

# O Método de Levenberg-Marquardt

Lucas da Penha Soares

Henrique Tapparello Moresco

14 de outubro de 2022

## 1 Introdução

Com recorrência usa-se sistemas de equações lineares e não lineares para modelar problemas do dia a dia. Da física à biologia, a matemática aplicada vem sido usada para ajudar a revolucionar o mundo. Deste modo, faz sentido tentar modelar sistemas onde dada uma matriz  $A \in \mathbb{R}^{m \times n}$  e um vetor  $b \in \mathbb{R}^m$ , buscamos resolver o sistema  $Ax = b$ , com  $x \in \mathbb{R}^n$ .

Estaremos interessados em sistemas não lineares, sendo assim vamos considerar a função  $R : \mathbb{R}^n \mapsto \mathbb{R}$  onde o objetivo é a busca por  $x$  em que  $R(x) = 0$ .

O algoritmo de Levenberg-Marquardt foi desenvolvido no início da década de 1960 para resolver problemas de mínimos quadrados não lineares. Problemas de mínimos quadrados surgiram para tentar ajustar um modelo matemático parametrizado a um conjunto de pontos de dados. Se a função não é linear em seus parâmetros, o problema dos mínimos quadrados requer um algoritmo de solução iterativa. Tais algoritmos reduzem a soma dos quadrados dos erros entre a função do modelo e os pontos de dados através de uma sequência de atualizações nos valores dos parâmetros do modelo.

O algoritmo de Levenberg-Marquardt combina dois algoritmos, o Método do Gradiente e o Método de Gauss-Newton. No Método do Gradiente, a soma dos quadrados dos erros é reduzida atualizando os parâmetros na direção da descida mais íngreme. No Método de Gauss-Newton, a soma dos quadrados dos erros é reduzida assumindo o menor valor da função localmente quadrática nos parâmetros, e assim, se encontra o mínimo desta quadrática. O método Levenberg-Marquardt age mais como o Método do Gradiente quando os parâmetros estão longe de seu valor ideal, e age mais como o Método de Gauss-Newton quando os parâmetros estão próximos de seu valor ótimo.

Antes de apresentarmos o método principal, vamos introduzir o método de Gauss-Newton como uma motivação ao leitor.

## 2 Gauss-Newton

Considere a seguinte aproximação de primeira ordem  $R(x)$  ao redor do ponto  $x_k$ ,

$$R(x) \approx R(x_k) + J(x_k)d,$$

onde  $d = x - x_k$ . Tal aproximação implica que  $\|R(x)\| \approx \|R(x_k) + J(x_k)d\|$ .

Deste modo, consideremos o seguinte problema

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} \|R(x_k) + J(x_k)d\|^2. \quad (1)$$

Reescrevendo tal função objetivo obtemos,

$$\frac{1}{2} \left( R(x_k)^T R(x_k) + 2d^T J(x_k) R(x_k) + d^T J(x_k)^T J(x_k) d \right) \quad (2)$$

A direção de minimização deve satisfazer o seguinte sistema

$$(J^T(x_k) J(x_k)) d_k = -J^T(x_k) R(x_k). \quad (3)$$

Tal direção é denominada **direção de Gauss-Newton**. O método está bem definido desde que  $J(x_k)^T J(x_k)$  seja inversível, isto é, definida positiva.

Com o intuito de apresentar uma motivação para o método de Levenberg-Marquardt, apresentaremos um exemplo em que o método apresentado acima pode falhar.

**Exemplo 2.1.** Considere a função  $R : \mathbb{R}^2 \mapsto \mathbb{R}^2$  definida por

$$R(x) = \begin{pmatrix} x_1^2 x_2 \\ x_2 + 7 \end{pmatrix}$$

e a aproximação inicial  $x_0 = (0, 1)$ . Assim, o Jacobiano de  $R$  é dado por

$$J(x) = \begin{pmatrix} 2x_1 x_2 & x_1^2 \\ 0 & 1 \end{pmatrix}$$

Assim, através do método de Gauss-Newton a partir do ponto  $x_0$ , a direção de minimização satisfaz o sistema

$$J^T(0, 1) J(0, 1) d = -J^T(0, 1) R(0, 1)$$

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} d = - \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 8 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} d = \begin{pmatrix} 0 \\ -8 \end{pmatrix}$$

Não é difícil ver que este sistema não está bem definido já que qualquer direção do tipo  $d = (d' - 8)^T$  é uma solução do sistema. Veja que isto ocorre pois  $J^T(0,1)J(0,1)$  não é uma matriz inversível, ou seja, definida positiva. Entretanto, se considerarmos o sistema

$$(J^T(0,1)J(0,1) + \lambda I)d = -J^T(0,1)R(0,1),$$

com  $\lambda = 1$  obtemos a direção  $d = (d' - 8)^T$ , que é uma direção de descida, pois

$$(J^T(0,1)R(0,1))^T d = -32 < 0.$$

O Exemplo anterior nos mostra que é possível contornar o problema de não invertibilidade de  $J^T(0,1)J(0,1)$ . Para tanto, dado um  $\lambda \geq 0$ , podemos tomar como aproximação da Hessiana de  $f$  a matriz  $(J^T(0,1)J(0,1) + \lambda I)$ . Este método é denominado **Método de Levenberg-Marquardt** que é o tema deste trabalho, por isso vamos apresentá-lo na próxima seção.

### 3 O Método de Levenberg-Marquardt

O Exemplo 2.1 mostrou que nem sempre a matriz  $J^T(x_k)J(x_k)$  é definida positiva. Assim, em 1944, Levenberg [4] propôs a introdução de um parâmetro  $\lambda_k \geq 0$  na diagonal de  $J^T(x_k)J(x_k)$ , de modo a estabelecer o sistema abaixo

$$(J^T(x_k)J(x_k) + \lambda_k I)d_k = -J^T(x_k)R(x_k). \quad (4)$$

Alguns anos depois, em 1963, Marquardt [1] criou o método que tem uma boa performance na iteração associando o método das séries de Taylor e o método do gradiente. Este possui uma rápida convergência ao ponto inicial que pode estar fora da região de convergência de outros métodos.

#### 3.1 Bases teóricas para o algoritmo

O primeiro passo para esta fundamentação teórica será apresentar o próximo lema que garante que  $J^T(x_k)J(x_k) + \lambda_k I$  sempre é definida positiva. Assim, justificamos que o método está bem definido, já obtemos a cada iteração uma única solução para o sistema 4.

**Lema 3.1.** *Considere as matrizes  $M, I \in \mathbb{R}^{m \times n}$ , para todo  $\lambda > 0$ , a matriz  $A^T A + \lambda I$  é definida positiva.*

O que chama atenção no método é a importância do número  $\lambda_k$  e, por isso, este recebe um nome especial. O parâmetro  $\lambda_k$  é denominado como parâmetro de **damping**. O parâmetro  $\lambda_k$  influencia tanto na direção quanto no tamanho do passo, não tornando necessária uma busca linear para descobrir o tamanho do passo ótimo a ser dado a cada iteração. Assim, definimos a direção

$$d(\lambda) = -(J^T(x_k)J(x_k) + \lambda I)^{-1}J^T(x_k)R(x_k). \quad (5)$$

Vale observar que quando  $\lambda = 0$ , obtemos a direção de Gauss-Newton, enquanto que se  $\lambda$  tender a infinito, teremos  $d$  como um múltiplo da direção de máxima descida, porém sua norma tenderá a zero como os teoremas abaixo, apresentados por Marquardt, demonstram. Ou seja, nas primeiras iterações o método converge a partir do gradiente, mas quando está próximo da convergência, o método de assemelha ao método de Gauss-Newton tendo uma ordem de convergência quadrática.

**Teorema 3.2** ([1], Teorema 2). *Seja  $d(\lambda)$ . Então,  $\|d(\lambda)\|$  é uma função contínua decrescente de  $\lambda$  tal que  $\lim_{\lambda \rightarrow +\infty} \|d(\lambda)\| = 0$ .*

**Teorema 3.3** ([1], Teorema 3). *Seja  $\gamma(\lambda)$  o ângulo entre  $d(\lambda)$  (direção de Gauss-Newton) e  $-g$  (direção de máxima descida). Então,  $\gamma(\lambda)$  é uma função contínua monótona decrescente de  $\lambda$ , tal que  $\gamma(\lambda) \rightarrow 0$  se  $\lambda \rightarrow +\infty$ . Além disso,  $d(\lambda)$  rotaciona na direção  $-g$  quando  $\lambda \rightarrow +\infty$ .*

O objetivo se resume a encontrar um parâmetro de damping suficientemente grande para que  $d_k$  proporcione uma boa redução no modelo, mas de forma que o passo não seja muito pequeno para garantir a convergência da sequência. É interessante que perto da vizinhança de convergência, o método se comporta como o de Gauss-Newton apresentando uma convergência quadrática. Dessa forma, escolher um bom parâmetro de damping é importante.

### 3.2 Escolha do parâmetro de damping

A grande ideia do Método de Levenberg-Marquardt foi conseguir através do parâmetro de damping "ajustar" a matriz  $J(x_k)^T J(x_k)$  adicionando  $\alpha_k$  em sua diagonal principal. Caso tal matriz já esteja definida positiva, voltamos essencialmente para o Método de Gauss-Newton.

Um modo de determinar se uma matriz é ou não definida positiva e aplicar a decomposição de Cholesky, que é aplicado somente para matrizes definidas positivas. Isto é, caso a decomposição falhe, devemos encontrar um parâmetro  $\alpha_k$  tal que  $J(x_k)^T J(x_k) + \alpha_k I$  seja definida positiva.

Lembre-se que o objetivo é encontrar  $x \in \mathbb{R}^n$  onde  $R(x) = 0$ , mais ainda lembre-se que  $\lambda$  influencia na direção e no tamanho do passo e, por isso definimos  $d(\lambda)$  como na Equação (5).

Vamos considerar o caso em que este  $x$  exista e  $R(x) = 0$ , em outras palavras, o caso em que  $R(x)$  é zero residual. Segundo Davis e Whittinh [5], a determinação do parâmetro pode ser feita em função da expressão

$$\lambda = \frac{(J(x)^T R(x))^T v}{f(x)}, \quad (6)$$

onde  $v = d'(0)$ .

Caso  $R(x)$  não seja zero residual, através do Método de Cauchy é possível mostrar que

$$\lambda = \frac{1}{t(J(x)^T R(x))^T v},$$

onde  $t > 0$  é o tamanho do passo a ser dado na direção oposta ao gradiente.

Existem varias propostas para o parâmetro de damping na literatura, a tabela a seguir apresenta 5 destes, os quais também poderão ser usados na aplicação do algoritmo.

Sigla	Parâmetro de damping
1	$\frac{2  J(x)^T R(x)  ^2}{  R(x_k)  ^2}$
2	$  J(x)^T R(x)  ^2$
3	$  J(x)^T R(x)  $
4	$  R(x)  ^2$
5	$\frac{  J^T(x)R(x)  ^2}{  R(x_k)  ^2}$

### Critérios de parada

Os experimentos feitos nesse trabalho utilizou dois tipos de critério de parada.

- **Quando o algoritmo converge para a solução do problema.** Colocamos como critério de parada  $||J(x)^T R(x)|| < \varepsilon$ .
- **Número máximo de iterações excedido.** Colocamos um número máximo de iterações  $itmax(= 1000)$ . Quando o número de iterações é excedido o algoritmo para e da como solução o valor  $x$  da última iteração, que pode não ser uma boa aproximação.

## 4 Análise Parâmetro de Damping

Foi feita uma análise utilizando os parâmetros supracitados e os resultados foram obtidos analisando individualmente por tipo: gaussiana, parabólica, cúbica e logarítmica. Apresentamos a média de iterações utilizadas para resolver os problemas em um total de 50 exemplos para cada tipo e cada parâmetro de Damping escolhido na seguinte tabela:

Tipo	1	2	3	4	5
<i>Gaussiana</i>	-	-	-	-	22.78
<i>Parabólica</i>	8.66	-	15.88	-	6.7
<i>Cúbica</i>	8.26	-	20.44	-	6.56
<i>Logarítmica</i>	-	-	-	-	-

O símbolo “-” representa que o número iterações excedeu um 1000 em algum dos 50 problemas.

## Referências

- [1] D.W. Marquardt, “An algorithm for least-squares estimation of nonlinear parameters,” *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431-441, 1963.
- [2] J. Revels, M. Lubin, e T. Papamarkou. Forward-mode automatic differentiation in Julia. arXiv:1607.07892, 2016.
- [3] K.A.Benatti e A.A.Ribeiro, O método de Levenberg-Marquardt para o problema de Quadrados Mínimos não Linear, SMNE, UFPR, 2017.
- [4] K. Levenberg, “A method for the solution of certain problems in least squares”. *Quart. Ap. Math.* 2, 1944, pp. 164–168.
- [5] M. Davies e I.J. Whitting, “A modified form of Levenberg’s correction”. Em *Numerical Methods for Nonlinear Optimization*, F. A. Lootsma (ed.) London, Academic Press, 1972, pp. 191–201.