

## Método de Levenberg-Marquardt

Problemas de mínimos quadrados têm como objetivo encontrar o melhor ajuste para um conjunto de dados de tal modo que a soma dos quadrados das distâncias (tomadas na vertical) entre o modelo (curva ajustada) e cada um dos pontos dados seja a menor possível. Essas diferenças entre a curva ajustada e cada um dos dados são denominadas resíduos.

Em linguagem matemática podemos escrever o problema de mínimos quadrados da seguinte forma:

Considere  $x = [x_1, x_2, x_3, \dots, x_n]$ .

Dada uma função vetorial  $f: \mathbf{R}^n \rightarrow \mathbf{R}^m$  com  $m \geq n$ , queremos minimizar  $\|f(x)\|$ , ou equivalente, para encontrar  $x^* = \text{mínimo local para } F(x)$ ,

onde

$$F(x) = \frac{1}{2} \sum_{i=1}^m (f_i(x))^2 = \frac{1}{2} \|f(x)\|^2 = \frac{1}{2} f(x)^T f(x).$$

Um bom motivo para se minimizar a soma dos quadrados dos resíduos ao invés de se minimizar a soma dos resíduos é o fato de que a função módulo não é diderenciável na origem e serão usadas derivadas para resolver os problemas de mínimos quadrados não-lineares.

Como exemplo dessa classe de problemas considere dados a coleção de pontos  $(t_1, y_1), \dots, (t_{m-1}, y_{m-1}), (t_m, y_m)$  obtidos em função de algum experimento e o modelo  $M(x, t) = x_3 e^{x_1 t} + x_4 e^{x_2 t}$  a ser ajustado. Esse exemplo foi retirado da referência [2].

Assume-se que existe um  $x^+$  tal que  $y_i = M(x^+, t_i) + \varepsilon_i$ , onde  $\{\varepsilon_i\}$  são erros de medida. Sendo assim, para qualquer escolha de  $x$  pode-se encontrar os valores dos resíduos da seguinte forma:

$$f_i(x) = y_i - M(x, t_i) = y_i - x_3 e^{x_1 t_i} - x_4 e^{x_2 t_i}, \text{ com } i = 1, \dots, m.$$

Pode-se perceber que o modelo depende dos parâmetros  $x = [x_1, x_2, x_3, x_4]^T$ . O objetivo do problema seria, então, encontrar  $x^* = [x_1^*, x_2^*, x_3^*, x_4^*]^T$  onde  $x_1^*, x_2^*, x_3^*, x_4^*$  são os valores que minimizam a soma dos quadrados dos resíduos, o que significa encontrar a melhor curva que se ajusta aos pontos fornecidos. A figura abaixo mostra exatamente os pontos fornecidos e a curva obtida no ponto  $x^*$ , isto é,  $M(x^*, t)$ .

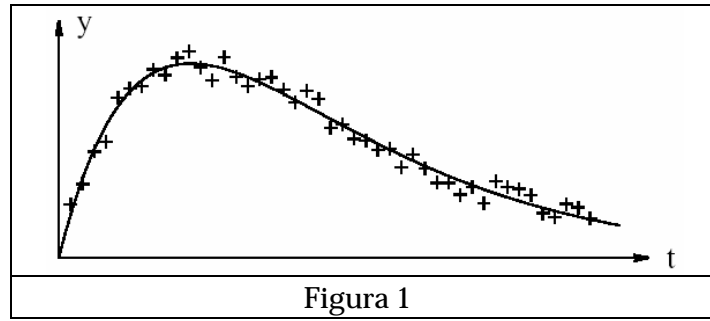


Figura 1

Existem diversas maneiras de se resolver problemas de mínimos quadrados não-lineares. Nesse trabalho foi implementado o método de Levenberg-Marquardt, que consiste em um aperfeiçoamento do método de Gauss-Newton que, por sua vez, é uma variante do método de Newton.

Assim como os métodos de Newton e Gauss-Newton, o método de Levenberg-Marquardt é iterativo. Isto significa que, dado um ponto inicial  $x_0$ , o método produz uma série de vetores  $x_1, x_2, \dots$ , que espera-se que vá convergir para  $x^*$ , um mínimo local para a função de entrada a ser ajustada.

O método de Newton se baseia no fato de que se  $x^*$  é um ponto crítico da função  $F$ , então  $F'(x^*) = 0$ . Expandindo-se  $F$  em uma série de Taylor obtém-se:

$$\begin{aligned} F(x+h) &= F(x) + F'(x)h + O(\|h\|^2) \\ &\approx F(x) + F'(x)h, \text{ para valores pequenos de } \|h\|. \end{aligned}$$

Desta forma, pode-se escrever:

$$F'(x+h) = F'(x) + F''(x)h.$$

Sabendo-se que  $F'(x^*) = 0$  e  $F'(x+h) = F'(x) + F''(x)h$ , pode-se encontrar o valor de  $h_n$ , isto é, a direção de busca no método de Newton, resolvendo-se o seguinte sistema de equações:

$$Hh_n = -F'(x)$$

$$\text{onde } H = F''(x) = H(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix} \text{ é conhecida como}$$

a matriz Hessiana.

Calculado a direção do passo  $h_n$  a próxima iteração do método de Newton é dada por:

$$x \leftarrow x + h_n.$$

No entanto, a avaliação da matriz Hessiana pode ser extremamente complexa para alguns problemas. Para contornar esse problema, foram propostos métodos que utilizam aproximações para essa matriz, denotados Quasi-Newton, dentre os quais se encontram o método de Gauss-Newton e Levenberg-Marquardt.

O método de Gauss-Newton se baseia na aproximação para os valores das componentes de  $f$  em uma vizinhança de  $x$ . Da expansão em série de Taylor pode-se perceber que:

$$f(x+h) \approx l(h) \equiv f(x) + J(x)h,$$

onde  $J \in \mathbf{R}^{m \times n}$  é a matriz Jacobiana, isto é, a matriz que contém as derivadas parciais de primeira ordem de cada componente da função vetorial  $f$ .

$$(J(x)) = \begin{pmatrix} \frac{\delta f_1}{\delta x_1} & \frac{\delta f_1}{\delta x_2} & \dots & \frac{\delta f_1}{\delta x_n} \\ \frac{\delta f_2}{\delta x_1} & \frac{\delta f_2}{\delta x_2} & \dots & \frac{\delta f_2}{\delta x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\delta f_m}{\delta x_1} & \frac{\delta f_m}{\delta x_2} & \dots & \frac{\delta f_m}{\delta x_n} \end{pmatrix}_{m \times n}$$

$$\text{Substituindo-se } f(x+h) \approx l(h) \equiv f(x) + J(x)h \quad \text{em} \quad F(x) = \frac{1}{2} f(x)^T f(x)$$

obtemos:

$$\begin{aligned} F(x+h) &\approx L(h) \equiv \frac{1}{2} l(h)^T l(h) \\ &= \frac{1}{2} f^T f + h^T J^T f + \frac{1}{2} h^T J^T J h \\ &= F(x) + h^T J^T f + \frac{1}{2} h^T J^T J h \end{aligned}$$

onde  $f = f(x)$  e  $J = J(x)$ .

É fácil perceber que o gradiente e a matriz Hessiana de  $L$  são dados respectivamente por:

$$L'(h) = J^T f + J^T J h \quad \text{e} \quad L''(h) = J^T J.$$

Portanto, a direção do passo calculada por meio do método de Gauss-Newton é dada pela seguinte equação:

$$(J^T J)h_{gn} = -J^T f .$$

Encontrado o valor de  $h_{gn}$  tem-se que a próxima iteração do método de Gauss-Newton é dada por:

$$x \leftarrow x + h_{gn} .$$

Apesar do método de Gauss-Newton resolver de maneira mais fácil a matriz Hessiana, pode-se perceber que, da mesma forma que acontecia no método de Newton, não é garantido que exista a inversa dessa matriz, necessária para o cálculo de  $h_{gn}$ .

O método de Levenberg-Marquardt, para contornar essa situação, propõe somar uma parcela  $\mu I$  à matriz Hessiana, aproximada pelo método de Gauss-Newton, onde  $\mu$  é um escalar denominado parâmetro de damping e  $I$  é a matriz identidade.

Com essa modificação, a direção do passo do método de Levenberg-Marquardt pode ser calculada da seguinte maneira:

$$(J^T J + \mu I)h_{lm} = -g , \text{ onde } g = J^T f \text{ e } \mu \geq 0 .$$

O parâmetro de damping  $\mu$  promove diferentes comportamentos do método, ou seja:

1. Para todo  $\mu > 0$  a matriz de coeficientes  $(J^T J + \mu I)$  é positiva definida (veja a demonstração no apêndice A), o que implica que  $h_{lm}$  é uma direção de descida.
2. Para valores grandes de  $\mu$  tem-se:

$h_{lm} = -\frac{1}{\mu} g = -\frac{1}{\mu} F'(x)$ , que é um pequeno passo na direção máxima de descida (ver apêndice B).

3. Se  $\mu$  é muito pequeno temos que  $h_{lm} \approx h_{gn}$ , o que é bom nos estágios finais da iteração quando  $x$  está próximo de  $x^*$ , pois, quando isso ocorre, o método de Levenberg-Marquardt consegue convergência quadrática.

Podemos perceber que o parâmetro  $\mu$  influencia tanto na direção quanto no tamanho do passo, permitindo assim que o método de Levenberg-Marquardt não necessite de uma busca linear para descobrir o tamanho ótimo do passo a ser dado em cada iteração.

O valor inicial de  $\mu$  pode ser relacionado ao tamanho dos elementos da matriz Hessiana aproximada que é calculada no método de Levenberg-Marquardt, isto é, ao tamanho dos elementos de  $A_0 = J(x_0)^T J(x_0)$ .

$$\mu_0 = \tau \max_i \{a_{ii}^{(0)}\},$$

onde  $\tau$  é escolhido pelo usuário. Para determinados problemas em que o valor de  $x_0$  é uma boa aproximação para  $x^*$  pode-se adotar  $\tau = 10^{-6}$ . Caso contrário, recomenda-se  $\tau = 10^{-3}$  ou até mesmo  $\tau = 1$ .

O valor de  $\mu$  pode ser modificado durante as iterações e isso é feito em função de um outro parâmetro  $\rho$ , denominado de gain ratio, ou seja:

$$\rho = \frac{F(x) - F(x + h_{lm})}{L(0) - L(h_{lm})},$$

onde  $L(0) - L(h_{lm})$  é o ganho obtido através do modelo linear.

$$\begin{aligned} L(0) - L(h_{lm}) &= -h_{lm}^T J^T f - \frac{1}{2} h_{lm}^T J^T J h_{lm} \\ &= -\frac{1}{2} h_{lm}^T (2g + (J^T J + \mu I - \mu I) h_{lm}) \\ &= \frac{1}{2} h_{lm}^T (\mu h_{lm} - g). \end{aligned}$$

Pode-se garantir que como  $h_{lm}^T h_{lm}$  e  $-h_{lm}^T g$  são positivos, então  $L(0) - L(h_{lm})$  também é positivo.

Se o valor de  $\rho$  for grande, isso indica que  $L(h_{lm})$  é uma boa aproximação para  $F(x + h_{lm})$  e pode-se diminuir o valor de  $\mu$  de modo que o próximo passo do Levenberg-Marquardt esteja próximo do passo de Gauss-Newton. Se  $\rho$  for pequeno isso significa que  $L(h_{lm})$  é uma aproximação ruim e precisamos aumentar  $\mu$  para buscar a direção máxima de descida e reduzir o tamanho do passo.

O valor de  $\mu$  é alterado de acordo com a seguinte regra:

se  $\rho > 0$ ,

$$\nu = 2.$$

caso contrário,

$$\mu = \mu\nu \text{ e } \nu = 2\nu.$$

onde  $\nu$  é inicializado com valor igual a 2.

Os critérios de convergência ou critérios de parada mais usuais são:

1. Norma do gradiente menor do que um determinado valor pré-determinado, isto é,  $\|g\|_{\infty} \leq \varepsilon_1$ , onde  $\varepsilon_1$  é um número positivo pequeno escolhido pelo usuário de modo a refletir que  $F'(x^*) = g(x^*) = 0$ .
2. Variação no valor de x muito pequena, isto é,  $\|x_{new} - x\| \leq \varepsilon_2(\|x\| + \varepsilon_2)$ .
3. Número máximo de iterações determinado pelo usuário. Isto é importante para evitar que o algoritmo entre em loop infinito:  $k \geq k_{max}$ , onde k é o número de iterações.

O pseudo-código do algoritmo de levenberg-marquardt pode ser visto na figura abaixo.

Método de Levenberg-Marquardt
<pre> k = 0;  ν = 2;  x = x<sub>0</sub>; A = J(x)<sup>T</sup> J(x);  g = J(x)<sup>T</sup> f(x); found = (  g  <sub>∞</sub>) ≤ ε<sub>1</sub>;  μ = τ · max{a<sub>ii</sub>}; while (not found) and (k &lt; k<sub>max</sub>) {     k=k+1; Solve (A + μI)h<sub>lm</sub> = -g ;     if   h<sub>lm</sub>   ≤ ε<sub>2</sub>(  x   + ε<sub>2</sub>)         found = true     else     {         x<sub>new</sub> = x + h<sub>lm</sub>         ρ = (F(x) - F(x<sub>new</sub>))/(L(0) - L(h<sub>lm</sub>))         if ρ &gt; 0         {             X = X<sub>new</sub>             A = J(x)<sup>T</sup> J(x);  g = J(x)<sup>T</sup> f(x);             found = (  g  <sub>∞</sub>) ≤ ε<sub>1</sub>;             μ = μ max{1/3, 1 - (2ρ - 1)<sup>3</sup>};  ν = 2;         }         else             μ = μν;  ν = 2ν;     } } </pre>

Os cálculos da matriz Jacobiana, necessária para se encontrar uma aproximação da matriz Hessiana, e do gradiente foram feitos pelo método de diferenças finitas, que pode ser descrito da seguinte forma:

Considere  $x = [x_1, \dots, x_n]^T$  e faça para  $i = 1, \dots, m$  e  $j = 1, \dots, n$ :

$$\frac{\delta f_i}{\delta x_j} \approx \frac{f_i(x + h e_j) - f_i(x)}{h},$$

onde  $h$  é o tamanho do passo e,  $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}_{n \times 1}$ ,  $e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}_{n \times 1}$ , sucessivamente.

A norma infinita representada por  $\|M\|_\infty$  e utilizada no pseudo-código acima, é dada por:

$$\|M\|_\infty = \max_i \sum_{j=1}^n |m_{ij}|, \text{ isto é, escolhe-se a linha que possui a maior soma dos}$$

valores absolutos de seus elementos. O resultado disso é, portanto, um escalar.

Como pode ser percebido no pseudo-código a estrutura do método segue o mesmo esquema do método de descida (descent method) onde, em cada iteração, encontra-se uma direção de descida e dá-se um passo nessa direção em que o valor de  $F$  é reduzido (veja o pseudo-código do método de descida no apêndice C). Tanto a direção quanto o tamanho do passo no método de Levenberg-Marquardt são influenciados pelo parâmetro de damping  $\mu$ , como já havia sido comentado anteriormente, dispensando assim o uso da busca linear para o cálculo do tamanho do passo, como ocorre no método de descida.

O método de Levenberg-Marquardt satisfaz a condição de descida em cada passo da iteração, a qual é dada por  $F(x_{k+1}) < F(x_k)$ .

Os critérios de parada e a alteração do valor do parâmetro de damping do método implementado já foram discutidos no item anterior.

O protótipo da função que implementa o método de levenberg-marquardt é dado por:

```
LMQ_MinimizarLevMarq( void (*pf)(int,int,double*,double*), void (*pfj)(Matriz*,
double*, int, int), int numMaxIteracoes, double * x_inicial, int tam_x, int
num_ptsAmostrados, double tam_passo, double e1, double e2, double t);
```

onde

pf – ponteiro para uma função f.

pfj – ponteiro para a função que fornece a jacobiana. Se pfj = NULL, então a jacobiana é calculada usando diferenças finitas.

numMaxIteracoes – limite máximo de iterações utilizadas.

x\_inicial – chute inicial para x. Em x\_inicial também é retornado o valor final de x encontrado pelo algoritmo de Levenberg-Marquardt.

tam\_x – dimensão de x.

num\_ptsAmostrados - número de pontos amostrados.

tam\_passo – tamanho do passo. Esse parâmetro é necessário para o cálculo das derivadas parciais.

e1 – valor de tolerância para  $F'(x^*)$ .

e2 – Penalização para quando a variação no valor de x é muito pequena.

t – usa-se  $t = 10^{-6}$  se  $x_0$  é uma boa aproximação para  $x^*$ , caso contrário usa-se  $t = 10^{-3}$  ou  $t = 1$ .

## Apêndices

### A. Matriz positiva definida

Uma matriz  $A$  é positiva definida quando se  $x^T A x > 0$  para todo  $x \in \mathbf{R}^n, x \neq 0$ , e positiva semidefinida se  $x^T A x \geq 0$  para todo  $x \in \mathbf{R}^n, x \neq 0$ .

Sendo  $J \in \mathbf{R}^{m \times n}$  faça  $A = J^T J$ . Temos que  $A^T = J^T (J^T)^T = A$ , logo a matriz  $A$  é simétrica. Além disso, para qualquer  $x \neq 0 \in \mathbf{R}^n$  tome  $y = Jx$ . Desta forma,  $x^T A x = x^T J^T J x = y^T y \geq 0$ . Com isso provamos que  $A$  é positiva semidefinida e conseqüentemente todos os seus autovalores são maiores ou iguais a zero. Dessa forma podemos escrever  $(A + \mu I)v_j = (\lambda_j + \mu)v_j, j = 1, \dots, n$  e  $\lambda_j$  são os autovalores de  $A$ . Isso implica que os autovalores de  $(A + \mu I)$  são positivos e essa matriz é garantidamente positiva definida para todo  $\mu > 0$ . Além disso, como  $A$  é simétrica temos que  $(A + \mu I)$  também é simétrica.

### B. Direção máxima de descida



Expandindo F em uma série de Taylor obtemos:

$$F(x + \alpha h) = F(x) + \alpha h^T F'(x) + O(\|h\|^2) \\ \approx F(x) + \alpha h^T F'(x) \quad , \quad \text{para valores pequenos e positivos de } \alpha \text{ (tamanho do passo).}$$

Dizemos que h é uma direção de descida para F se  $h^T F'(x) < 0$ . O ganho relativo no valor da função satisfaz:

$$\lim_{\alpha \rightarrow 0} \frac{F(x) - F(x + \alpha h)}{\alpha \|h\|} = -\frac{1}{\|h\|} h^T F'(x) = -\|F'(x)\| \cos \theta .$$

Podemos perceber que o ganho máximo acontece quando  $\theta = \pi$ , o que implica em  $h_{sd} = -F'(x)$ , onde  $h_{sd}$  é a direção máxima de descida.

## C. Métodos de descida

```

k = 0;  x = x0;
while (not found) and (K < Kmax)
{
    hd = search_direction(x)
    if (hd não existe )
        found = true
    else
    {
        α = step_length(x, hd)
        x = x + αhd;  k = k + 1;
    }
}

```

onde  $h_d$  é a direção de descida e  $\alpha$  é o tamanho do passo. O tamanho do passo no método da descida é encontrado fazendo-se uma busca linear e a direção de descida utilizando método de Newton ou da máxima descida.

## Referências Bibliográficas

[1] ZHANG, Zhengyou. A Flexible New Technique for Camera Calibration [online]. MicrosoftResearch, December 2, 1998. Last updated on Aug. 10, 2002. Available from World Wide Web: <<http://research.microsoft.com/~zhang/calib/>>.

[2] K. Madsen, H.B. Nielsen, O. Tingleff. Methods for non-linear least squares problems [online]. Technical University of Denmark, April 2004, 2<sup>nd</sup> Edition. Available from World Wide Web: <<http://www2.imm.dtu.dk>>.

[3] Mokhtar S. Bazaraa, Hanif D. Sherali, C. M. Shetty. Nonlinear Programming Theory and Algorithms. Second Edition. John Wiley & Sons, Inc. ISBN 0-471-55793-5.

[4] Sérgio Luciano Ávila, Charles Borges de Lima e Walter Pereira Carpes Junior. Otimização Conceitos Básicos, Ferramentas e Aplicações. GRUCAD Departamento de Engenharia Elétrica – UFSC. Available from World Wide Web: <<http://www.grucad.ufsc.br/avila/>>.

[5] Luiz M. G. Golçalves. Visão Computacional Calibração de Câmeras. UFRN. Available from World Wide Web: <<http://www.dca.ufrn.br/~lmarcos/courses/visao/notes/index.html>>.

[6] Iaci Malta, Sinésio Pesco, Hélio Lopes. Cálculo a uma variável: Derivada e Integral volume 2. Primeira Edição. Editora PUC-rio. ISBN 85-15-02442-4.

[7] Humberto José Bortolossi. Cálculo diferencial a várias variáveis: uma introdução à teoria de otimização. Primeira Edição. Editora PUC-rio. ISBN 85-15-02442-X.

[8] William E. Boyce, Richard C. DiPrima. Equações Diferenciais Elementares e Problemas de Valores de Contorno. Sétima Edição. LTC – Livros Técnicos e Científicos Editora S.A.

[9] Hugo Iver V. Gonçalves, Rodrigo C.M. Coimbra, Marcelo Ladeira e Hércules Antônio do Prado. Ferramenta de Regressão Não-linear Múltipla com Redes Neurais Artificiais [online]. Available from World Wide Web: < <http://www.sbc.org.br/>>.

[10] Marcos Craizer, Geovan Tavares. Cálculo Integral a Várias Variáveis. Primeira Edição. Editora PUC-rio. ISBN 85-15-02441-1.

[11] Charles C. Dyer and Peter S. S. Ip. An Elementary Introduction to Scientific Computing [online]. Division of Physical Sciences , University of Toronto at Scarborough. Revised January 3, 2000. Available from World Wide Web: <[http://pathfinder.scar.utoronto.ca/~dyer/csca57/book\\_P/book.html](http://pathfinder.scar.utoronto.ca/~dyer/csca57/book_P/book.html)>.