figura 4

Veamos, agora, o caso em que $a_1 b_2 < a_2 c_1$ e $a_1 c_2 < a_2 b_1$. Neste caso, a reta $dy_2/dt = 0$ está acima da reta $dy_1/dt = 0$, como mostra a figura 5 e é fácil ver que a espécie 2 será a vencedora e a espécie 1 extinta.

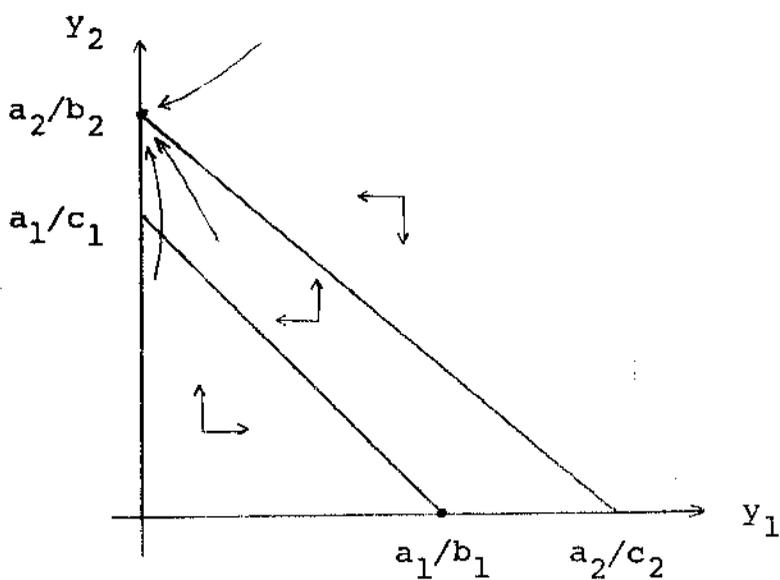


Figura 5

Nestes dois casos não existe um par de valores y_1 e y_2 , ambos positivos, que satisfaçam simultaneamente as equações $dy_1/dt = 0$ e $dy_2/dt = 0$. Portanto é impossível encontrar uma condição de equilíbrio, onde as populações permanecem constantes. Vejamos, agora, quando isto é possível.

Consideremos a situação em que as retas $dy_1/dt = 0$ e $dy_2/dt = 0$ se interceptam e seja E este ponto de interseção. Se $a_1 b_2 > a_2 c_1$ e $a_1 c_2 < a_2 b_1$, como mostra a figura 6, observamos que, independente da grandeza das populações iniciais $y_1(t_0)$ e $y_2(t_0)$, a população combinada aproxima-se do ponto E à medida que o tempo passa. Desta forma a população combinada atingirá um estágio, no ponto E , onde permanecerá inalterada indefinidamente. Aqui dizemos que a coexistência das espécies é possível.

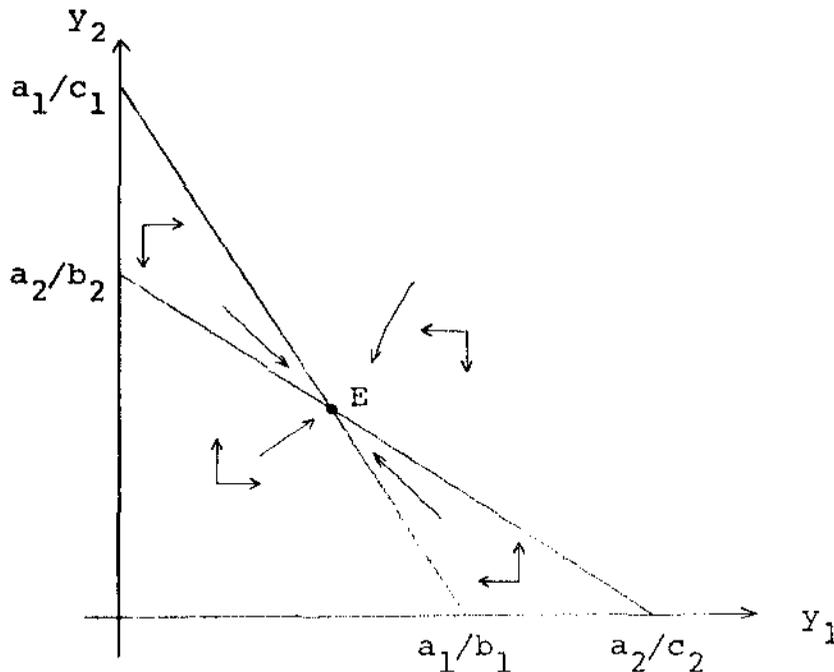


Figura 6

Se $a_1 b_2 < a_2 c_1$ e $a_1 c_2 > a_2 b_1$, como mostra a figura 7, observamos que, de um modo geral, qualquer população combinada atingirá um estágio em que uma das espécies será extinta. Portanto, o resultado da competição não depende somente dos parâmetros, mas também das populações iniciais $y_1(t_0)$ e $y_2(t_0)$. Na figura 7 mostramos isto para diferentes populações combinadas iniciais.

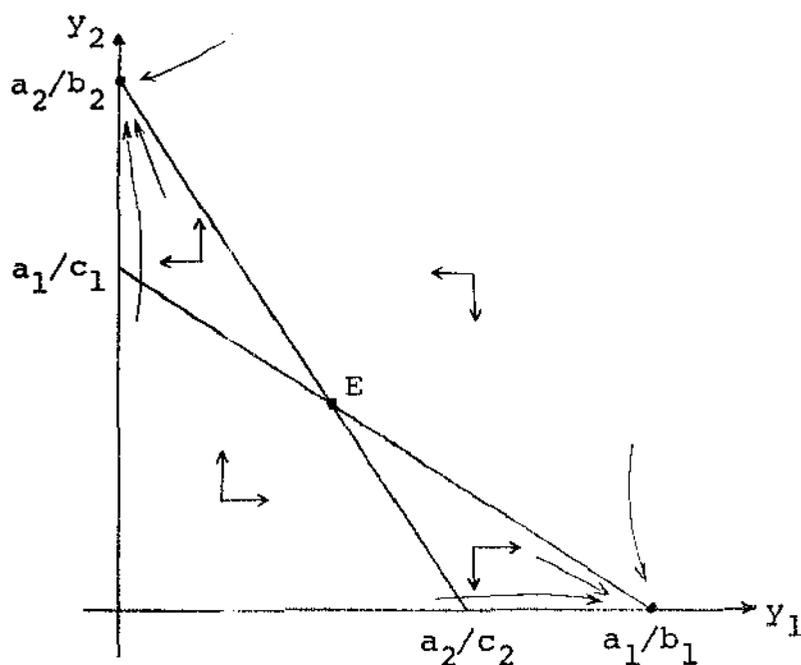


Figura 7

Dos quatro casos estudados, apenas o último apresenta dependência das populações iniciais. Neste caso, para determinar qual a espécie que será vencedora, ou extinta, precisamos conhecer as variações das populações com o tempo. Vejamos como podemos obter estas variações.

1.6) O MODELO DISCRETO:

Como já foi dito, as equações simultâneas (1.4.3) não

podem ser explicitamente resolvidas. Portanto é necessário encontrar equações de diferença que nos permitam prever as grandezas das populações no tempo $t+1$, dadas as populações no tempo t . Para isto, consideremos uma espécie y e sejam $y(t)$ e $y(t+1)$ as grandezas desta população nos instantes t e $t+1$. Como sabemos a "taxa de crescimento por unidade de tempo" é dada por

$$k = \frac{y(t+1) - y(t)}{y(t)}$$

a qual multiplicada por 100, fornece o percentual de variação da população na unidade de tempo considerada. Assim, se conhecemos a população no instante t , $y(t)$, e a taxa de crescimento k , podemos determinar a população no instante $t+1$ pela fórmula

$$y(t+1) = y(t) + y(t) \cdot k$$

Deste modo, as equações de diferença para as duas espécies em competição serão escritas na forma

$$y_1(t+1) = y_1(t) + y_1(t) \cdot k_1$$

$$y_2(t+1) = y_2(t) + y_2(t) \cdot k_2$$

onde k_1 e k_2 são as taxas de crescimento para as espécies 1 e 2, respectivamente. Visto que as taxas de crescimento k_1 e k_2 são dadas por

$$k_1 = a_1 - b_1 y_1(t) - c_1 y_2(t)$$

$$k_2 = a_2 - b_2 y_2(t) - c_2 y_1(t)$$

podemos, finalmente, escrever

$$\begin{aligned} y_1(t+1) &= y_1(t) + y_1(t)(a_1 - b_1 y_1(t) - c_1 y_2(t)) \\ y_2(t+1) &= y_2(t) + y_2(t)(a_2 - b_2 y_2(t) - c_2 y_1(t)) \end{aligned} \quad (1.6.1)$$

Assim, dados $y_1(0)$ e $y_2(0)$, podemos calcular $y_1(1)$, $y_2(1)$, $y_1(2)$, $y_2(2)$, ... e assim por diante. Observe nas equações (1.6.1), que se $k_1 > 0$ a espécie 1 cresce, se $k_1 < 0$ decresce e que se $k_1 = 0$ a espécie 1 permanece inalterada. O mesmo ocorre com a espécie 2 com relação a k_2 . Isto mostra que estas equações estão de acordo com o estudo do comportamento da solução mostrado anteriormente.

Nesta seção, mostramos o modelo discreto que nos permite calcular as variações das populações com o tempo. Vejamos agora, o problema inverso.

1.7) O PROBLEMA INVERSO:

Consideremos duas populações competindo pelos mesmos recursos limitados. Conhecidas as populações iniciais $y_1(0)$ e $y_2(0)$, determinar os parâmetros a_1, b_1, c_1 e a_2, b_2, c_2 os quais substituídos em (1.6.1), produzem soluções mais próximas quanto possível dos dados observados $\bar{y}_1(t+1)$ e $\bar{y}_2(t+1)$ ($t = 0, \dots, t-1$). Obviamente este problema é um problema de estimação de parâmetros o qual pode ser formulado como segue

$$\text{Min} \quad \sum_{t=0}^{t-1} [(y_1(t+1) - \bar{y}_1(t+1))^2 + (y_2(t+1) - \bar{y}_2(t+1))^2]$$

onde $y_1(t+1)$ e $y_2(t+1)$ são calculados através de (1.6.1). A resolução deste problema, bem como dos métodos utilizados para resolvê-lo serão abordados nos próximos capítulos.

CAPÍTULO II

MÉTODO DOS QUADRADOS MÍNIMOS LINEAR

2.1) INTRODUÇÃO:

A primeira idéia que ocorre para resolver o problema apresentado no capítulo anterior é aplicar o método dos quadrados mínimos linear já que, à primeira vista, o modelo parece linear. Entretanto, como veremos posteriormente, o modelo é não linear e portanto precisamos de métodos que levem em consideração as características da não linearidade do modelo. Tais métodos, porém, necessitam de uma boa estimativa inicial dos parâmetros para resolver o problema. Estas estimativas foram obtidas, para o problema em questão, pelo método dos quadrados mínimos linear.

2.2) MÉTODO DOS QUADRADOS MÍNIMOS LINEAR:

Consideremos o problema de encontrar os parâmetros x_1, x_2, \dots, x_n que minimizam

$$Q(x) = \sum_{i=1}^m (f_i(x) - \bar{y}_i)^2, \quad m \geq n \quad (2.2.1)$$

onde $x = (x_1, \dots, x_n)$ é o vetor dos parâmetros, \bar{y}_i os valores observados e f_i o modelo. Consideremos, ainda, que cada f_i é linear nos parâmetros, ou seja, f_i é da forma

$$f_i(x) = a_{i1} x_1 + \dots + a_{in} x_n$$

ou

$$f_i(x) = \sum_{j=1}^n a_{ij} x_j \quad (2.2.2)$$

Feitas estas considerações, vejamos como podemos encontrar a solução de (2.2.1). Substituindo (2.2.2) em (2.2.1) podemos escrever $Q(x)$ na forma

$$Q(x) = \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} x_j - \bar{y}_i \right)^2$$

cuja forma matricial é

$$Q(x) = \|Ax - \bar{y}\|^2 \quad (2.2.3)$$

onde $A = (a_{ij})$ é uma matriz $(m \times n)$ e \bar{y} é o vetor m -dimensional dos valores observados. Por conveniência expressaremos $Q(x)$ na forma

$$Q(x) = \langle Ax - \bar{y}, Ax - \bar{y} \rangle$$

ou ainda

$$Q(x) = x^t (A^t A) x - 2(A^t \bar{y}) x + \bar{y}^t \bar{y} \quad (2.2.4)$$

Da expressão acima, observamos que $Q(x)$ é uma quadrática nos parâmetros. Como $A^t A$ é positiva definida, exceto quando existem colunas proporcionais em A , podemos afirmar que $Q(x)$ tem um mínimo e ele satisfaz

$$\frac{\partial Q}{\partial x_j} = 0 \quad j = 1, 2, \dots, n$$

de onde obtemos o sistema linear

$$(A^t A) x = A^t \bar{y} \quad (2.2.5)$$

Portanto, a solução do problema (2.2.1) é dado pela solução deste sistema. As equações que constituem este sistema

são chamadas "equações normais".

2.3) INTERPRETAÇÃO GEOMÉTRICA:

Consideremos o problema (2.2.1) escrito na forma

$$\text{Min } \|Ax - \bar{y}\|^2$$

Deste modo, é fácil ver que o problema consiste em encontrar um vetor $y = Ax$ de tal modo que a distância entre ele e o vetor das observações \bar{y} seja a menor possível. Observando que o vetor estimado y pode ser escrito na forma

$$y = x_1 A_1 + \dots + x_n A_n \quad (2.3.1)$$

onde A_j ($j = 1, \dots, n$) são as colunas da matriz A , podemos notar que para diferentes valores dos x_{is} , os vetores estimados y estão localizados no subespaço, digamos H , determinado pelos vetores A_j . Assim, a norma $\|y - \bar{y}\|$ será mínima quando o vetor $(y - \bar{y})$ for ortogonal ao subespaço H , como ilustra a figura abaixo.

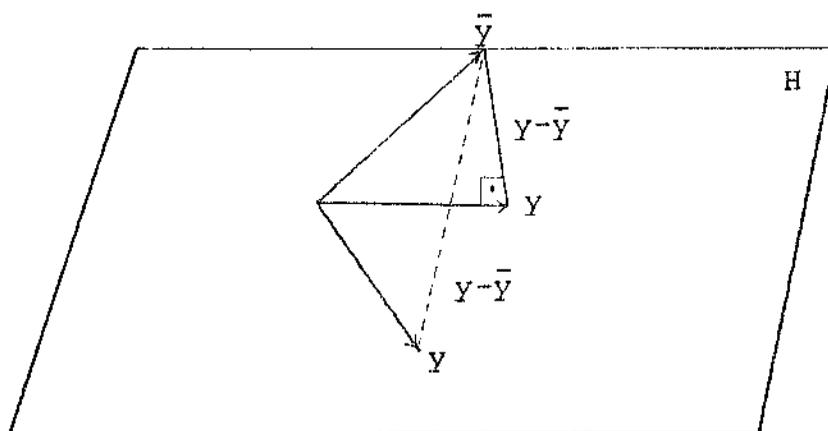


Figura 8

Visto que, um vetor é ortogonal a um subespaço, se e só se ele é ortogonal a todos os vetores deste subespaço, em particular aos vetores que o determinam, o vetor $(y - \bar{y})$ terá norma mínima desde que ele seja ortogonal a cada vetor A_j . Da definição de ortogonalidade segue que

$$\langle y - \bar{y}, A_j \rangle = 0 \quad j = 1, \dots, n$$

ou

$$\langle y, A_j \rangle = \langle \bar{y}, A_j \rangle \quad (2.3.2)$$

Substituindo (2.3.1) em (2.3.2) e fazendo j variar de 1 a n , obtemos

$$\begin{array}{rcl} x_1 \langle A_1, A_1 \rangle + \dots + x_n \langle A_n, A_1 \rangle & = & \langle \bar{y}, A_1 \rangle \\ \vdots & & \vdots \\ x_1 \langle A_1, A_n \rangle + \dots + x_n \langle A_n, A_n \rangle & = & \langle \bar{y}, A_n \rangle \end{array}$$

Estas equações são as equações normais apresentadas na seção anterior, cuja forma matricial é

$$(A^t A)x = A^t \bar{y}$$

2.4) APLICAÇÃO AO PROBLEMA:

O nosso problema consiste em encontrar os parâmetros a_{is} , b_{is} , c_{is} que minimizam

$$\sum_{t=0}^{t-1} [(y_1(t+1) - \bar{y}_1(t+1))^2 + (y_2(t+1) - \bar{y}_2(t+1))^2]$$

onde $\bar{y}_1(t+1)$, $\bar{y}_2(t+1)$ ($t = 0, \dots, t-1$) são os valores observa-

dos e $y_1(t+1)$ e $y_2(t+1)$ são os valores estimados pelo modelo

$$y_1(t+1) = y_1(t) + a_1 y_1(t) - b_1 y_1(t)^2 - c_1 y_1(t) y_2(t)$$

$$y_2(t+1) = y_2(t) + a_2 y_2(t) - b_2 y_2(t)^2 - c_2 y_1(t) y_2(t)$$

para dados valores de $y_1(0)$ e $y_2(0)$. Vejamos como podemos considerar este modelo linear. Fazendo $y_1(t) = \bar{y}_1(t)$ e $y_2(t) = \bar{y}_2(t)$ ($t = 0, \dots, t-1$), o modelo pode ser escrito como segue

$$y_1(t+1) = \bar{y}_1(t) + a_1 \bar{y}_1(t) - b_1 \bar{y}_1(t)^2 - c_1 \bar{y}_1(t) \bar{y}_2(t) \quad (2.4.1)$$

$$y_2(t+1) = \bar{y}_2(t) + a_2 \bar{y}_2(t) - b_2 \bar{y}_2(t)^2 - c_2 \bar{y}_1(t) \bar{y}_2(t)$$

Desta forma, o modelo deixa de ser recursivo e assim os valores estimados $y_1(t+1)$ e $y_2(t+1)$ ficam dependendo somente dos parâmetros, já que os valores $\bar{y}_1(t)$, $\bar{y}_2(t)$ ($t = 0, \dots, t$) são constantes conhecidas. É fácil ver que as expressões obtidas de (2.4.1), fazendo t variar de 0 a $t-1$, são lineares nos parâmetros. Deste modo, podemos formular o problema como segue

$$\text{Min } \|Ax - \bar{y}\|$$

onde

$$A = \begin{bmatrix} \bar{y}_1(0) & -\bar{y}_1(0)^2 & -\bar{y}_1(0)\bar{y}_2(0) & & & \\ \vdots & \vdots & \vdots & & & \\ \bar{y}_1(t-1) & -\bar{y}_1(t-1)^2 & -\bar{y}_1(t-1)\bar{y}_2(t-1) & & & \\ & & & \bar{y}_2(0) & -\bar{y}_2(0) & -\bar{y}_2(0)\bar{y}_1(0) \\ & & & \vdots & \vdots & \vdots \\ & & & \bar{y}_2(t-1) & -\bar{y}_2(t-1) & -\bar{y}_2(t-1)\bar{y}_1(t-1) \end{bmatrix}$$

$$\bar{y} = \begin{bmatrix} \bar{y}_1(1) - \bar{y}_1(0) \\ \vdots \\ \bar{y}_1(t) - \bar{y}_1(t-1) \\ \bar{y}_2(1) - \bar{y}_2(0) \\ \vdots \\ \bar{y}_2(t) - \bar{y}_2(t-1) \end{bmatrix}, \quad x = \begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ a_2 \\ b_2 \\ c_2 \end{bmatrix}$$

Assim, dados os valores $\bar{y}_1(t)$, $\bar{y}_2(t)$ ($t = 0, \dots, t$), determinamos a matriz A e o vetor \bar{y} , como definidos anteriormente, e o sistema linear

$$A^t Ax = A^t \bar{y}$$

que fornecerá a solução procurada.

CAPÍTULO III

MÉTODO DE LEVENBERG-MARQUARDT

3.1) INTRODUÇÃO:

Consideremos o problema proposto, o qual foi formulado como se segue

$$\text{Minimizar } \sum_{t=0}^{t-1} [(y_1(t+1) - \bar{y}_1(t+1))^2 + (y_2(t+1) - \bar{y}_2(t+1))^2]$$

onde $y_1(t+1)$ e $y_2(t+1)$ são calculados por

$$y_1(t+1) = y_1(t) + a_1 y_1(t) - b_1 y_1(t)^2 - c_1 y_1(t) y_2(t)$$

$$y_2(t+1) = y_2(t) + a_2 y_2(t) - b_2 y_2(t)^2 - c_2 y_1(t) y_2(t)$$

para dados valores de $y_1(0)$ e $y_2(0)$.

Em virtude deste modelo ser recursivo, observamos que $y_i(t+1)$ ($i = 1, 2$) não dependem somente dos parâmetros, mas também dos valores estimados $y_i(t)$ ($i = 1, 2$). Como $y_i(t)$ também dependem dos parâmetros, pois $y_i(t)$ são calculados do mesmo modo, os valores $y_i(t+1)$ são funções não lineares nos parâmetros. Assim, o problema proposto é um "problema de estimação de parâmetros não lineares".

Um dos métodos específicos para resolver este tipo de problema é o método de Levenberg-Marquardt ([14], [15]). Este método foi utilizado na resolução do nosso problema.

3.2) MÉTODO DE LEVENBERG-MARQUARDT:

O método de Levenberg-Marquardt fundamenta-se no método de Gauss-Newton [23] e no método do Gradiente [4]. O método de Gauss-Newton consiste em resolver, a cada iteração, um proble-

ma de quadrados mínimos linear obtido pela linearização do modelo, através de uma série de Taylor, onde são negligenciados os termos lineares. Assim, se $x^0 = (x_1^0, x_2^0, \dots, x_n^0)$ é uma estimativa inicial dos parâmetros que minimizam

$$F = \sum_{i=1}^m (\bar{y}_i - f_i(x))^2 \quad (3.2.1)$$

onde f_i é o modelo considerado, começamos expandindo f em série de Taylor como segue

$$f \cong f_0 + \left(\frac{\partial f}{\partial x_1} \right)_0 (x_1 - x_1^0) + \dots + \left(\frac{\partial f}{\partial x_n} \right)_0 (x_n - x_n^0)$$

ou

$$f \cong f_0 + \sum_{j=1}^n \left(\frac{\partial f}{\partial x_j} \right)_0 \delta_j^0 \quad (3.2.2)$$

onde $\delta_j^0 = x_j - x_j^0$ e o sub-índice "0" em f e $\frac{\partial f}{\partial x_j}$ indica os valores de f e $\frac{\partial f}{\partial x_j}$ calculados no ponto x^0 . Após a linearização do modelo, substituímos (3.2.2) em (3.2.1) e fazemos as derivadas parciais de F , em relação aos parâmetros, iguais a zero, ou seja:

$$\frac{\partial F}{\partial x_j} = 0 \quad j = 1, \dots, n \quad (3.2.3)$$

de onde obtemos um conjunto de n equações lineares correspondentes às equações normais, vistas no capítulo anterior. Fazendo $\bar{y}_i - f_i = E_i$, estas equações podem ser escritas, na forma matricial, como segue

$$(J_0^t J_0) \delta^0 = J_0^t E \quad (3.2.4)$$

onde

$$J_0 = \begin{bmatrix} \left(\frac{\partial f_1}{\partial x_1} \right)_0 & \dots & \left(\frac{\partial f_1}{\partial x_n} \right)_0 \\ \left(\frac{\partial f_m}{\partial x_1} \right)_0 & \dots & \left(\frac{\partial f_m}{\partial x_n} \right)_0 \end{bmatrix}$$

$$\delta^0 = \begin{bmatrix} x_1 - x_1^0 \\ x_n - x_n^0 \end{bmatrix}$$

Calculado o valor de δ^0 em (3.2.4), obtemos a nova estimativa pela seguinte fórmula

$$x^1 = x^0 + \alpha^0 \delta^0$$

Novas estimativas são obtidas repetindo-se o processo, ou seja, se x^r é a estimativa anterior, calculamos a matriz das derivadas parciais J_r , neste ponto, e definimos o sistema

$$(J_r^t J_r) \delta^r = J_r^t E \quad (3.2.5)$$

o qual fornece o valor de δ^r . Este valor é então usado para obter a nova estimativa pela seguinte fórmula:

$$x^{r+1} = x^r + \alpha^r \delta^r$$

A constante α é um fator de correção para δ , pois de outro modo a extrapolação pode estar além da região onde a li-

nearização representa adequadamente a função não linear.

O método do gradiente consiste em minimizar a função F , dada em (3.2.1), na direção do gradiente negativo ($-\nabla F$), visto que $-\nabla F$, num ponto qualquer, fornece a direção de máxima descida de F naquele ponto. A figura 9 mostra isto geometricamente no plano.

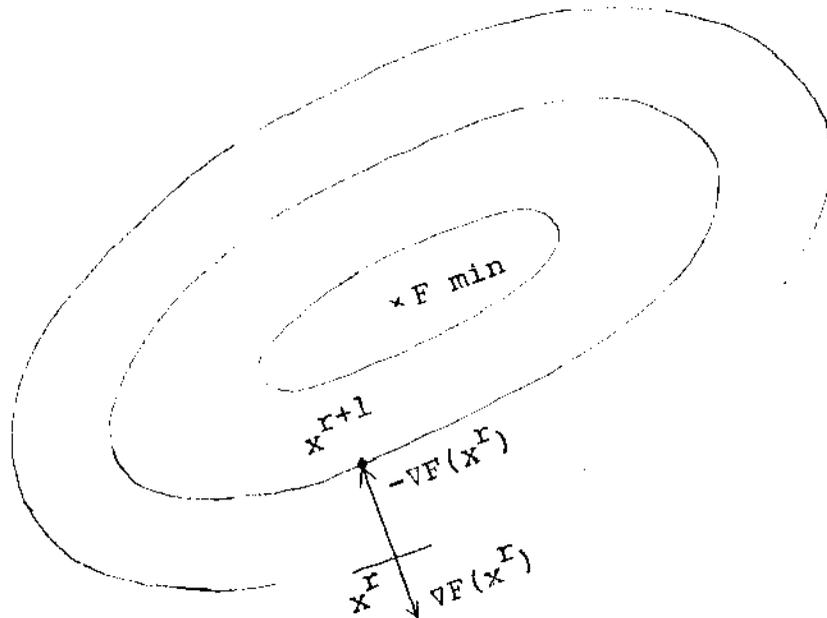


Figura 9

Se a função F tem um mínimo na região de busca, é possível reduzir o seu valor até que seja atingido este mínimo. Desta forma, se x^r é uma estimativa de parâmetros, podemos obter nova estimativa pela seguinte fórmula

$$x^{r+1} = x^r + \alpha^r d^r$$

onde $d^r = -\frac{\nabla F(x^r)}{\|\nabla F(x^r)\|}$ e α é o tamanho do passo.

Os dois métodos abordados apresentam algumas dificul-

dades práticas. No método de Gauss-Newton, a matriz $J^t J$ pode ser singular. Outra dificuldade encontrada neste método, é a escolha do fator α , pois uma escolha imprópria pode causar divergência nas iterações. O método do gradiente, por sua vez, pode levar um tempo muito grande para convergir. Isto ocorre pelo fato das curvas de nível de F não serem elípticas, mas deformadas e bastante alongadas como mostra a Figura 10.

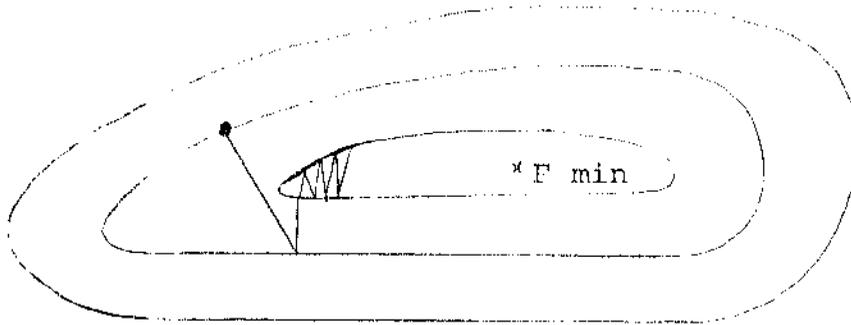
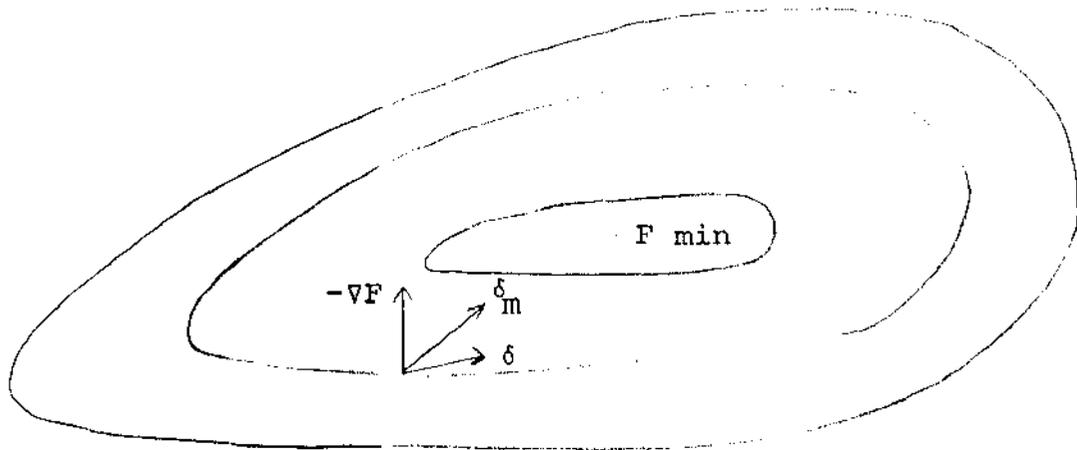


Figura 10

Fora estas dificuldades, estes métodos apresentam boas características. O método do gradiente apresenta a vantagem de convergir partindo de uma estimativa inicial a qual pode estar fora da região de convergência de outros métodos. O método de Gauss-Newton fornece sempre direção de descida, se $J^t J$ é positiva definida, além de atingir a convergência quadrática na resolução de problemas onde $F = 0$ na solução. Assim, Marquardt propôs um método que conserva estas características enquanto evita as dificuldades citadas.

Marquardt observou, na prática, que a direção do método do gradiente é aproximadamente ortogonal a direção do método de Gauss-Newton. Considerando que qualquer método, para resolver este tipo de problema, precisa de uma direção que esteja dentro de um ângulo de 90° com o gradiente negativo, pois de outro modo os valores de F serão maiores que os desejados, ele sugeriu uma combinação destes métodos. A figura 11 ilustra esta situação geometricamente no plano.



- δ - direção do método de Gauss-Newton.
 δ_m - direção do método de Marquardt.

Figura 11

Vejam agora, como podemos obter esta combinação. Consideremos a equação (3.2.5) obtida no método de Gauss-Newton, ou seja

$$(J^t J) \delta = J^t E$$

Suponhamos que a matriz diagonal λI , ($\lambda \geq 0$) é adicionada à matriz $J^t J$. Assim a equação (3.2.5) será escrita na forma

$$(J^t J + \lambda I) \delta = J^t E \quad (3.2.6)$$

Quando $\lambda = 0$, esta equação é idêntica a equação (3.2.5) e, portanto, a direção δ é calculada do mesmo modo que no método de Gauss-Newton. Quando $\lambda \rightarrow \infty$ a matriz $(J^t J + \lambda I)$ é dominada pela diagonal λI . Assim, δ é calculada, essencialmente, como

$$\delta = \frac{1}{\lambda} (J^t E)$$

Visto que $J^t E = -\nabla F$, podemos escrever

$$\delta = \frac{1}{\lambda} (-\nabla F)$$

Deste modo, a direção δ é equivalente a direção dada pelo método do gradiente. Portanto, a equação (3.2.6) tem como casos limites o método de Gauss-Newton e o Método do Gradiente. Para valores intermediários de λ , Marquardt [15] prova que a direção obtida está entre a direção do método do Gradiente e a direção do método de Gauss-Newton. Deste modo, a equação (3.2.6) representa a composição destes métodos.

Como vimos, o escalar λ é de máxima importância no método de Marquardt, visto que o mesmo determina a direção deste método. Marquardt prova na referência [15] que $\|\delta\|$ é uma função monótona decrescente de λ . Assim, o propósito do método é escolher um λ de tal modo que se extenda o máximo possível a região de validade da série de Taylor, utilizada no método de Gauss-Newton, mantendo a direção próxima a direção do gradiente para evitar divergências. Para isso Marquardt propõe a seguinte estratégia para a escolha de λ .

Seja $v > 1$ ($v = 10$) e $\lambda^{(r-1)}$ o valor de λ na

iteração anterior ($\lambda^{(0)} = 10^{-2}$). Calcule $F(\lambda^{(r-1)})$ e $F(\lambda^{(r-1)}/v)$. Existem três condições que controlam a escolha de λ , a saber:

- 1) Se $F(\lambda^{(r-1)}/v) < F^{(r)}$, então faça $\lambda^{(r)} = \lambda^{(r-1)}/v$.
- 2) Se $F(\lambda^{(r-1)}/v) > F^{(r)}$ e $F(\lambda^{(r-1)}) < F^{(r)}$, então faça $\lambda^{(r)} = \lambda^{(r-1)}$.
- 3) Se $F(\lambda^{(r-1)}/v) > F^{(r)}$ e $F(\lambda^{(r-1)}) > F^{(r)}$, então faça $\lambda^{(r)} = \lambda^{(r-1)} \cdot v^k$, onde k é o menor inteiro positivo que satisfaz $F(\lambda^{(r-1)} \cdot v^k) \leq F^{(r)}$.

Entretanto esta estratégia, conforme Fletcher [26], é pouco eficiente. Fletcher cita, por exemplo, que a redução de λ , usando $v = 10$ como recomenda Marquardt, pode ser excessiva, causando mais de duas avaliações de F por iteração, além de impedir que seja atingida a convergência quadrática na resolução de problema onde $F = 0$ na solução. Para eliminar dificuldades como estas, Fletcher propôs uma estratégia para a escolha de λ . Esta estratégia é similar a usada por Marquardt, ou seja, consiste em aumentar ou reduzir λ multiplicando por v ou $1/v$, conforme seja necessário, para obter $F(x^{(r)} + \delta) < F(x^R)$. O critério utilizado para determinar quando λ deve ser aumentado ou reduzido é dado pela razão R entre a redução real da soma dos quadrados $F(x^R) - F(x^R + \delta)$ e a redução estimada da soma dos quadrados pelo modelo linearizado. Assim, se $\beta(x^R)$ representa a soma estimada dos quadrados, obtemos a redução estimada pela fórmula

$$\vartheta(x^r) - \vartheta(x^r + \delta) = -2\delta^t v - \delta^t A \delta$$

onde $v = J_r^t E_r$ e $A = J_r^t J_r$. Então calculamos R pela fórmula

$$R = \frac{F(x^r) - F(x^r + \delta)}{\vartheta(x^r) - \vartheta(x^r + \delta)}$$

Se R é um valor próximo de 1, λ é reduzido e se R é um valor próximo ou menor que zero, λ é aumentado. Na prática, entretanto, foi observado que, para valores intermediários, seria melhor manter o valor de λ para a próxima iteração. Assim, decidiu-se escolher constantes arbitrárias ρ e σ tais que $0 < \rho < \sigma < 1$ e reduzir λ se $R > \sigma$ e aumentar λ se $R < \rho$. Os valores de ρ e σ recomendados por Fletcher são $\rho = 0.25$ e $\sigma = 0.75$.

O método para aumentar λ é similar ao usado por Marquardt, multiplicando λ por v , mas com a diferença de que o fator v não é fixo. Foi observado, na prática, que nas primeiras iterações o valor $v = 2$ era adequado, entretanto quando λ era pequeno, um fator grande, dito $v = 10$, era necessário. Assim, decidiu-se escolher um fator v entre 2 e 10. Devido ao fato de que, para valores grandes de λ , crescer λ para $v\lambda$ corresponde aproximadamente a reduzir δ para δ/v , e como é possível estimar uma correção ótima $\alpha\delta$ na direção δ pela fórmula

$$\alpha = 1 / (2 - (F(x^r + \delta) - F(x^r)) / \delta^t v^r),$$

o fator $v = 1/\alpha$ foi escolhido para aumentar λ . Este fator é substituído por 2 ou 10 conforme seja menor que 2 ou maior que 10, respectivamente.

Para reduzir λ , Fletcher usa a mesma idéia de Marquardt, multiplicando λ por $1/v$, mas usando $v = 2$ e definindo adicionalmente um valor λ_c tal que qualquer valor reduzido $\lambda < \lambda_c$ é substituído por $\lambda = 0$. O valor λ_c foi escolhido de modo a satisfazer $\|\delta(\lambda_c)\| / \|\delta(0)\| = 1/2$. Fletcher demonstra que uma estimativa deste λ_c pode ser encontrada pela fórmula

$$\lambda_c = 1 / \|A^{-1}\|$$

Como o cálculo de λ_c exige a avaliação de A^{-1} , o que é dispendioso, Fletcher propõe que λ_c seja recalculado somente quando for necessário crescer λ de zero para algum número positivo. Isto evita o cálculo repetido de λ_c na estratégia de redução, como também favorece o uso da estratégia de crescimento, quando $\lambda=0$ na iteração anterior.

Em resumo podemos descrever a estratégia de Fletcher como se segue:

- 1) Inicie com $\lambda = 0$.
- 2) Resolva $(A+\lambda I)\delta = v$ para δ e avalie $F' = F(x+\delta)$.
- 3) Calcule R .
- 4) Se $R > \rho$, calcule α e faça $v = 1/\alpha$ ($2 \leq v \leq 10$). Se $\lambda = 0$, calcule λ_c e faça $\lambda = (v/2) \cdot \lambda_c$. Em caso contrário, faça $\lambda = v\lambda$.
- 5) Se $\rho \leq R \leq \sigma$, mantenha o valor de λ da iteração anterior.
- 6) Se $R > \sigma$, faça $\lambda = \lambda/2$. Se $\lambda < \lambda_c$ faça $\lambda = 0$.
- 7) Se $F' < F(x)$, faça $x = x+\delta$ e recalcule A e v . Em caso contrário volte ao passo 2.

A equação $(A+\lambda I)\delta = v$ pode ser alternativamente substituída pela equação

$$(A + \lambda D)\delta = v$$

onde D é uma matriz diagonal, cujos elementos D_{ii} são os elementos da diagonal principal da matriz A . Ball [24] mostrou que a substituição da matriz I por D , como definida acima, é exatamente equivalente ao escalamento proposto por Marquardt para melhorar os aspectos numéricos dos processos computacionais. Marquardt define a matriz escalada A^* e o vetor escalado v^* como segue

$$\begin{aligned} A^* &= (a_{ij}^*) = (\epsilon_i \epsilon_j a_{ij}) \\ v^* &= (v_j^*) = (\epsilon_j v_j) \end{aligned}$$

onde $\epsilon_j = (a_{jj})^{-1/2}$ é o fator escalar. Em virtude desta mudança de escala, os elementos do vetor δ são dados por

$$\delta_j = \epsilon_j \delta_j^*$$

onde δ^* é a solução de

$$(A^* + \lambda I)\delta^* = v^* .$$

Na verdade, a matriz D pode ser tomada como qualquer matriz diagonal com $D_{ii} > 0$ para todo i . Fletcher usa na implementação de sua estratégia a matriz D como a diagonal de A , avaliada no ponto inicial x^0 , embora deixe como opção ao usuário a utilização de $D = I$ ou outra matriz diagonal qualquer. Um estudo mais detalhado da escolha da matriz D é dado por Moré, J. na referência [25].

3.3) APLICAÇÃO AO PROBLEMA:

Para aplicarmos o método de Marquardt necessitamos das derivadas parciais do modelo e de uma estimativa inicial dos parâmetros com o propósito de obter a matriz $A = J^t J$ e o vetor $v = J^t E$.

A estimativa inicial, para o problema em questão, foi obtida através do método dos quadrados mínimos linear, do modo abordado no capítulo anterior. As derivadas parciais do modelo em relação aos parâmetros foram obtidas do modo descrito abaixo.

Consideremos o modelo proposto para o nosso problema, ou seja,

$$\begin{aligned} y_1(t+1) &= y_1(t) + a_1 y_1(t) - b_1 y_1(t)^2 - c_1 y_1(t) y_2(t) \\ y_2(t+1) &= y_2(t) + a_2 y_2(t) - b_2 y_2(t)^2 - c_2 y_2(t) y_1(t) \quad (3.3.1) \\ y_1(0) &\text{ e } y_2(0) \text{ dados.} \end{aligned}$$

Por conveniência, denotaremos os parâmetros a_1, b_1, c_1 e a_2, b_2, c_2 por x_1, x_2, x_3 e x_4, x_5, x_6 respectivamente. Como foi frisado no início deste capítulo, $y_1(t+1)$ e $y_2(t+1)$ são funções não apenas dos parâmetros, mas também de $y_1(t)$ e $y_2(t)$. Assim podemos escrever

$$\begin{aligned} y_1(t+1) &= G_1(y_1(t), y_2(t), x) \\ y_2(t+1) &= G_2(y_1(t), y_2(t), x) \end{aligned}$$

onde x denota o vetor dos parâmetros e G_1 e G_2 denotam as expressões do lado direito em (3.3.1). Visto que $y_1(t)$ e $y_2(t)$

também dependem dos parâmetros, podemos obter as derivadas parciais em relação a um dado parâmetro, digamos x_j , usando a regra da cadeia como segue

$$\frac{\partial y_1(t+1)}{\partial x_j} = \frac{\partial G_1}{\partial y_1(t)} \cdot \frac{\partial y_1(t)}{\partial x_j} + \frac{\partial G_1}{\partial y_2(t)} \cdot \frac{\partial y_2(t)}{\partial x_j} + \frac{\partial G_1}{\partial x_j} \quad (3.3.2)$$

$$\frac{\partial y_2(t+1)}{\partial x_j} = \frac{\partial G_2}{\partial y_1(t)} \cdot \frac{\partial y_1(t)}{\partial x_j} + \frac{\partial G_2}{\partial y_2(t)} \cdot \frac{\partial y_2(t)}{\partial x_j} + \frac{\partial G_2}{\partial x_j}$$

As derivadas $\partial G_1/\partial y_i(t)$, $\partial G_2/\partial y_i(t)$ ($i = 1, 2$) e $\partial G_1/\partial x_j$, $\partial G_2/\partial x_j$ ($j = 1, \dots, 6$) são obtidas diretamente de (3.3.1). Temos então:

$$\frac{\partial G_1}{\partial y_1(t)} = 1 + x_1 - 2x_2 y_1(t) - x_3 y_2(t)$$

$$\frac{\partial G_2}{\partial y_1(t)} = -x_6 y_2(t)$$

$$\frac{\partial G_1}{\partial y_2(t)} = -x_3 y_1(t)$$

$$\frac{\partial G_2}{\partial y_2(t)} = 1 + x_4 - 2x_5 y_2(t) - x_6 y_1(t)$$

$$\frac{\partial G_1}{\partial x_1} = y_1(t)$$

$$\frac{\partial G_2}{\partial x_1} = \frac{\partial G_2}{\partial x_2} = \frac{\partial G_2}{\partial x_3} = 0$$

$$\frac{\partial G_1}{\partial x_2} = -y_1(t)^2$$

$$\frac{\partial G_2}{\partial x_4} = y_2(t)$$

$$\frac{\partial G_1}{\partial x_3} = -y_1(t) y_2(t)$$

$$\frac{\partial G_2}{\partial x_5} = -y_2(t)^2$$

$$\frac{\partial G_1}{\partial x_4} = \frac{\partial G_1}{\partial x_5} = \frac{\partial G_1}{\partial x_6} = 0$$

$$\frac{\partial G_2}{\partial x_6} = -y_2(t) y_1(t)$$

Assim, para obtermos $\partial y_1(t+1)/\partial x_j$ e $\partial y_2(t+1)/\partial x_j$ ($t=0, \dots, t-1$) precisamos resolver (3.3.2) conjuntamente com (3.3.1). Para isto precisamos dos valores iniciais $\partial y_1(0)/\partial x_j$ e $\partial y_2(0)/\partial x_j$. Visto que $y_1(0)$ e $y_2(0)$ são constantes, podemos então escrever

$$\frac{\partial y_1(0)}{\partial x_j} = 0 \quad \text{e} \quad \frac{\partial y_2(0)}{\partial x_j} = 0 \quad (j = 1, \dots, 6)$$

Assim, dados $y_1(0)$ e $y_2(0)$, resolvemos (3.3.1) conjuntamente com (3.3.2) e obtemos $\frac{\partial y_1(t+1)}{\partial x_j}$ e $\frac{\partial y_2(t+1)}{\partial x_j}$ ($t = 0, \dots, t-1$) recursivamente.

CAPÍTULO IV
MÉTODO DE BREMERMANN

4.1) INTRODUÇÃO:

Uma das dificuldades encontradas na resolução de certos problemas de estimação de parâmetros não lineares é o cálculo das derivadas. Em alguns problemas, por exemplo, onde os modelos são dados por equações diferenciais ordinárias, este cálculo não somente é difícil como também dispendioso. Em vista disto, vários métodos que dispensam o cálculo das derivadas ([16], [17], [18], [19]) têm sido usados para resolver este tipo de problema. Em particular, na estimação de parâmetros em modelos biológicos, o método de Bremermann [20] tem se mostrado eficiente conforme Swartz-Bremermann [22]. Assim, este método também foi usado para resolver o problema proposto.

4.2) MÉTODO DE BREMERMANN:

O método de Bremermann se propõe a resolver o seguinte tipo de problema:

$$\text{Min } F(x) , \quad F : \mathbb{R}^n \rightarrow \mathbb{R} \quad (4.2.1)$$

Este método consiste, basicamente, em resolver, a cada iteração, o seguinte problema

$$\begin{aligned} &\text{Min } F(x) \\ &\text{s/a } x^r + \lambda d , \quad \lambda \in \mathbb{R} \end{aligned} \quad (4.2.2)$$

onde x^r é uma estimativa da solução de (4.2.1) e d uma direção aleatória, cujas componentes são obtidas com distribuição -

gaussiana. Para resolver o problema (4.2.2), o método aproxima a função F , na restrição $x^r + \lambda d$, por um polinômio de grau 4, o qual é obtido pela interpolação Lagrangiana de F , através de cinco pontos igualmente espaçados com distância h e centro em x^r . O valor de λ , dito x^r , que minimiza este polinômio, é então usado para determinar a nova estimativa pela seguinte fórmula

$$x^{r+1} = x^r + \lambda^r h d$$

Nova direção d é obtida e o processo é repetido a partir de x^{r+1} se $F(x^{r+1}) < F(x^r)$ ou a partir de x^r se a desigualdade não é satisfeita. A figura 12 ilustra o processo no plano.

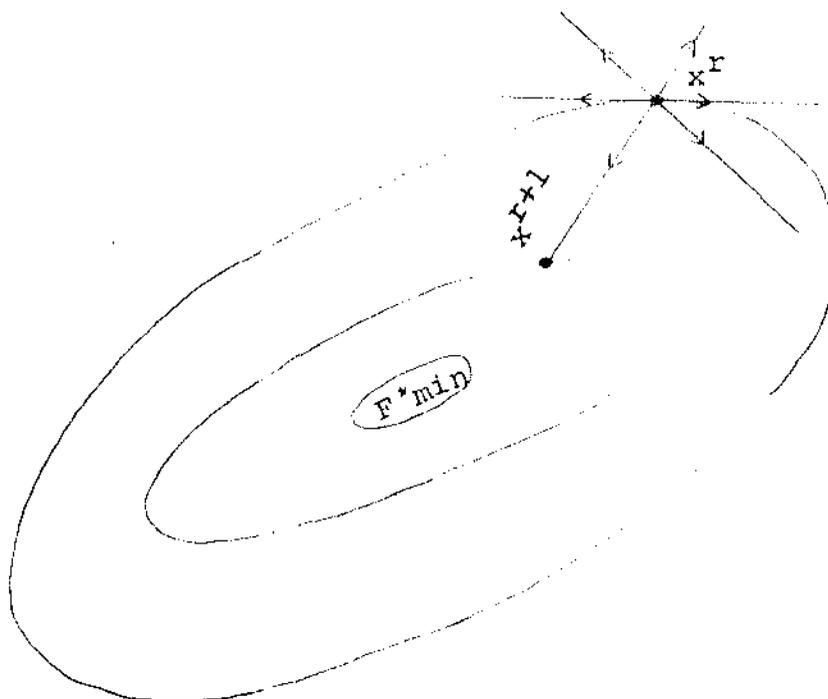


Figura 12

Este método tem convergência garantida para polinômios de várias variáveis de, no máximo, grau 4. A discussão das propriedades de convergência, bem como um programa em FORTRAN do método, podem ser encontrados na referência [20]. Daremos agora,

a descrição do algoritmo utilizado por Bremermann na implementação do método. O algoritmo consiste dos seguintes passos:

- 1) Avalie F na estimativa inicial x^0 .
- 2) Obtenha uma direção aleatória d , gerando as componentes d_{is} com distribuição gaussiana.
- 3) Avalie F nos pontos $x^0 - 2hd$, $x^0 - hd$, x^0 , $x^0 + hd$ e $x^0 + 2hd$. A constante h é um parâmetro do método fornecido pelo usuário.
- 4) Calcule os cinco coeficientes do polinômio de grau 4, $P(\lambda)$, através da interpolação Lagrangiana.
- 5) Calcule as raízes da derivada deste polinômio pela fórmula de Cardan.
- 6) Escolha a raiz λ que produz o menor valor de $p(\lambda)$. Se existe apenas uma raiz, considere esta como o mínimo de $p(\lambda)$.
- 7) Faça $x = x^0 + \lambda hd$ e avalie $F(x)$.
- 8) Se $F(x) < F(x^0)$, faça $x^0 = x$ e volte ao passo 2. Em caso contrário, mantenha o ponto x^0 e volte ao passo 2.
- 9) Após um número determinado de iterações, pare o processo.

O critério de parada dado pelo algoritmo pode ser alternativamente substituído. Em particular, para problemas de estimação de parâmetros, Swartz-Bremermann [22] sugerem que o processo seja parado, após F permanecer aproximadamente constante para várias iterações. Este critério foi utilizado nas experiências numéricas.

4.3) APLICAÇÃO AO PROBLEMA:

Para aplicarmos o método de Bremermann, abordamos o problema proposto, como um problema geral de otimização sem restrição. Assim, definimos a função a ser minimizada como segue

$$F(x) = \sum_{t=0}^{t-1} [(y_1(t+1) - \bar{y}_1(t+1))^2 + (y_2(t+1) - \bar{y}_2(t+1))^2]$$

onde x denota o vetor dos parâmetros. Os valores $y_1(t+1)$ e $y_2(t+1)$ são calculados pelo modelo proposto para dados valores de $y_1(0)$ e $y_2(0)$.

O processo utilizado para encontrar a solução deste problema consiste dos seguintes passos:

- 1) Obtemos uma boa estimativa inicial dos parâmetros pelo método dos quadrados mínimos linear.
- 2) Aplicamos a técnica de otimização de Bremermann para melhorar esta estimativa.
- 3) Iteramos o processo até que F permaneça aproximadamente constante para várias iterações.

CAPÍTULO V
EXPERIÊNCIAS NUMÉRICAS

5.1) INTRODUÇÃO:

Para testar a eficiência dos métodos anteriormente descritos é necessário conhecermos os valores reais dos parâmetros, o que não é possível com dados experimentais. Assim escolhemos um sistema teste com os parâmetros conhecidos de onde geramos dados artificiais. Estes dados foram perturbados, simulando erros aleatórios, pela seguinte fórmula

$$\bar{y}_i(t) = y_i(t) (1 + (R-0.5) \cdot 2 \cdot P/100) \quad i = 1,2$$

onde $y_i(t)$ é o valor encontrado pelo modelo proposto, $\bar{y}_i(t)$ o valor perturbado e R é um número aleatório entre 0 e 1 gerado com distribuição uniforme. Assim, se a máxima porcentagem de erro para um dado experimento é P , então o erro em cada ponto estará entre $P/100$ e $-P/100$ vezes o ponto dado.

As experiências numéricas foram realizadas com as subrotinas LLSQF (IMSL [28]) para o método dos quadrados mínimos linear e EO4GAF (NAG [29]) para o método de Marquardt. O programa para o método de Bremermann foi obtido no apêndice da referência [20]. Todos os testes foram realizados no computador PDP-10 da UNICAMP.

5.2) RESULTADOS DAS EXPERIÊNCIAS NUMÉRICAS:

O sistema teste escolhido para as experiências numéricas, considerado em Pielou [1] é o seguinte

$$y_1(t+1) = y_1(t) + 0.1y_1(t) - 0.0007y_1(t)^2 - 0.001y_1(t)y_2(t)$$

$$y_2(t+1) = y_2(t) + 0.075y_1(t) - 0.0007y_2(t)^2 - 0.0007y_2(t)y_1(t)$$

$$y_1(0) = 4$$

$$y_2(0) = 10$$

A Figura 13 mostra a variação da população com o tempo no intervalo de 0 a 50, obtida pelo modelo.

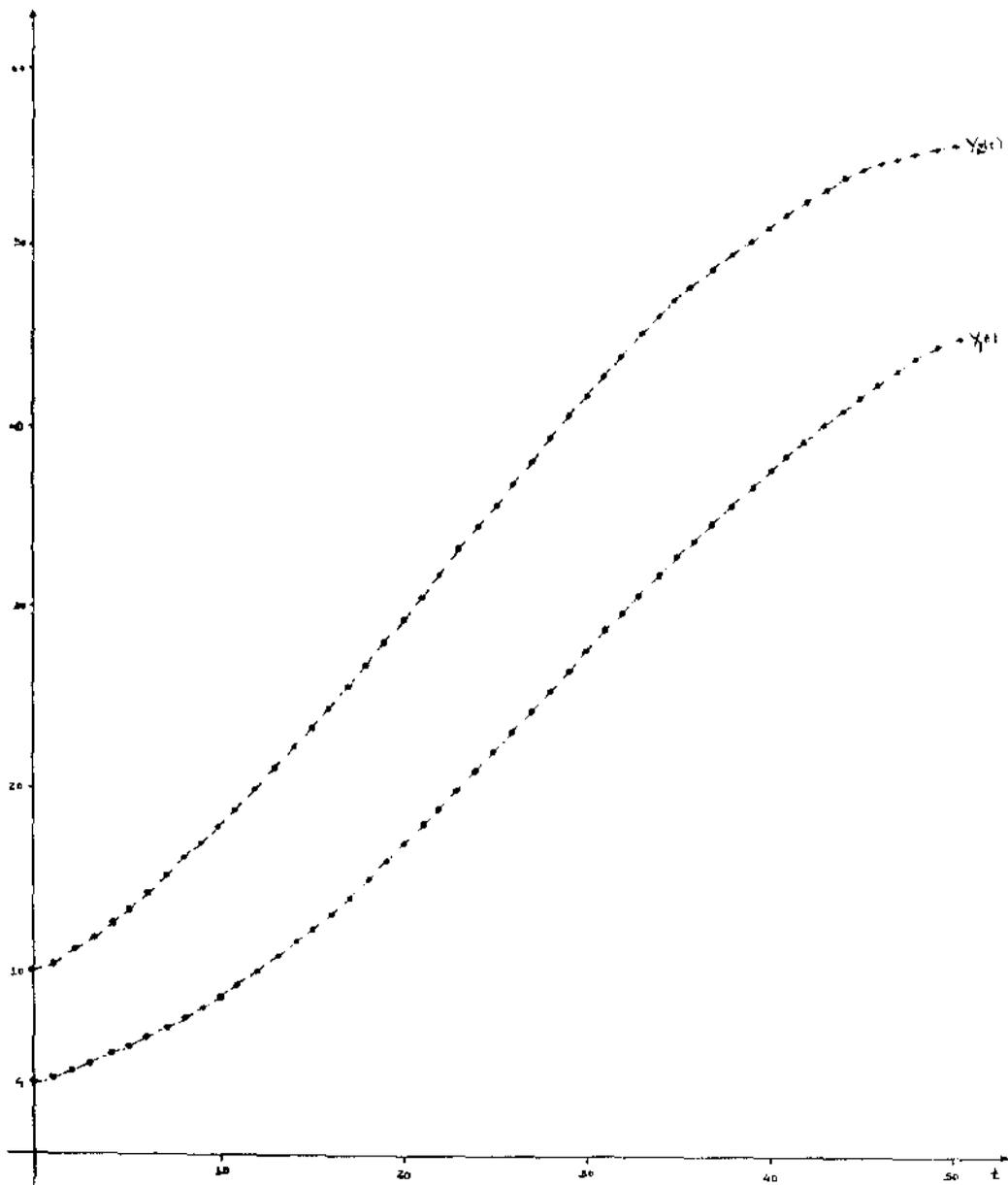


Figura 13

Na Tabela 1, mostramos os resultados obtidos com os dados sem perturbação.

PERTURBAÇÃO DE 0%			
PARÂMETROS	QUADRADOS MÍNIMOS LINEAR	MÉTODO DE LEVENBERG-MARQUARDT	MÉTODO DE BREMERMAN
x_1	0.1000000E+00	0.1000000E+00	0.1000000E+00
x_2	0.6999999E-03	0.6999995E-03	0.7000000E-03
x_3	0.1000000E-02	0.1000000E-02	0.1000000E-02
x_4	0.7499999E-01	0.7500000E-01	0.7499999E-01
x_5	0.6999994E-03	0.6999997E-03	0.6999991E-03
x_6	0.7000006E-03	0.7000003E-03	0.7000010E-03
SOMA DOS QUADRADOS	0.2640391E-08	0.1446772E-09	0.5073275E-11

Tabela 1

Desta tabela, observamos que os resultados apresentados pelos três métodos são bons. Em particular, a solução obtida pelo método dos quadrados mínimos linear já pode ser considerada a solução do problema, pois apresenta ótima precisão. Isto pode nos levar, a princípio, pensar que o problema pode ser abordado como um problema linear nos parâmetros. Entretanto, quando os dados são perturbados isto não ocorre. A Tabela 2, a seguir, mostra isto com os dados perturbados de 5%.

Nesta tabela, observamos que o resíduo produzido pela solução do método dos quadrados mínimos linear é bem maior que os resíduos produzidos pelas soluções dos métodos de Marquardt e Bremermann. Na verdade, à medida que aumentamos a perturbação, obte

PERTURBAÇÃO DE 5%			
PARÂMETROS	QUADRADOS MÍNIMOS LINEAR	MÉTODO DE LEVENBERG- MARQUARDT	MÉTODO DE BREMERMAN
x_1	0.1685271	0.9780967E-01	0.1683176
x_2	-0.3571623E-02	0.8127656E-03	-0.4606315E-02
x_3	0.5596261E-02	0.8744583E-03	0.6665066E-02
x_4	0.1202774	0.7351701E-01	0.1203260
x_5	0.4092217E-02	0.6467602E-03	0.4724305E-02
x_6	-0.2588738E-02	0.7303711E-03	-0.3168648E-02
SOMA DOS QUADRADOS	8052.598	68.15652	474.2981

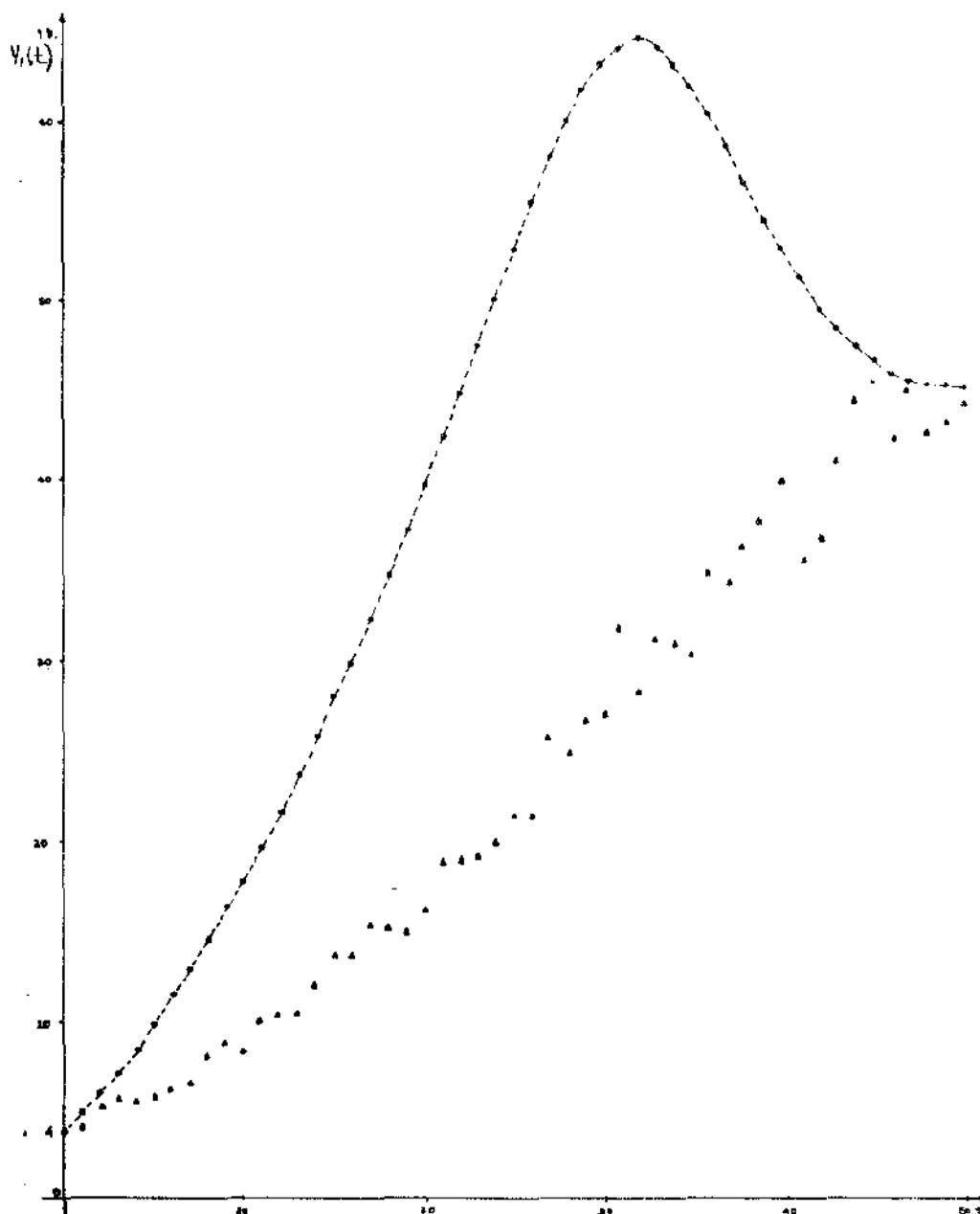
Tabela 2

mos um maior resíduo para este método, comparando com os métodos de Marquardt e de Bremermann. A Tabela 3 confirma isto, onde os dados foram perturbados de 10%.

PERTURBAÇÃO DE 10%			
PARÂMETROS	QUADRADOS MÍNIMOS LINEAR	MÉTODO DE LEVENBERG- MARQUARDT	MÉTODO DE BREMERMAN
	0.3418535	0.9571299E-01	0.3425742
	-0.1539097E-01	0.9118345E-03	-0.1547208E-01
	0.1795254E-01	0.7597689E-03	0.1960321E-01
	0.2512925	0.7177553E-01	0.2517152
	0.1441242E-01	0.5622904E-03	0.1476208E-01
	-0.1282540E-01	0.7959696E-03	-0.1172059E-01
SOMA DOS QUADRADOS	37866.92	272.6257	2107.209

Tabela 3

Dos resultados apresentados, fica fácil ver que os métodos de Marquardt e de Bremermann, onde são levados em conta as características não lineares do problema, apresentam melhores resultados. Observamos, também, que o método de Marquardt é o que apresenta o menor resíduo em todos os casos onde os dados foram perturbados. Para que fique mais claro a comparação dos métodos citados, mostramos nas Figuras 14, 15, 16, 17, 18 e 19 as curvas obtidas pelas soluções destes métodos no caso da perturbação de 10%.

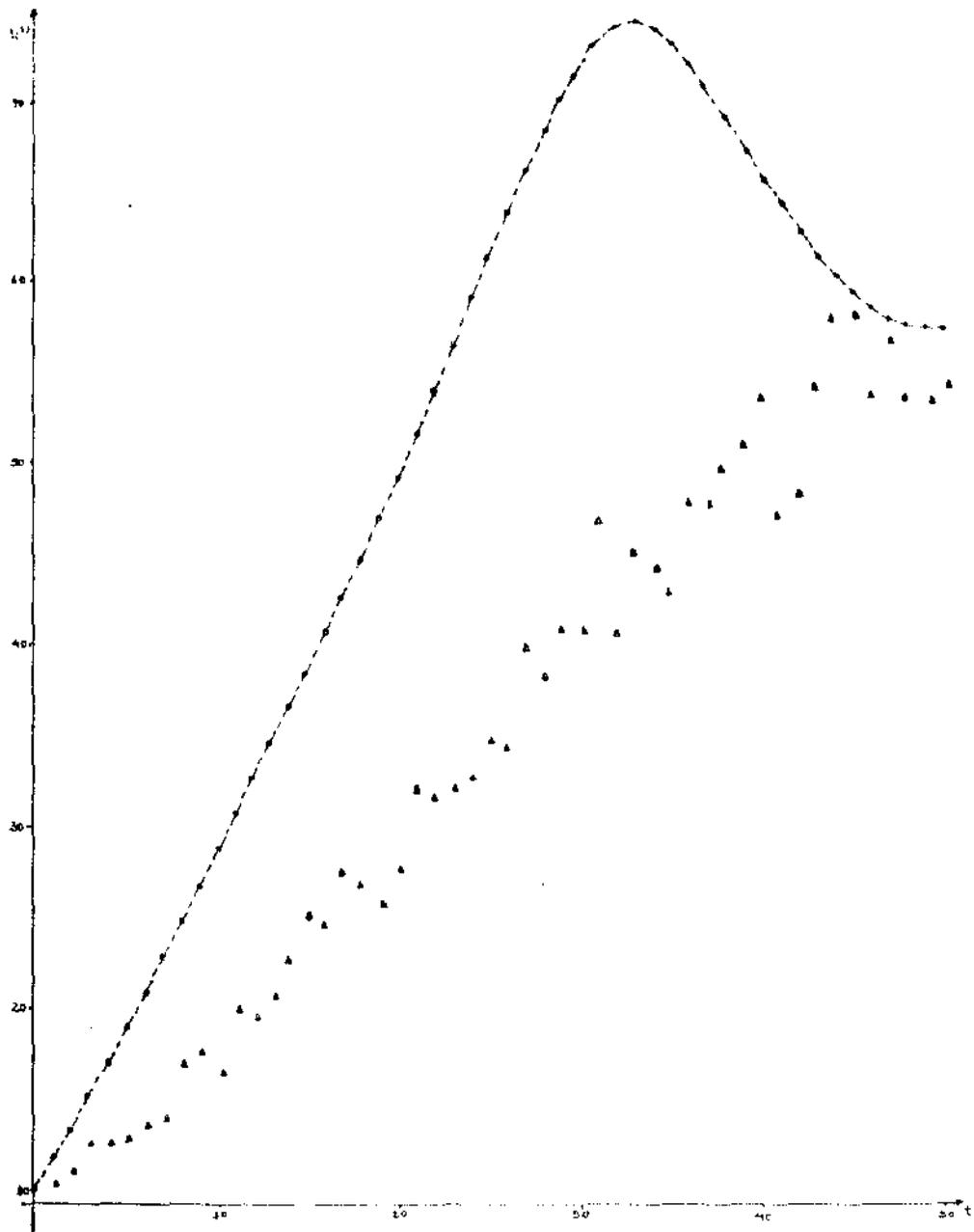


▲ Pontos dados.

● Pontos estimados.

Curva obtida pelo método dos Quadrados Mínimos para a espécie 1.

Figura 14

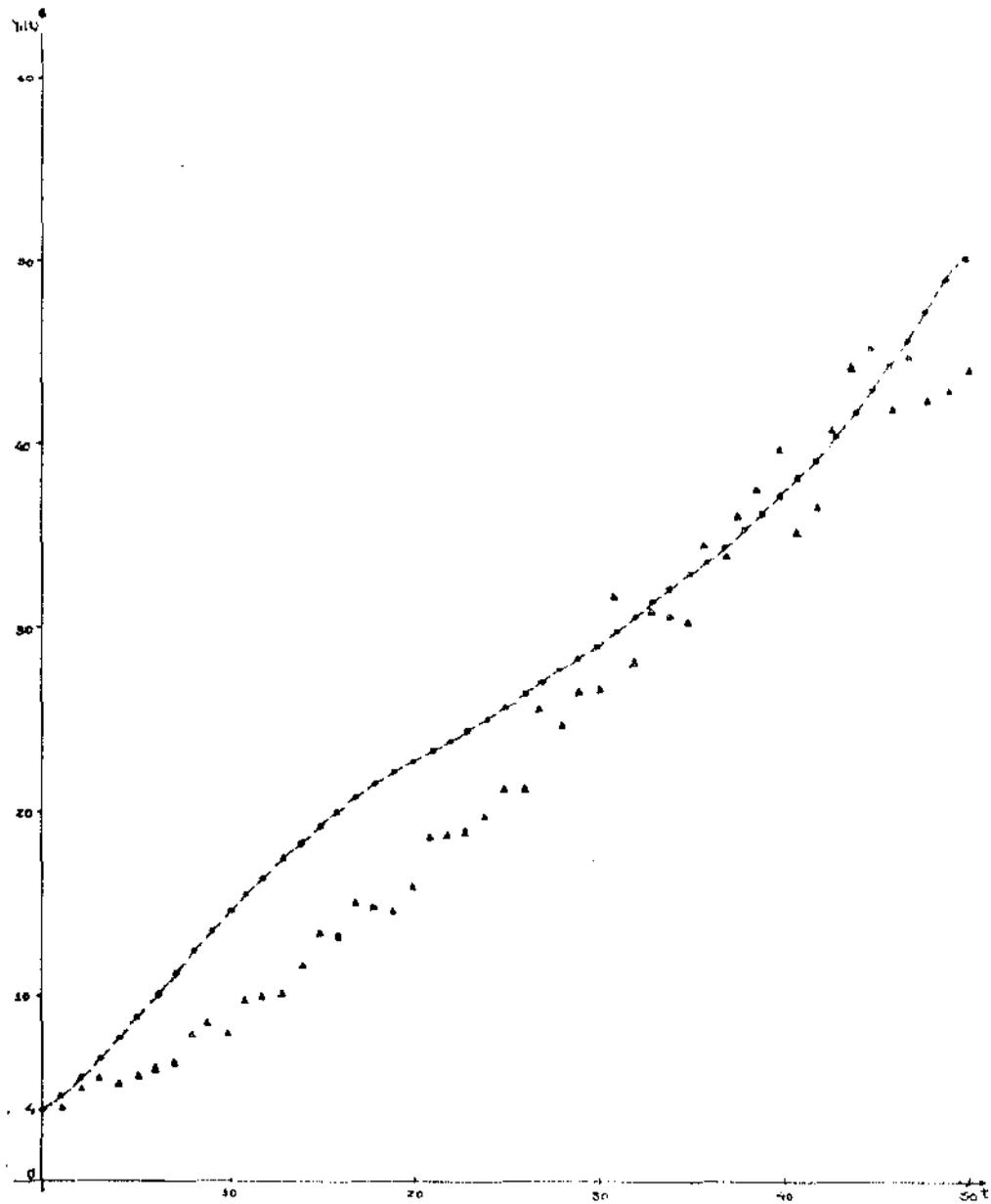


▲ Pontos dados.

● Pontos Estimados.

Curva obtida pelo método dos quadrados mínimos para a espécie 2.

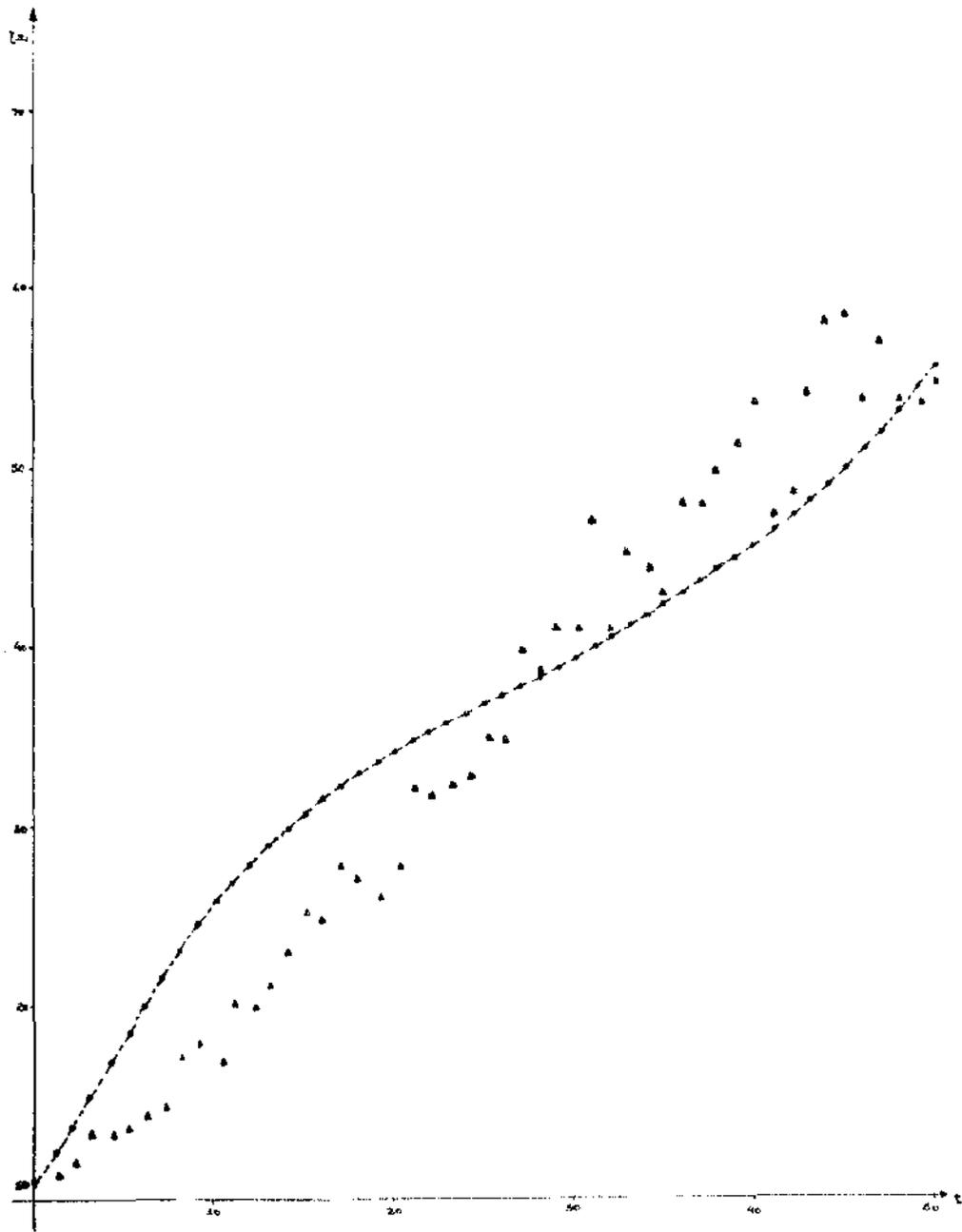
Figura 15



- ▲ Pontos dados.
- Pontos estimados.

Curva obtida pelo método de Bremermann para a espécie 1.

Figura 16

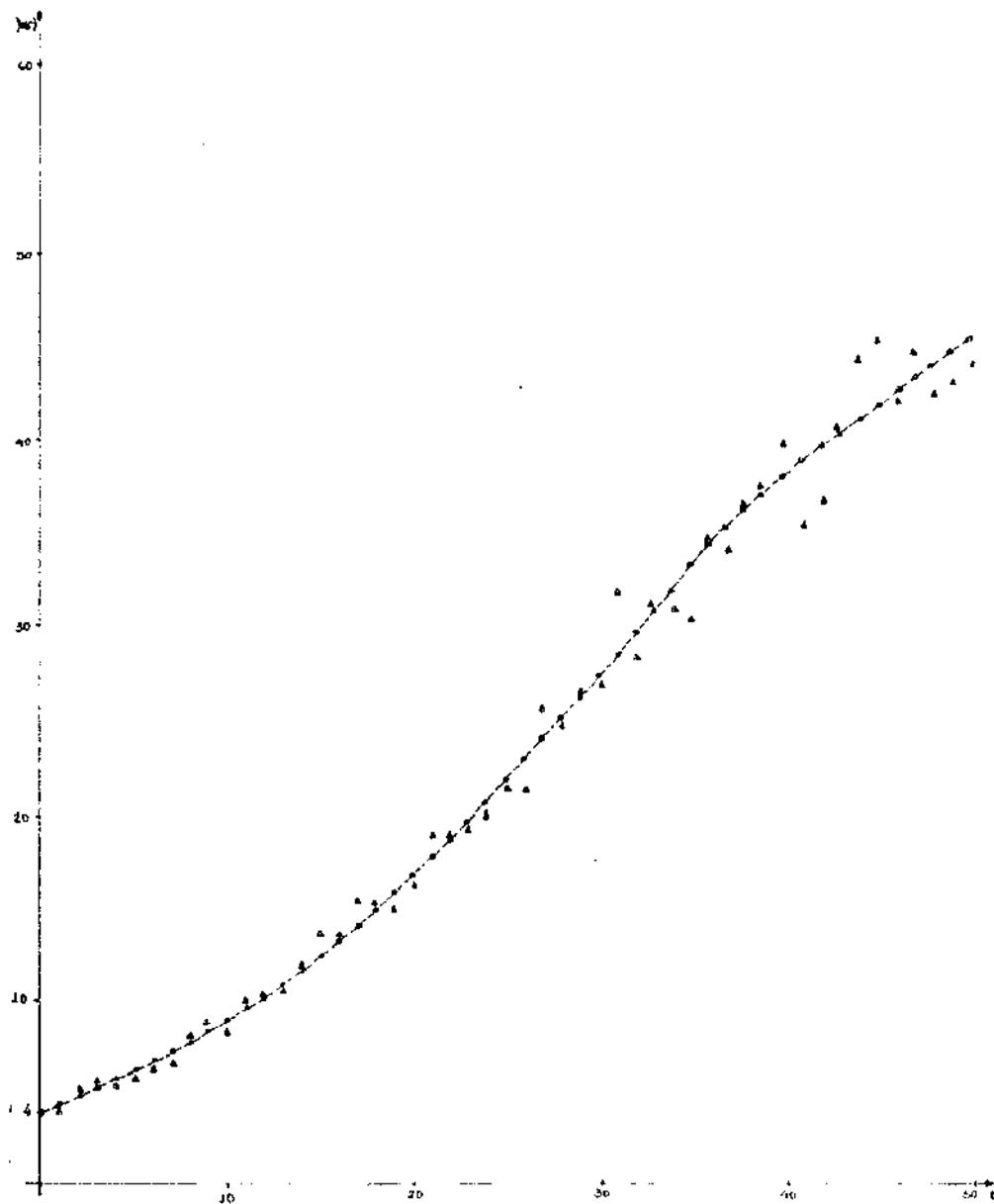


▲ Pontos dados.

● Pontos estimados.

Curva obtida pelo método de Bremermann para a es
pécie 2.

Figura 17

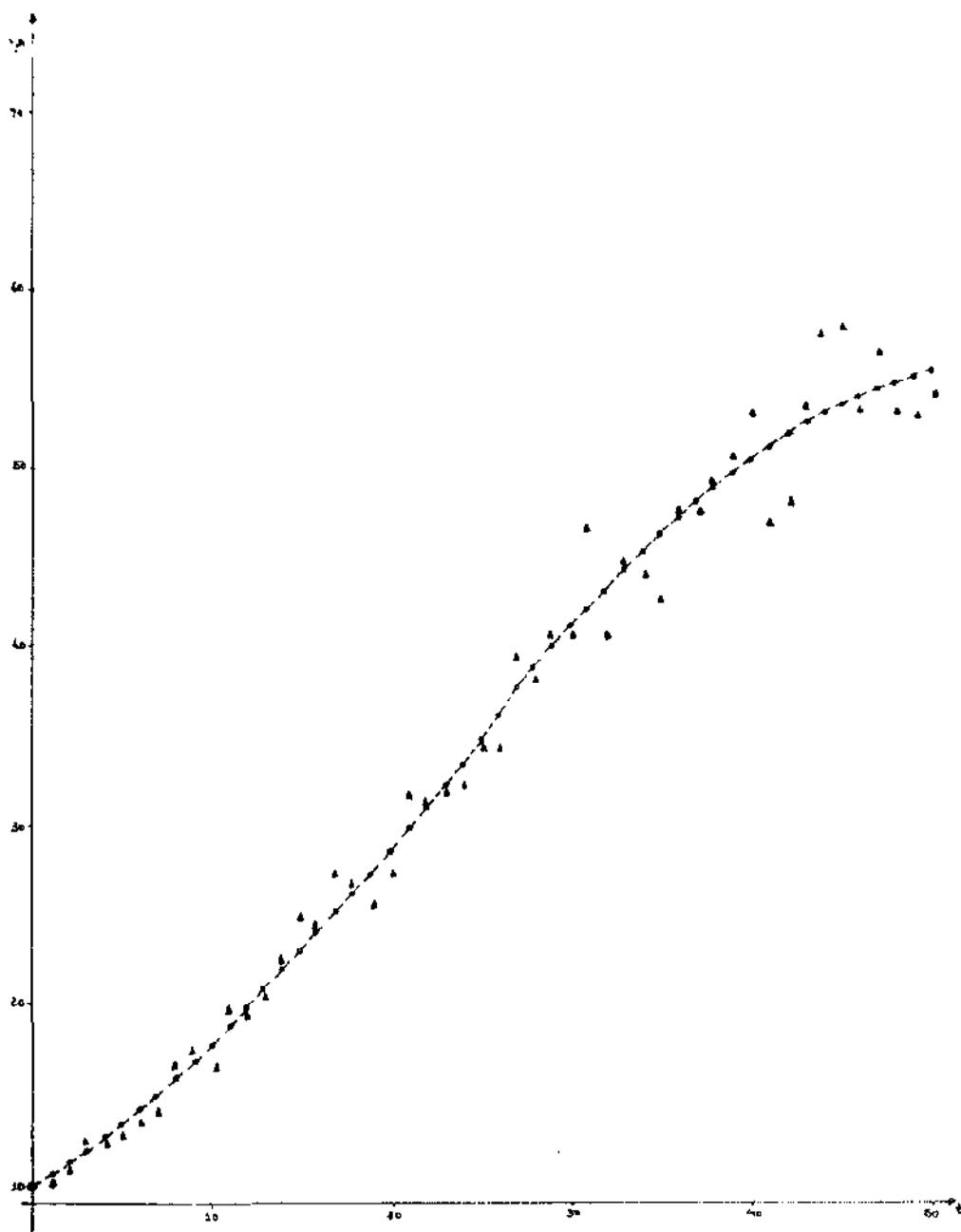


▲ Pontos dados.

● Pontos estimados.

Curva obtida pelo método de Levenberg-Marquardt
para a espécie 1.

Figura 18



▲ Pontos dados.

● Pontos estimados.

Curva obtida pelo método de Levenberg-Marquardt
para a espécie 2.

Figura 19

O segundo sistema teste considerado, descrito por Pielou [1], é o seguinte:

$$y_1(t+1) = y_1(t) + 0.1y_1(t) - 0.0014y_1(t)^2 - 0.0012y_1(t)y_2(t)$$

$$y_2(t+1) = y_2(t) + 0.08y_2(t) - 0.0009y_2(t)^2 - 0.001y_2(t)y_1(t)$$

$$y_1(0) = 10 \quad \text{e} \quad y_2(0) = 20$$

As Tabelas 4 e 5, mostram os resultados obtidos pelos três métodos com os dados perturbados de 5% e 10% respectivamente. Note nestas tabelas, que os resultados são similares aos resultados obtidos no problema anterior, onde o método de Levenberg-Marquardt mostrou-se mais eficiente.

PERTURBAÇÃO DE 5%			
PARÂMETROS	QUADRADOS MÍNIMOS LINEAR	MÉTODO DE LEVENBERG-MARQUARDT	MÉTODO DE BREMERMAN
x_1	0.1269380	0.9938138E-01	0.1269602
x_2	0.4304434E-02	0.1968373E-02	0.4348110E-02
x_3	0.3212018E-03	0.9042355E-03	0.3791798E-03
x_4	0.1072210	0.7876761E-01	0.1071353
x_5	0.1387272E-03	0.7473910E-03	0.1455355E-03
x_6	0.3681828E-02	0.1258780E-02	0.3785051E-02
SOMA DOS QUADRADOS	304.3636	76.15818	245.8196

Tabela 4

PERTURBAÇÃO DE 10%			
PARÂMETROS	QUADRADOS MÍNIMOS LINEAR	MÉTODO DE LEVENBERG- MARQUARDT	MÉTODO DE BREMERMAN
x_1	0.2052062	0.9903805E-01	0.2057770
x_2	0.1363615E-01	0.2634818E-02	0.1365748E-01
x_3	-0.2715541E-02	0.5637981E-03	-0.2401885E-02
x_4	0.1870636	0.7756431E-01	0.1874491
x_5	-0.2771380E-02	0.5990707E-03	-0.1898328E-02
x_6	0.1283036E-01	0.1510181E-02	0.1280465E-01
SOMA DOS QUADRADOS	2942.358	304.6056	1083.161

Tabela 5

O terceiro sistema teste considerado é um sistema Predador-Presa, cujo modelo é análogo ao modelo considerado. Este sistema é descrito por Bellman [21] e Swartz-Bremermann [22]. As equações deste sistema são:

$$y_1(t+1) = y_1(t) + x_1 y_1(t) - x_2 y_1(t)y_2(t)$$

$$y_2(t+1) = y_2(t) - x_3 y_2(t) + x_4 y_2(t)y_1(t)$$

$$y_1(0) = 1.2$$

$$y_2(0) = 1.1$$

$$x_j = 1.0 \quad (j = 1, \dots, 4)$$

Neste problema ocorreu uma dificuldade que vale a pena ser citada. Na resolução do modelo acima, ocorreu overflow devido o tamanho do passo, usado na discretização, o qual está implícito neste modelo ser igual a 1. Assim, seria necessário reescre

ver o modelo na forma

$$\begin{aligned} y_1(t+1) &= y_1(t) + h \cdot (x_2 y_1(t) - x_2 y_1(t) y_2(t)) \\ y_2(t+1) &= y_2(t) + h \cdot (-x_3 y_2(t) + x_4 y_2(t) y_1(t)) \end{aligned} ,$$

e usar $h = 0.1$. Entretanto esta dificuldade foi eliminada, observando que, multiplicar as taxas de crescimento abaixo, por h ,

$$(x_1 y_1(t) - x_2 y_1(t) y_2(t)) \quad \text{e} \quad (-x_3 y_2(t) + x_4 y_2(t) y_1(t))$$

corresponde a um escalamento dos parâmetros, onde h é o fator escala. Deste modo foram considerados os parâmetros escalados, ou seja

$$x'_j = h \cdot x_j \quad (j = 1, \dots, 4)$$

Visto que $h = 0.1$, os valores dos parâmetros x'_j são todos iguais a 0.1. Assim o modelo usado nos testes foi

$$\begin{aligned} y_1(t+1) &= y_1(t) + 0.1 y_1(t) - 0.1 y_1(t) y_2(t) \\ y_2(t+1) &= y_2(t) - 0.1 y_2(t) - 0.1 y_2(t) y_1(t) \\ y_1(0) &= 1.2 \\ y_2(0) &= 1.1 \end{aligned}$$

As Tabelas 6 e 7 mostram os resultados obtidos com os dados perturbados de 5% e 10% respectivamente.

Na Tabela 6, observamos o resíduo obtido pelo método de Levenberg-Marquardt é ligeiramente menor que o resíduo obtido pelo método de Bremermann. Situação oposta é verificada na Tabela 7. Isto mostra que a eficiência do método de Bremermann depende do problema considerado, contrastando com o método de Leven-

berg-Marquardt que obteve bons resultados em todos os problemas testes considerados.

PERTURBAÇÃO DE 5%			
PARÂMETROS	QUADRADOS MÍNIMOS LINEAR	MÉTODO DE LEVENBERG-MARQUARDT	MÉTODO DE BREMERMAN
x_1	0.1143882	0.1021205	0.1086856
x_2	0.1161240	0.1028078	0.1092730
x_3	0.6479416E-01	0.9622310E-01	0.9036741E-01
x_4	0.6213309E-01	0.9642146E-01	0.9077900E-01
SOMA DOS QUADRADOS	0.2821343	0.6291322E-01	0.6662717E-01

Tabela 6

PERTURBAÇÃO DE 10%			
PARÂMETROS	QUADRADOS MÍNIMOS LINEAR	MÉTODO DE LEVENBERG-MARQUARDT	MÉTODO DE BREMERMAN
x_1	0.1572176	0.1972900	0.1085306
x_2	0.1618839	0.2066164	0.1099377
x_3	-0.2415118E-01	0.1296503E-01	0.8864037E-01
x_4	-0.3523299E-01	0.7785139E-01	0.8908706E-01
SOMA DOS QUADRADOS	831.3657	0.8692188	0.2532408

Tabela 7

CONCLUSÃO

Os resultados das experiências numéricas confirmam que o modelo não pode ser considerado linear. Entretanto, as soluções obtidas, neste caso, constituem boa estimativa inicial para resolver o problema.

O método de Bremermann, apesar de reduzir de modo significativo os resíduos obtidos pelo método dos quadrados mínimos linear, não obteve na maioria dos testes, resultados melhores que os obtidos pelo método de Levenberg-Marquardt, o qual é sem dúvida mais eficiente. Isto nos leva a concluir que a melhor técnica para estimar os parâmetros do modelo proposto é aplicar o método de Levenberg-Marquardt com o cálculo recursivo das derivadas analíticas e com aproximação inicial obtida pelo método dos quadrados mínimos linear.

BIBLIOGRAFIA

- [1] PIELOU, E.C.: *Mathematical Ecology*, John Wiley & Sons (1977), (7 - 100).
- [2] BOYCE, W.E. and DIPRIMA, R.C.: *Elementary Differential Equations and Boundary Value Problems*, John Wiley & Sons, Inc. (1965) - (369 - 413).
- [3] LAWSON, C.L. and HANSON, R.J.: *Solving Least Squares Problems*, Prentice Hall, Inc., Englewood Cliffs, NJ, (1974).
- [4] HIMMELBLAN, D.M.: *Process Analysis by Statistical Methods*, John Wiley & Sons, Inc. (1970) - (176 - 324).
- [5] BARD, Y.: *Nonlinear Parameter Estimation*, Academic Press, Inc., (1974) - (83 - 139).
- [6] BARD, Y.: *Comparasion of Gradiante Methods for the Solution of Nonlinear Parameter Estimation*, SIAM J. Numer. Anal., 7, (1970), (157 - 186).
- [7] BECK, J.Y. and ARNOLD, K.J.: *Parameter Estimation in Engineering and Science*, John Wiley & Sons, (1977), (334 - 414).
- [8] LEE, E.S.: *Quasilinearization and Invariant Imbedding*, Academic Press (1968), (83 - 126).
- [9] DENIS, J.E.: *Nonlinear Least Squares and Equations - The State of the Art in Numerical Analysis*, Edited by D. Jacobs, Academic Press (1977), (269 - 306).
- [10] BETTS, J.T.: *Solving the Nonlinear Least Squares Problems: Application of the General Method*. Journal of Optimization Theory and Applications, 18 (1976), 469 - 483.
- [11] BROWN, K.M. and DENNIS, J.E.: *Derivative-free Analogues of the Levenberg-Marquardt and Gauss Algorithms for nonlinear*

- Least Squares Approximation*. Numer. Math., 18 (1972), (289 - 297).
- [12] DAVIDON, W.C.: *New Least Squares Algorithms*. Journal of Optimization Theory and Application, 18 (1976), (187 - 197).
- [13] JONES, A.P.: *Spiral - A New Algorithm for Nonlinear Parameter Estimation using Least Squares*. Computer J. 13 (1970), (301 - 308).
- [14] LEVENBERG, K.: *A Method for Solution of Certain Nonlinear Problems in Least Squares*. Quart. Appl. Math., 2, (1944), pp. (164 - 168).
- [15] MARQUARDT, D.W.: *An Algorithm for Least Squares Estimation of Nonlinear Parameter*. J. SIAM, 11 (1963), (437 - 441).
- [16] NELDER, J.A. and MEAD, R.: *A Simplex Method for Function Minimization*. Computer J., 7 (1965), (308 - 313).
- [17] PECKHAM, G.: *A New Method for Minimizing a Sum of Squares without Calculating Gradients*. Computer J. 13 (1970), (418 - 420).
- [18] POWELL, M.J.D.: *A Method for Minimizing a Sum of Squares of Nonlinear Functions without Calculating Derivatives*. Computer J., 7 (1965), (303 - 307).
- [19] RALSTON, M.L. and JENRICH, R.I.: *Dud, A Derivative-Free Algorithm for Nonlinear Least Squares*. Technometrics, 20, (1978), (7 - 13).
- [20] BREMERMAN, H.: *A Method of Unconstrained Global Optimization*. Math. Biasi, 9 (1970).
- [21] BELLMAN, R. and al.: *Inverse Problems in Ecology*. J. Theoret Biol. 11, 164 (1966).
- [22] SWARTZ, J. and BREMERMAN, H.: *Dimension of Parameter Es-*

- timation in Biological Modelling: Algorithms for Estimation and Evaluation of Estimates.* Journal of Math. Biology, 1, (1975), (241 - 257).
- [23] HARTLEY, H.O.: *The Modified Gauss-Newton Method for Fitting of Nonlinear Regression Functions by Least Squares.* Technometrics, 3 (1961), (269 - 280).
- [24] BALL, W.E.: Ind. Eng. Chem. Fundamentals 6, 475 (1967).
- [25] MORE, J.J.: *The Levenberg-Marquardt Algorithm - Implementation and Theory.* Lecture e Notes in Mathematics, Numerical Analysis, Dundee 1977 .
- [26] FLETCHER, R.: *A Modified Marquardt Subroutine for Nonlinear Least Squares.* Atomic Energy Establishment report R6799, Harwell, England (1971).
- [27] NAG FORTRAN Library Manual, Mark 6, *The Numerical Algorithms Group* (USA) Inc., Downers Grove Ill, 1978.
- [28] IMSL Library Reference Manual, 8th ed. IMSL, Houston, Tex., 1980.