



**Simpósio de Métodos  
Numéricos em Engenharia**

**25 a 27 de outubro, 2017**

# *O Método de Levenberg-Marquardt para o Problema de Quadrados Mínimos não Linear*

Kléber Aderaldo Benatti

Ademir Alves Ribeiro

Departamento de Matemática, Universidade Federal do Paraná

Curitiba, Brasil

kleber.benatti@hotmail.com , ademir.ribeiro@ufpr.br

**Resumo**—Neste trabalho apresentamos uma revisão de conceitos acerca do método de Levenberg-Marquardt, utilizado para solução do problema de quadrados mínimos não linear. Além da abordagem clássica do método, constam neste trabalho duas contribuições por nós estabelecidas. A primeira contribuição é a sugestão de um novo parâmetro de *damping*, ou parâmetro de Levenberg-Marquardt, que está diretamente ligado ao desempenho do método. A segunda contribuição estabelece uma nova maneira de resolução do subproblema relacionado ao método para problemas mal escalados, utilizando decomposições matriciais pautadas em direções conjugadas.

**Palavras-chave**—Quadrados Mínimos Não Linear, Levenberg-Marquardt, Parâmetros de *damping*, Problemas mal escalados.

## I. INTRODUÇÃO

A modelagem de problemas através de equações é comumente utilizada em ciências experimentais, como a Física, Biologia, Economia, entre outras. Por isso, é importante o desenvolvimento da teoria matemática para tratamento desses sistemas. Um caso particular de tais sistemas é o linear, onde dada uma matriz  $A \in \mathbb{R}^{m \times n}$  e um vetor  $b \in \mathbb{R}^m$  nosso objetivo é encontrar  $x \in \mathbb{R}^n$  tal que

$$Ax = b. \quad (1)$$

Apesar da grande aplicabilidade desses modelos a problemas reais, há casos que não podem ser tratados a partir de modelos lineares.

Nesses casos, podemos considerar a função  $R : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , sendo nosso objetivo a busca por  $x \in \mathbb{R}^n$  tal que

$$R(x) = 0. \quad (2)$$

Nos sistemas lineares, é natural considerar o caso em que o sistema (1) não tem solução, sendo uma proposta alternativa encontrar uma solução aproximada em determinado sentido. Uma das propostas gera o problema de quadrados mínimos. Assim como no caso dos sistemas lineares, é possível considerar no caso não linear um tratamento alternativo para o problema quando o mesmo não tem solução. Este tratamento gera o problema de quadrados mínimos não linear. Veremos a seguir como isso é feito, e de que maneira o problema de quadrados mínimos não linear pode ser resolvido. Tal abordagem pode modelar diversos problemas reais, como o tratamento de modelos de difusão de produtos [18], análise de dados [7] e tratamento de imagens [19].

## II. PROBLEMA DE QUADRADOS MÍNIMOS NÃO LINEAR

Como citado anteriormente, contemplando a possibilidade de não haver solução para o sistema (2), nosso intuito consiste em resolver o problema

$$\min_{x \in \mathbb{R}^n} f(x), \quad (3)$$

com

$$f(x) = \frac{1}{2} \|R(x)\|^2, \quad (4)$$

denominado problema de quadrados mínimos não linear. Note que se o sistema (2) tem solução,  $x^*$ , então há uma equivalência

entre os problemas (2) e (3), no sentido de que as soluções dos problemas coincidem. Veremos a seguir métodos iterativos clássicos para resolução do problema de quadrados mínimos não linear, como o método de Newton e o método de Gauss-Newton.

#### A. Método de Newton

Pela definição de  $f$  apresentada em (4), temos que

$$\nabla f(x) = J^T(x)R(x) \quad (5)$$

e

$$\nabla^2 f(x) = J^T(x)J(x) + S(x), \quad (6)$$

onde  $[J(x)]_{i,j} = \frac{\partial r_i(x)}{\partial x_j}$  e  $S(x) = \sum_{k=1}^m r_k(x)\nabla^2 r_k(x)$ , sendo  $r_k(x)$  a  $k$ -ésima coordenada de  $R(x)$ ,  $k = 1, \dots, m$ . Assim, considerando o Método de Newton clássico para minimização de problemas irrestritos, temos que a direção de minimização, dita direção de Newton, é tal que

$$(J^T(x_k)J(x_k) + S(x_k))d_k = -J^T(x_k)R(x_k). \quad (7)$$

Nestas circunstâncias, haverá apenas uma solução para (7) se  $\nabla^2 f(x)$  for inversível, e caso contrário, não haverá solução ou haverá infinitas direções satisfazendo o sistema. Assim, esta condição é fundamental para que o método esteja bem definido.

É provado em [17, Seção 5.2.3] que, mediante algumas hipóteses sobre  $f$ , o método de Newton gera uma sequência que converge localmente com taxa de convergência quadrática.

Pela dificuldade em calcular  $S(x)$  em problemas de grande porte, foram desenvolvidas técnicas de aproximação de tal matriz, gerando os métodos Quasi-Newton, que não serão abordados em nosso estudo. Uma outra alternativa é aproximar  $R(x)$  por um modelo afim e tentar resolver o problema de quadrados mínimos associado, como mostraremos a seguir.

#### B. Método de Gauss-Newton

Consideremos a aproximação de primeira ordem de  $R(x)$  em torno de  $x_k$ ,

$$R(x) \approx R(x_k) + J(x_k)d, \quad (8)$$

onde  $d = x - x_k$ . Desta forma, o problema (3) é substituído por

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} \|R(x_k) + J(x_k)d\|^2, \quad (9)$$

que tem por solução a direção  $d_k$  que satisfaz

$$(J^T(x_k)J(x_k))d_k = -J^T(x_k)R(x_k), \quad (10)$$

denominada direção de Gauss-Newton.

É provado em [3] que mediante algumas hipóteses o método de Gauss-Newton converge localmente para solução do problema (3), com taxa de convergência quadrática. O método está bem definido desde que  $J^T(x_k)J(x_k)$  seja inversível, e portanto definida positiva.

Veremos na seção a seguir uma modificação do método de Gauss-Newton para garantir a boa definição de seus iterandos.

### III. MÉTODO DE LEVENBERG-MARQUARDT

Para garantir a boa definição das direções de minimização, Levenberg [8] propôs inicialmente a introdução de um parâmetro  $\lambda_k \geq 0$  na diagonal de  $J^T(x_k)J(x_k)$  ao sistema de Gauss-Newton, estabelecendo assim o seguinte sistema:

$$(J^T(x_k)J(x_k) + \lambda_k I)d_k = -J^T(x_k)R(x_k). \quad (11)$$

Tal método foi denominado Método de Levenberg-Marquardt. Uma observação importante é que o parâmetro  $\lambda$  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. É provado em [10] que a medida em que o parâmetro  $\lambda$  cresce no sistema (11), a direção  $d_k$  tende à direção de máxima descida, enquanto que se  $\lambda$  se aproximar de zero a direção de minimização tende à direção de Gauss-Newton. Desta forma, tal parâmetro deve ser grande o suficiente para proporcionar um decréscimo na função objetivo, e deve se aproximar de zero nos estágios finais do algoritmo, de maneira que a convergência quadrática do Método de Gauss-Newton seja resgatada. Assim, a escolha deste parâmetro é um objeto de estudo do nosso trabalho.

Segundo Davis e Whitting [2], a determinação do parâmetro  $\lambda_k$  pode ser feita em função da expressão

$$\lambda_k = \frac{g_k^T v_k}{f(x_k)}, \quad (12)$$

onde  $v_k \in \mathbb{R}^n$  é escolhido de maneira conveniente. Vejamos a seguir algumas propostas para o parâmetro de *damping* existentes na literatura.

P1 Levenberg [8].

$$\lambda_k = \frac{\|g_k\|^2}{f_k}, \text{ obtido com } v_k = g_k.$$

P2 Levenberg modificado 1.

$$\lambda_k = \|g_k\|^2, \text{ obtido com } v_k = f_k g_k.$$

P3 Levenberg modificado 2.

$$\lambda_k = \|g_k\|, \text{ obtido com } v_k = \frac{f_k}{\|g_k\|} g_k.$$

Tais valores são muito convenientes, já que próximo à solução é esperado que a norma  $\|g_k\|$  seja pequena, de forma que  $\lambda_k$  será pequeno e o algoritmo herdará a boa convergência local do método de Gauss-Newton.

P4 Fan e Yuan [5].

$$\lambda_k = \|R(x_k)\|^2 = 2f_k, \text{ obtido com}$$

$$(v_k)_i = \begin{cases} \frac{2f_k^2}{p(g_k)_i} & \text{se } (g_k)_i \neq 0 \\ 0 & \text{se } (g_k)_i = 0 \end{cases}.$$

onde  $p$  é o número de componentes não nulas do gradiente  $g_k$ .

P5 Yamashita e Fukushima [20].

$$\lambda_k = \|R(x_k)\| = 2\sqrt{f_k}, \text{ obtido com}$$

$$(v_k)_i = \begin{cases} \frac{2\sqrt{f_k^3}}{p(g_k)_i} & \text{se } (g_k)_i \neq 0 \\ 0 & \text{se } (g_k)_i = 0 \end{cases}.$$

As expressões P4 e P5 fazem sentido nos casos em que  $f$  tem de fato um zero, e portanto nos estágios finais  $\lambda_k$  será pequeno. É possível encontrar a teoria de convergência do método de Levenberg-Marquardt para os parâmetros propostos por Fan e Yuan [5] e Yamashita e Fukushima [20], em seus respectivos artigos.

P6 Moré [12].

$$\lambda_k = |g_k^T v_k|,$$

com  $(v_0)_i = \left\| \frac{\partial f(x_0)}{\partial x_i} \right\|$  (a norma da  $i$ -ésima coluna da matriz Jacobiana)

$$\text{e } (v_k)_i = \max \left\{ (v_{k-1})_i, \left\| \frac{\partial f(x_0)}{\partial x_i} \right\| \right\}.$$

Nossa primeira contribuição para o trabalho consiste na criação de uma estimativa para o parâmetro  $\lambda_k$ , ao qual denotaremos P7. Dado  $\lambda_0 \geq 0$ , tomaremos a estimativa

$$\lambda_k = \frac{2\|g_k\|}{3k}, \text{ obtido com } v_k = \frac{2f_k}{3k\|g_k\|}g_k; \quad k > 0.$$

Desta forma, forçamos que o parâmetro diminua ao longo das iterações, o que é atrativo, como já foi abordado no trabalho. No final deste trabalho apresentaremos o perfil de desempenho destas estimativas.

Há ainda outras técnicas para obtenção dos parâmetros  $\lambda_k$ , consistindo na atualização do parâmetro utilizado na iteração anterior, mediante ao monitoramento do decréscimo da função. Tal estratégia necessita de um parâmetro inicial  $\lambda_0$ , e tenta tornar o parâmetro menor conforme a proximidade entre o iterando e a solução.

Consideremos a aproximação quadrática de  $f$  no ponto  $x_k$ , dada por

$$m_k(d) = f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T (J^T(x_k)J(x_k) + \lambda_k I) d, \quad (13)$$

e o coeficiente de redução relativa

$$\rho_k = \frac{ared_k}{pred_k} = \frac{f(x_k) - f(x_k + d_k)}{m_k(0) - m_k(d_k)}. \quad (14)$$

Através do parâmetro  $\rho_k$ , podemos atualizar o parâmetro  $\lambda_k$  para  $\lambda_{k+1}$ . Fazemos tal atualização da seguinte forma:

- $\rho_k$  é suficientemente grande: Indica que o modelo aproximou bem a função, e portanto aceitamos a estimativa do minimizador de  $f$ , isto é, tomamos  $x_{k+1} = x_k + d_k$ . Além disso, nosso intuito é que  $\lambda_{k+1}$  seja menor que  $\lambda_k$ . Para tal atualização multiplicaremos  $\lambda_k$  por um número real  $\alpha$ , com  $0 < \alpha < 1$  e tomamos  $\lambda_{k+1} = \alpha\lambda_k$ .
- $\rho_k$  próximo de 0 ou negativo: Neste caso, tomaremos  $x_{k+1} = x_k$  e desejamos aumentar  $\lambda_{k+1}$ , multiplicando por um parâmetro maior que 1.

Desta forma, podemos considerar parâmetros  $\eta > 0$ ,  $\alpha > 0$  e  $\beta > 0$ , e atualizar  $\lambda$  da seguinte forma:

Se  $\rho_k > \eta$ , faça

$$\lambda_{k+1} = \alpha\lambda_k \quad \delta = \beta. \quad (15)$$

Senão,

$$\lambda_{k+1} = v\lambda_k \quad \delta = 2v. \quad (16)$$

Utilizaremos  $\eta = \frac{1}{4}$  para os testes computacionais. Marquardt [10] utiliza  $\alpha = \frac{1}{\beta}$ , com  $\beta = 10$ . Esta estratégia será denominada aqui por P8, tomando  $\lambda_0 = 0.01$ . Outros autores tomam a mesma ideia de Marquardt, porém com  $\beta = 2$ . Osborne [16] propõe  $\alpha = \frac{1}{3}$  e  $\beta = \frac{3}{2}$ , além do parâmetro inicial  $\lambda_0 = 0.01$ , sendo tal estratégia denominada em nosso trabalho por P9.

Há ainda uma forma de atualização que se mostrou bastante competitiva na prática, apresentada por Nielsen [15]. Tal estratégia consiste em evitar grandes saltos do escalar  $\lambda$  entre iterações consecutivas, pois grandes oscilações do parâmetro de Levenberg-Marquardt podem exigir maiores esforços para obtenção de um valor admissível. Consideremos as constantes reais  $\eta > 0$ ,  $\mu > 1$  e  $\beta > 0$ , e um inteiro positivo ímpar  $p$ , sendo então o processo descrito por Nielsen dado por:

Se  $\rho_k > \eta$ , faça

$$\lambda_{k+1} = \lambda_k \max \left\{ \frac{1}{\mu}, 1 - (\beta - 1)(2\rho_k - 1)^p \right\}; \quad v = \beta. \quad (17)$$

Senão,

$$\lambda_{k+1} = v\lambda_k \quad v = 2v. \quad (18)$$

Madsen, Nielsen e Tingleff [9] propõem  $\beta = 2$ , e tomaremos  $\mu = 2$  e  $p = 3$ . Esta estratégia será denotada por P10, considerando o parâmetro inicial para  $\lambda$  como  $\lambda_0 = \max\{\epsilon_{abs} \max((B_0)_{ii}), \epsilon_{rel}\}$ , com  $\epsilon_{abs} = 10^{-10}$  e  $\epsilon_{rel} = 10^{-12}$ .

### A. Levenberg-Marquardt e o método de região de confiança

Muitos autores, como [12], [11] e [6], relacionam o problema de Levenberg-Marquardt a um problema de região de confiança, dado como

$$\min_{\text{s.a}} \quad \frac{1}{2} \|R(x_k) + J(x_k)d\|^2, \quad (19)$$

$$\|d\| \leq \Delta_k$$

definindo  $\Delta_k$  por

$$\Delta_k = \| -(J^T(x_k)J(x_k) + \lambda_k I)^{-1}g_k \|. \quad (20)$$

Desta forma, o método de Levenberg-Marquardt clássico é equivalente ao problema (19) desde que a solução esteja na fronteira de  $\Omega_k = \{d : \|d\| \leq \Delta_k\}$ . Tal relação é estabelecida pois, pela teoria de região de confiança, a solução do subproblema deverá satisfazer  $(J^T(x_k)J(x_k) + \lambda_k I)d_k = -J^T(x_k)R(x_k)$  e  $(\Delta_k - \|d_k\|)\lambda_k = 0$  para algum  $\lambda_k \geq 0$ .

### B. Escalamento no método de Levenberg-Marquardt

Apesar da eficácia do método de Levenberg-Marquardt perante outros métodos para solução de quadrados mínimos não linear, ainda podemos obter resultados ruins quando trabalhamos com problemas mal escalados, isto é, quando há uma grande desproporcionalidade entre as colunas ou linhas da matriz Jacobiana. Moré [14] considerou para cada iteração  $k$ , o sistema escalado

$$(J_k^T J_k + \lambda D_k^T D_k) d_k = J_k^T R_k, \quad (21)$$

onde  $D_k \in \mathbb{R}^{n \times n}$  é uma matriz diagonal de elementos

$$D_k(i, i) = \max\{(J_k^T J_k)(i, i), a\} \quad (22)$$

onde  $a > 0$ . No contexto de região de confiança, tal sistema é equivalente a resolver

$$\min_{\text{s.a}} \quad \frac{1}{2} \|J_k d + R_k\|^2 \quad (23)$$

$$\|D_k d\| \leq \Delta$$

para algum valor de  $\Delta > 0$ . Desta forma, é suposto que o mau escalamento da matriz Jacobiana  $J_k$  seja remediado. Dado o sistema acima, como a matriz  $J_k^T J_k + \lambda D_k^T D_k$  é definida positiva, podemos aplicar a decomposição de Cholesky ou o método de Gradientes Conjugados para resolver o sistema.

Outra técnica aplicada aos problemas mal escalados foi introduzida por Zhou e Si [22], e visava inicialmente uma melhor implementação para os problemas de redes neurais. Para o problema de quadrados mínimos não linear, esta técnica objetiva tanto a aceleração do método aplicado a problemas mal escalados quanto a solução do subproblema de quadrados mínimos linear. Para tal, utilizam-se decomposições matriciais específicas, diminuindo o custo computacional e memória requeridas. Vale ressaltar que há duas variações desta técnica apresentadas em [22], e que serão utilizadas nos testes numéricos como métodos para comparação.

Utilizando ideias semelhantes à técnica apresentada por Zhou e Si, nosso objetivo consiste em desenvolver uma técnica de resolução para o Método de Levenberg-Marquardt aplicado a problemas mal escalados, utilizando a decomposição matricial pautada em direções

conjugadas. Mais detalhes acerca desta decomposição se encontram em [21] e [1]. Este método será denominado Conjugate Gradient Levenberg-Marquardt (CG-LM).

Consideremos o problema de Levenberg-Marquardt escalado

$$\min_{d \in \mathbb{R}^n} \quad \frac{1}{2} \|J_k^T d + R_k\|^2 \quad (24)$$

$$\text{s.a} \quad \|S_k d\| \leq \Delta$$

para algum valor de  $\Delta > 0$ . Assim, pela teoria de Região de Confiança, a solução do problema está caracterizada através do sistema linear

$$(J_k^T J_k + \lambda S_k^T S_k)d = -J_k^T R_k \quad (25)$$

para algum valor de  $\lambda \geq 0$ . Utilizando a decomposição matricial pautada em direções conjugadas, temos que

$$P_k^T J_k^T J_k P_k = D_k \quad (26)$$

onde  $P$  tem os vetores gradientes conjugados de  $J_k^T J_k$ , e

$$D_k = \begin{pmatrix} D_{1k} & 0 \\ 0 & 0 \end{pmatrix} \quad (27)$$

com  $D_{1k}$  matriz diagonal, onde os elementos da diagonal são da forma  $d_{ii} = p_i^T J_k^T J_k p_i$ . Assim, temos que

$$J_k^T J_k = P_k^{-T} D_k P_k^{-1}. \quad (28)$$

Utilizando (28) e (25), obtemos

$$(P_k^{-T} D_k P_k^{-1} + \lambda S_k^T S_k)d = -J_k^T R_k. \quad (29)$$

Seja  $b$  tal que  $P_k^{-T} b = -J_k^T R_k$ , isto é,  $b = -P_k^T J_k^T R_k$ , obtendo assim

$$(P_k^{-T} D_k P_k^{-1} + \lambda S_k^T S_k)d = P_k^{-T} b. \quad (30)$$

Desta maneira, podemos tomar  $S_k = P_k^{-1}$ , e portanto a direção  $d$  deve satisfazer

$$(P_k^{-T} D_k P_k^{-1} + \lambda P_k^{-T} P_k^{-1})d = P_k^{-T} b, \quad (31)$$

ou de maneira equivalente

$$(D_k + \lambda I)P_k^{-1}d = b. \quad (32)$$

Considerando  $\tilde{d} = P_k^{-1}d$  e reescrevendo o sistema em blocos, obtemos

$$\begin{pmatrix} D_{1k} + \lambda I & 0 \\ 0 & \lambda I \end{pmatrix} \begin{pmatrix} \tilde{d}_1 \\ \tilde{d}_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad (33)$$

e portando

$$\tilde{d} = \begin{pmatrix} \tilde{d}_1 \\ \tilde{d}_2 \end{pmatrix} = \begin{pmatrix} (D_{1k} + \lambda I)^{-1}b_1 \\ \frac{1}{\lambda}b_2 \end{pmatrix}. \quad (34)$$

Dada a solução em (34), é fácil voltar à definição de  $d$ , pois

$$d = P_k \tilde{d}. \quad (35)$$

Vejam o algoritmo da técnica a seguir.

**Algoritmo 1 : Método CG-LM**

Dados  $x_0 \in \mathbb{R}^n$  e a sequência de parâmetros  $\lambda_k$ , soluções das raízes seculares associadas ao problema, defina  $k = 0$ . Faça

Compute a decomposição  $P_k^T J_k^T J_k P_k = D_k$ ;  
 $b = -P_k^T J_k^T R_k$ ;  
 $\tilde{d} = \begin{pmatrix} (D_{1k} + \lambda_k I)^{-1} b_1 \\ \frac{1}{\lambda_k} b_2 \end{pmatrix}$ ;  
 $d = P_k \tilde{d}$ ;  
 $x_{k+1} = x_k + d$ ;  
 $k = k + 1$ ;

fim

Note que o algoritmo não necessita da inversa de  $P_k$ , mas apenas da computação da própria decomposição  $P_k^T J_k^T J_k P_k = D_k$ , tornando o algoritmo mais barato em relação ao que a teoria aparentemente apresenta.

É possível provar a convergência de nosso método a partir da convergência do método de Região de Confiança, como enunciado no próximo teorema.

**Teorema** Seja  $R \in \mathcal{C}^1$ , e suponhamos que a sequência  $(x_k)$  gerada pelo método proposto cumpra:

- (1)  $\nabla f$  Lipschitz, com  $f(x) = \frac{1}{2} \|R(x)\|^2$ .
- (2)  $\exists c > 0$  tal que  $c \leq \sqrt{\alpha_{min}^k}, \forall k \in \mathbb{N}$ , onde  $\alpha_{min}^k$  é o menor autovalor de  $S_k^T S_k$ .
- (3) As Jacobianas são uniformemente limitadas, isto é,  $\exists \mu > 0$  tal que  $\|J_k\| \leq \mu \forall k \in \mathbb{N}$ .

Nestas condições, temos que a sequência  $(x_k)$  gerada pelo Algoritmo 1 cumpre

$$\lim_{k \rightarrow \infty} \nabla f(x_k) = 0. \quad (36)$$

Em outras palavras, nosso método converge globalmente para solução do problema.

*Demonstração:* Apresentada no Teorema 3.6 de [1]

### C. Experimentos numéricos

Baseados na teoria desenvolvida ao longo deste trabalho, abordaremos alguns testes numéricos relativamente à escolha do parâmetro de Levenberg-Marquardt e o desempenho do método CG-LM aplicado a problemas mal escalados.

Os testes foram realizados em um computador portátil com processador Intel Core i3-2367M, velocidade do clock de 1.40 GHz, 4 GB de memória RAM e sistema operacional Windows 7 Home Premium com arquitetura 64 bits. Os algoritmos foram implementados no software Matlab versão 7.14 (R2012a). Em todos nossos experimentos numéricos utilizaremos todas as funções do banco de proposto por Moré, Garbow e Hillstrom [13]. Todos os experimentos feitos neste trabalho tiveram como critérios de parada a norma do gradiente menor que uma tolerância ( $\epsilon = 10^{-5}$ ) e o número

máximo de iterações ( $K_{max} = 1000$ ). As análises serão feitas através de perfis de desempenho [4]. Vale ainda ressaltar que o tempo de execução dos algoritmos foi calculado através dos comandos tic e toc, próprios do Matlab.

Primeiramente, apresentaremos os testes efetuados utilizando as escolhas P1 a P7, que são definidas sobre a iteração  $k$ , sem a análise do coeficiente de redução relativa. Serão apresentados os perfis de desempenho associados aos métodos, relativamente ao tempo e ao número de iterações. Vale salientar que todos perfis de desempenho apresentados neste trabalho estão em escala logarítmica.

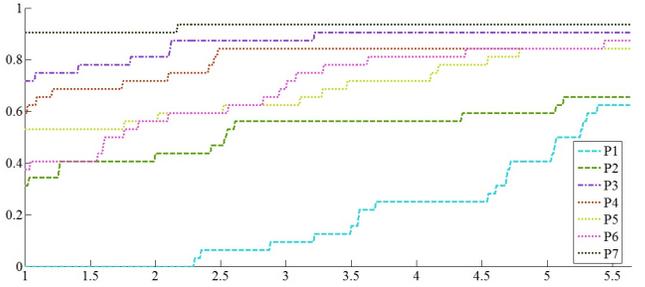


Figura 1: Perfil de desempenho dos parâmetros P1-P7 de Levenberg-Marquardt em função do tempo.

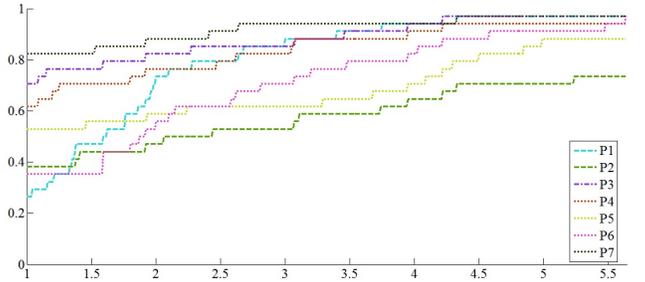


Figura 2: Perfil de desempenho dos parâmetros P1-P7 de Levenberg-Marquardt em função do número de iterações.

Através desta análise de desempenho, podemos notar que nossa proposta se mostrou muito eficiente e robusta em relação aos parâmetros clássicos. Agora nosso objetivo é avaliar o desempenho dos parâmetros de Levenberg-Marquardt em um contexto geral, contendo os parâmetros P8 a P10 definidos a partir da análise do coeficiente de redução relativa. Apresentaremos o perfil de desempenho dos parâmetros de Levenberg-Marquardt, considerando uma seleção contendo o pior e os melhores desempenhos feita a priori.

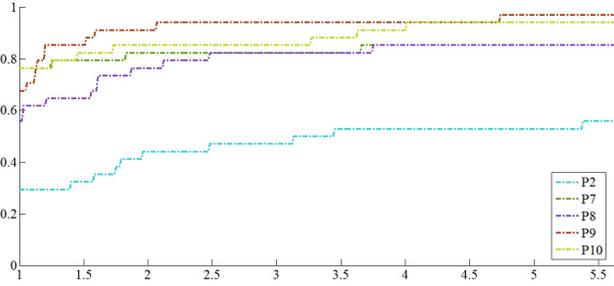


Figura 3: Perfil de desempenho geral dos parâmetros de Levenberg-Marquardt em função do tempo.

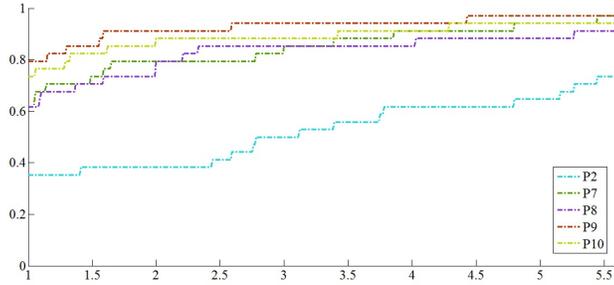


Figura 4: Perfil de desempenho geral dos parâmetros de Levenberg-Marquardt em função do número de iterações.

Analisando os resultados, podemos ver que relativamente ao tempo, nossa proposta de parâmetro, P7, continua sendo o mais eficiente, resolvendo 77% dos problemas mais rapidamente, juntamente com o parâmetro P10. Relativamente a robustez, o parâmetro P7 resolveu 85% dos problemas, mesmo índice obtido por P8. Podemos notar que o parâmetro mais robusto nesta análise foi P9, tendo resolvido 98% dos problemas.

Segundo a análise acima, podemos concluir que nossa proposta de parâmetro é promissora relativamente aos parâmetros expressos na literatura.

Agora, nosso objetivo é discorrer acerca do desempenho computacional das técnicas utilizadas para solução do problema de Levenberg-Marquardt aplicado à problemas mal escalados. Antes de apresentar o perfil de desempenho dos algoritmos, discorreremos acerca da computação dos vetores A-conjugados exigidos na técnica CG-LM.

Notamos que na computação dos vetores A-conjugados é possível utilizar a informação obtida das iterações anteriores, de forma a construir um algoritmo mais barato computacionalmente. Apresentaremos agora um algoritmo para construção dos vetores A-conjugados explorando essa característica.

**Algoritmo 2 : Obtenção dos vetores A-conjugados**

Dada matriz  $A$  simétrica semidefinida positiva com  $\text{posto}(A) = r$  e  $\{v_1, v_2, \dots, v_r\}$  base de  $\text{Im}(A)$ .

Defina

$p_1 = v_1$ ;  
 $aux = p_1^T A$ ;  
 $Aux = p_1 \frac{aux}{aux.p_1}$ .

De  $i = 2 : r$  faça

$p_i = (I - Aux)v_i$ ;  
 $aux = p_i^T A$ ;  
 $Aux = Aux + p_i \frac{aux}{aux.p_i}$ .

Fim

Agora, a fim de fazermos a comparação de desempenho das técnicas, denominaremos a técnica de escalamento clássico definido em (21) e resolvido por decomposição de Cholesky por  $E1$ , utilizando  $a = 1$ . A mesma técnica resolvida por gradientes conjugados será denotada por  $E2$ . As variações da técnica proposta por Zhou e Si apresentadas em [22] serão denotadas por  $E3$  e por  $E4$ , respectivamente. Todas estas técnicas utilizarão o parâmetro  $\lambda$  proposto em P9. Já nossa proposta de resolução, apresentada pelo Algoritmo 1, será denotada por  $E5$ , utilizando o parâmetro  $\lambda$  proposto em P7. Assim, analisemos o desempenho dos algoritmos relativamente ao tempo e ao número de iterações, como mostra a seguir.

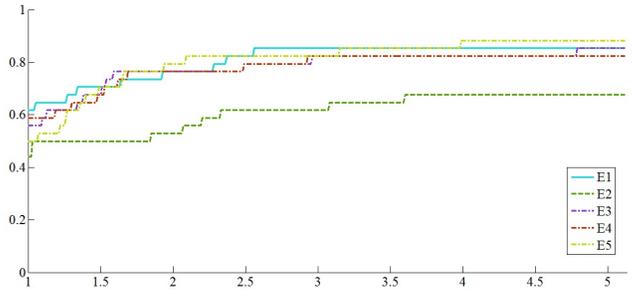


Figura 5: Perfil de desempenho dos algoritmos E1 a E5 em função do tempo. Escala de eficiência.

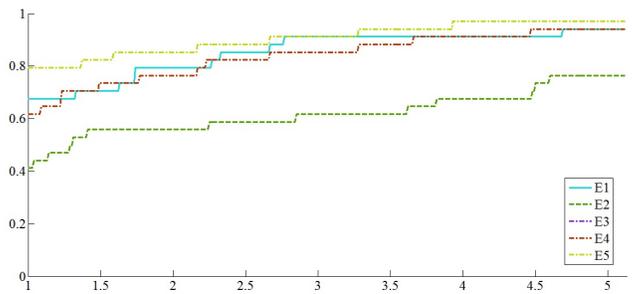


Figura 6: Perfil de desempenho dos algoritmos E1 a E5 em função do tempo. Escala de robustez.

Podemos notar nos resultados acima que, relativamente ao tempo, o método CG-LM (E5) resolveu 50% dos problemas mais rapidamente, enquanto que o mais eficiente resolveu 62% dos problemas de forma mais veloz. Já no quesito robustez, o método CG-LM obteve o melhor resultado, tendo resolvido um total de 90% dos problemas. Os métodos E1 e E3 resolveram 86% dos problemas.

Através da análise dos perfis de desempenho, pode-se notar que nossa proposta de extensão se mostrou a mais robusta. Cabe salientar que desde que haja uma maneira eficiente para o cálculo de uma base para imagem e uma base para o núcleo de  $J_k^T J_k$ , nossa proposta apresenta um baixo custo computacional, não envolvendo soluções de sistemas nem inversas de matrizes, exceto a inversa de uma matriz diagonal. Podemos assim concluir que a aplicação de nossa proposta é uma boa escolha para o método de Levenberg-Marquardt para problemas mal escalados.

#### REFERÊNCIAS

- [1] K. A. Benatti. *O Método de Levenberg-Marquardt para o Problema de Quadrados Mínimos não Linear*. Dissertação de Mestrado, Universidade Federal do Paraná, Curitiba, 2016.
- [2] 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.
- [3] J. E. Dennis e R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. SIAM, Philadelphia, PA, 1996.
- [4] E. D. Dolan e J. J. Moré, “Benchmarking optimization software with performance profiles”. *Math. Programming* (Ser. A) 91, 1981, pp. 201–213.
- [5] J. Fan e Y. Yuan, “On the convergence of the a new Levenberg-Marquardt method”. Technical Report, AMSS, Chinese Academy of Sciences, Beijing, China, 2001.
- [6] J. L. C. Gardenghi e S. A. Santos, “Sistemas não lineares via região de confiança: o algoritmo de Levenberg-Marquardt”. Campinas: UNICAMP, 45. Relatório de Pesquisa. Departamento de Matemática Aplicada, Universidade de Campinas, 2011.
- [7] M. L. Johnson, J. J. Correia, D. A. Yphantis e H. R. HalvorSON, “Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques”. *Biophysical Journal*, 36 (3), 1981, pp. 575-588.
- [8] K. Levenberg, “A method for the solution of certain problems in least squares”. *Quart. Ap. Math.* 2, 1944, pp. 164–168.
- [9] K. Madsen, H. B. Nielsen e O. Tingleff, “Methods for non-linear least-squares problems”, 2.ed. *Lecture note, Informatics and Mathematical Modelling*, Technical University of Denmark, Lyngby, Dinamarca, 2004.
- [10] D. Marquardt, “An algorithm for least squares estimation of nonlinear parameters”. *J. Soc. Indust. Ap. Math.* 11, 1963, pp. 431–441.
- [11] L. A. da S. Medeiros, “Acelerando o método de Levenberg-Marquardt para a minimização da soma de quadrados de funções com restrições em caixa”. Campinas: UNICAMP, 120. Tese (Doutorado em Matemática Aplicada) - Programa de Pós-Graduação em Matemática Aplicada, Departamento de Matemática Aplicada, Universidade de Campinas, 2008.
- [12] J. J. Moré, “The Levenberg-Marquardt algorithm: implementation and theory”. Em *Proceedings of the 1977 Dundee conference on numerical analysis*, Lecture Notes in Mathematics 630. G. A. Watson. (ed.) Berlin, Springer, 1978, pp. 105–116.
- [13] J. J. Moré, B. S. Garbow e K. E. Hillstrom, “Testing unconstrained optimization software”. *ACM Trans. Math. Software* 7, 1981, pp. 17–41.
- [14] J. J. Moré, “Lecture notes in mathematics”. 630, 1977, pp. 105-116.
- [15] H. B. Nielsen, “Damping parameter in Marquardt’s method”. Report IMM-REP-1999-05, Technical University of Denmark, Lyngby, Dinamarca, 2005.
- [16] M. R. Osborne, “Nonlinear least squares—the Levenberg algorithm revisited”. *J. Austral. Math. Soc. (Ser. B)* 19, 1976, pp. 343–357.
- [17] A. A. Ribeiro e E. W. Karas, *Otimização Contínua: aspectos teóricos e computacionais*. Cengage Learning, São Paulo, Brasil, 2013.
- [18] V. Srinivasan e C. H. Mason, “Nonlinear least squares estimation of new product diffusion models”. *Marketing science*, 5(2), 1986, pp. 169-178.
- [19] R. Szeliski e S. B. Kang, “Recovering 3D shape and motion from image streams using nonlinear least squares”. *Journal of Visual Communication and Image Representation*, 5(1), pp. 1994, 10-28.
- [20] N. Yamashita e M. Fukushima, “On the rate of convergence of the Levenberg-Marquardt method”. Technical Report 2000-008, Kyoto University, Kyoto, Japão, 2000.
- [21] L. P. Wang e Y. Jinyun “Conjugate decomposition and its applications.” *Journal of the Operations Research Society of China* 1.2 (2013): 199-215.
- [22] G. Zhou e J. Si, “Advanced neural-network training algorithm with reduced complexity based on Jacobian deficiency”. *IEEE Trans. on Neural Net.* 9, 1998, pp. 448–453.