

UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE MATEMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA
(Mestrado)

EDILAINE DOS SANTOS DURAN SCHWERTNER

Uma classe de métodos do tipo Levenberg-Marquardt com passos múltiplos para problemas de Otimização de Menor Valor Ordenado

Maringá-PR

2020

Edilaine dos Santos Duran Schwertner

Uma classe de métodos do tipo Levenberg-Marquardt com passos múltiplos para problemas de Otimização de Menor Valor Ordenado

Dissertação apresentada ao Programa de Pós-Graduação em Matemática do Departamento de Matemática, Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito para obtenção do título de Mestre em Matemática.

Área de concentração: Matemática Aplicada.

Orientador: Dr. Emerson Vitor Castelani

Coorientador: Dr. Francisco Nogueira Calmon Sobral

Maringá

2020

Dados Internacionais de Catalogação na Publicação (CIP)
(Biblioteca Setorial BSE-DMA-UEM, Maringá, PR, Brasil)

S415u Schwertner, Edilaine dos Santos Duran
Uma classe de métodos do tipo Levenberg-Marquardt com passos múltiplos para problemas de Otimização de Menor Valor Ordenado / Edilaine dos Santos Duran Schwertner. -- Maringá, 2020.
xx 76p. il., color.

Orientador: Prof. Dr. Emerson Vitor Castelani.
Coorientador: Prof. Dr. Francisco Nogueira Calmon Sobral.

Dissertação (mestrado) - Universidade Estadual de Maringá, Centro de Ciências Exatas, Programa de Pós-Graduação em Matemática - Área de Concentração: Matemática Aplicada, 2020.

1. Levenberg-Marquardt com passos múltiplos. 2. Otimização de menor valor ordenado. 3. Método de Levenberg-Marquardt. 4. Quadrados mínimos não linear. 5. Levenberg-Marquardt with multiple steps. 6. Low order-value optimization. 7. Non-linear least squares. I. Castelani, Emerson Vitor, orient. II. Sobral, Francisco Nogueira Calmon, orient. III. Universidade Estadual de Maringá. Centro de Ciências Exatas. Programa de Pós-Graduação em Matemática - Área de Concentração: Matemática Aplicada. IV. Título.

CDD 22.ed. 519.6

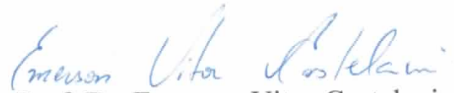
Edilson Damasio CRB9-1.123

EDILAINE DOS SANTOS DURAN SCHWERTNER

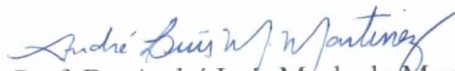
**UMA CLASSE DE MÉTODOS DO TIPO LEVENBERG-MARQUARDT COM PASSOS
MÚLTIPLOS PARA PROBLEMAS DE OTIMIZAÇÃO DE MENOR VALOR ORDENADO**

Dissertação apresentada ao Programa de Pós-Graduação em Matemática do Departamento de Matemática, Centro de Ciências Exatas da Universidade Estadual de Maringá, como parte dos requisitos necessários para a obtenção do título de Mestre em Matemática tendo a Comissão Julgadora composta pelos membros:

COMISSÃO JULGADORA:



Prof. Dr. Emerson Vitor Castelani
Universidade Estadual de Maringá (Presidente)



Prof. Dr. André Luís Machado Martinez
Universidade Tecnológica Federal do Paraná



Prof. Dr. Emerson Luiz do Monte Carmelo
Universidade Estadual de Maringá

Aprovada em 27 de fevereiro de 2020.

Local de defesa: Auditório DMA, Bloco F67, campus da Universidade Estadual de Maringá.

Dedico esta dissertação aos meus pais, Candido e Neuza, que mesmo na simplicidade me mostraram que o caminho do conhecimento é libertador.

AGRADECIMENTOS

Agradeço primeiramente à Deus por sempre estar comigo nos momentos mais difíceis, em minhas dúvidas internas, e pelo seu amor por todos nós.

À meu marido, por sempre estar ao meu lado, me ajudando, me incentivando e por nunca me deixar desistir.

Aos professores Emerson Vitor Castelani e Francisco Nogueira Calmon Sobral, por aceitaram me orientar, e por todo apoio na realização deste trabalho.

Aos demais professores e funcionários do Programa de Pós-Graduação em Matemática, e todos os que contribuíram para que este trabalho fosse concluído.

Ao PMA e à UEM, pela oportunidade que me foi dada.

À CAPES, pelo apoio financeiro.

“Tudo que a mente humana pode conceber,
ela pode conquistar.”

Napoleon Hill

RESUMO

Neste trabalho dissertamos acerca do problema de quadrados mínimos, do Método de Levenberg-Marquardt e de três métodos do tipo Levenberg-Marquardt com passos múltiplos, chamados Levenberg-Marquardt Modificado, Levenberg-Marquardt Modificado Acelerado e Levenberg-Marquardt Modificado Acelerado Adaptado. Realizamos uma revisão literária do problema de Otimização de Menor Valor Ordenado (LOVO). Em seguida, trabalhamos com adaptações dos métodos do tipo Levenberg-Marquardt com passos múltiplos, abordados como uma nova estratégia para resolução de problemas do tipo LOVO irrestritos com convergência para pontos fracamente críticos. Neste contexto, junto com técnicas de região de confiança, apresentamos resultados acerca da convergência global do método. Por fim, por meio de experimentos numéricos comparamos o desempenho dos novos métodos com a variação do Método de Gauss-Newton para problemas LOVO e com uma variação do Método de Levenberg-Marquardt com busca linear para problemas LOVO, já conhecidos na literatura.

Palavras-chave: Levenberg-Marquardt com passos múltiplos, Otimização de Menor Valor Ordenado, Quadrados Mínimos Não Linear.

ABSTRACT

In this work we talk about the Least Squares problem, the Levenberg-Marquardt method and three methods of the Levenberg-Marquardt type with multiple steps, called Levenberg-Marquardt Modified, Levenberg-Marquardt Modified Accelerated and Levenberg-Marquardt Modified Accelerated Adapted. We performed a literary review of the Lower Ordered Value Optimization problem (LOVO). Then, we work with adaptations of the Levenberg-Marquardt type methods with multiple steps, approached as a new strategy for solving unrestricted LOVO problems with convergence to weakly critical points. In this context, together with trust-region techniques, we present results about the global convergence of the method. Finally, through numerical experiments, we compared the performance of the new methods with the variation of the Gauss-Newton method for LOVO problems and with a variation of the Levenberg-Marquardt method with linear search for LOVO problems, already known in the literature.

Keywords: Levenberg-Marquardt with multiple steps, Low Order-Value Optimization, Nonlinear Least Squares.

LISTA DE FIGURAS

2.1	Exemplo de iteração do Método de Região de Confiança.	14
5.1	Gráfico da função LOVO.	39

LISTA DE TABELAS

5.1	Dados para o exemplo	38
7.1	Parâmetros utilizados.	62
7.2	Modelos utilizados para os testes.	63
7.3	Polinomial de Grau 1 para 100 pontos, com ponto inicial $x_0 = (0.0, 0.0)$	67
7.4	Polinomial de Grau 1 para 1000 pontos, com ponto inicial $x_0 = (0.0, 0.0)$. . .	67
7.5	Polinomial de Grau 1 para 5000 pontos, com ponto inicial $x_0 = (0.0, 0.0)$. . .	67
7.6	Polinomial de Grau 3 para 100 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0, 0.0)$. 68	
7.7	Polinomial de Grau 3 para 1000 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0, 0.0)$. 68	
7.8	Polinomial de Grau 3 para 5000 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0, 0.0)$. 68	
7.9	Exponencial para 100 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0, 0.0)$	69
7.10	Exponencial para 1000 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0, 0.0)$	69
7.11	Exponencial para 5000 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0, 0.0)$	69
7.12	Senoidal Tipo 01 para 100 pontos, com ponto inicial $x_0 = (1.0, 1.0, 1.0, 1.0)$. .	70
7.13	Senoidal Tipo 01 para 1000 pontos, com ponto inicial $x_0 = (1.0, 1.0, 1.0, 1.0)$. 70	
7.14	Senoidal Tipo 01 para 5000 pontos, com ponto inicial $x_0 = (1.0, 1.0, 1.0, 1.0)$. 70	
7.15	Senoidal Tipo 02 para 100 pontos, com ponto inicial $x_0 = (5.0, 5.0, 5.0, 5.0, 5.0)$. 71	
7.16	Senoidal Tipo 02 para 1000 pontos, com ponto inicial $x_0 = (5.0, 5.0, 5.0, 5.0, 5.0)$. 71	
7.17	Senoidal Tipo 02 para 5000 pontos, com ponto inicial $x_0 = (5.0, 5.0, 5.0, 5.0, 5.0)$. 71	
7.18	Logístico para 100 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0)$	72
7.19	Logístico para 1000 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0)$	72

7.20	Logístico para 5000 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0)$	72
------	---	----

SUMÁRIO

1	Introdução	1
2	Conceitos básicos	4
2.1	Álgebra Linear	4
2.1.1	Desigualdade de Cauchy-Bunyakovskii-Schwarz	4
2.1.2	Normas Matriciais	5
2.1.3	Positividade de uma matriz simétrica	7
2.1.4	Decomposição de Matrizes	7
2.2	Otimização	9
2.2.1	Condições de otimalidade para problemas de minimização sem restrições	10
2.2.2	Minimização Convexa	12
2.2.3	O Método de Região de Confiança	12
3	Problema de quadrados mínimos	18
3.1	A formulação do problema de quadrados mínimos	18
3.2	O problema de quadrados mínimos não-linear	20
4	O Método de Levenberg-Marquardt	23
4.1	Levenberg-Marquardt	23
4.2	Levenberg-Marquardt Modificado	26
4.3	Levenberg-Marquardt Modificado Acelerado	30
4.4	Levenberg-Marquardt Modificado Acelerado Adaptado	33

5	Revisão Literária do Problema LOVO	36
5.1	Otimização de Menor Valor Ordenado	36
5.2	Otimalidade do problema LOVO	39
5.3	Métodos para resolver problemas LOVO no contexto de ajuste	41
6	Levenberg-Marquardt com passos múltiplos para LOVO	44
6.1	Conceitos iniciais	44
6.2	Método de Levenberg-Marquardt Modificado para problemas LOVO	46
6.3	O Método de Levenberg-Marquardt Modificado Acelerado para problemas LOVO	57
6.4	O Método de Levenberg-Marquardt Modificado Acelerado Adaptado para problemas LOVO	59
7	Experimentos Numéricos	61
7.1	Análise dos Resultados	62
7.1.1	Funções testadas	62
7.1.2	Análises	63
7.1.3	Tabelas com Resultados	66
8	Considerações Finais e Trabalhos Futuros	73
	Bibliografia	74

INTRODUÇÃO

Dadas r funções reais $F_i : \mathbb{R}^n \rightarrow \mathbb{R}$, com $i = 1, 2, \dots, r$, e um inteiro $p \in \{1, 2, \dots, r\}$, o problema de Otimização de Menor Valor Ordenado (*Low Order-Value Optimization problem* - LOVO) procura minimizar a soma das p menores funções, tal problema foi inicialmente proposto por Andreani et al [4], em 2005, e um ano após Yano [35] defendeu sua tese de doutorado no tema. A utilização da teoria LOVO é um importante recurso para a resolução de problemas em diversas áreas, tais como Ciências, Tecnologia e Economia, sendo exemplos clássicos o alinhamento estrutural de proteínas, a determinação de padrões ocultos em conjuntos de dados, encontrados em [4] e *Value at Risk* (VaR), tema abordado em [6]. Grande parte dos trabalhos relacionados ao problema LOVO são baseados em métodos de otimização contínua que são modificados para contornar o problema da não diferenciabilidade, como veremos no decorrer do trabalho. Duas adaptações clássicas para resolver LOVO em contexto de funções de ajuste são Gauss-Newton [3] e Levenberg-Marquardt com busca linear [32]. Recentemente, foi desenvolvido e publicado do pacote *Robust Algebraic Fitting Function in Julia* (RAFF.jl), escrito na linguagem de programação Julia que emprega o Método de Levenberg-Marquardt no contexto de problemas LOVO para encontrar ajustes robustos, livres da influência de possíveis dados discrepantes, conforme [8].

Uma das principais dificuldades para resolver o problema LOVO está relacionada à avaliação da função objetivo, o qual é o passo mais caro em termos computacionais deste tipo de problema. Dessa forma, faz sentido indagarmos sobre o que aconteceria no uso de métodos com ordem de convergência mais elevada, como por exemplo, convergência cúbica ou maior, e como eles poderiam ser tratados no ambiente LOVO, tanto do ponto de vista teórico, quanto prático. Visto isto, procuramos na literatura métodos do tipo Levenberg-

Marquardt com passos múltiplos e com convergência cúbica, dos quais abordaremos neste trabalho o Método de Levenberg-Marquardt Modificado e o Método de Levenberg-Marquardt Modificado Acelerado propostas por Fan em [16] e [17] e o Método de Levenberg-Marquardt Modificado Acelerado Adaptado, proposto por Amini e Rostami em [2]. No entanto não há evidências de que existam versões de tais métodos para problemas LOVO. Nosso intuito com este trabalho é adaptar os métodos de Levenberg-Marquardt com passos múltiplos para a resolução de problemas do tipo LOVO irrestritos com convergência para pontos fracamente críticos. Além disso, empregando técnicas de região de confiança, apresentamos resultados acerca da convergência global do método, e realizamos experimentos numéricos de modo a comparar o desempenho dos novos métodos com as variações do métodos de Gauss-Newton e Levenberg-Marquardt com busca linear para problemas LOVO.

A dissertação é dividida em oito capítulos, o Capítulo 2 revisa resultados básicos acerca de Álgebra Linear e Otimização importantes para o desenvolvimento deste trabalho, enfatizando principalmente conceitos sobre Matrizes, Minimização Irrestrita e o Método de Região de Confiança. No Capítulo 3 introduzimos o problema de quadrados mínimos não linear, e apresentamos o método de Gauss-Newton.

No Capítulo 4 abordamos alguns conceitos teóricos a respeito do método de Levenberg-Marquardt e três métodos do tipo Levenberg-Marquardt com passos múltiplos, a saber Levenberg-Marquardt Modificado, Levenberg-Marquardt Modificado Acelerado e Levenberg-Marquardt Modificado Acelerado Adaptado.

O Capítulo 5 explica o problema de Otimização de Menor Valor Ordenado (LOVO), onde é apresentado o problema LOVO com convergência a pontos fracamente críticos. Descrevemos o Método de Gauss-Newton para problemas LOVO e o Método Levenberg-Marquardt com Busca Linear para problemas LOVO.

No Capítulo 6, estão presentes as contribuições deste trabalho. Nele apresentamos variações dos métodos Levenberg-Marquardt Modificado, Levenberg-Marquardt Modificado Acelerado e Levenberg-Marquardt Modificado Acelerado Adaptado no contexto dos problemas do tipo LOVO, assim como resultados de convergência global para esses métodos.

Por fim, no Capítulo 7 são apresentadas as análises dos testes numéricos compa-

rando os algoritmos LOVO Gauss-Newton e LOVO Levenberg-Marquardt que usam busca linear com os algoritmos que utilizam região de confiança.

CONCEITOS BÁSICOS

Neste capítulo apresentaremos definições e resultados importantes para o desenvolvimento deste trabalho, como resultados conhecidos de Álgebra Linear e Otimização irrestrita.

A Seção 2.1 aborda conceitos relacionados à Álgebra Linear, como a desigualdade geral de Cauchy-Bunyakovskii-Schwarz, normas matriciais, positividade de matriz e algumas decomposições matriciais clássicas. Como principais referências, citamos [28] e [33].

Já a Seção 2.2 desenvolve conceitos básicos de Otimização, como as condições de primeira e segunda ordem para minimização sem restrições, e o método de região de confiança. Como principais referências adotamos [18], [26] e [29].

2.1 Álgebra Linear

A Álgebra Linear é uma importante área da matemática devido a sua versatilidade tanto na modelagem e resolução de uma gama de problemas na engenharia, economia, física, cientistas da computação, programadores, estatísticos, biólogos, entre outros.

2.1.1 Desigualdade de Cauchy-Bunyakovskii-Schwarz

Antes de enunciar o principal resultado desta subseção, segue-se a seguinte definição.

Definição 2.1. [28, Definição 5.2.1] Seja \mathbb{V} um espaço vetorial real. Um *produto interno* é uma função que aplica um par ordenado de vetores x, y em um escalar real $\langle x, y \rangle$ de modo

que satisfaça as seguintes propriedades

- i) $\langle x, x \rangle \geq 0$ e $\langle x, x \rangle = 0 \iff x = 0$;
- ii) $\langle x, \alpha y \rangle = \alpha \langle x, y \rangle$, para todo escalar α ;
- iii) $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$;
- iv) $\langle x, y \rangle = \langle y, x \rangle$.

Qualquer espaço vetorial real que admite produto interno é chamado *espaço com produto interno*.

Seja \mathbb{V} um espaço vetorial com produto interno sobre \mathbb{R} . Dado $x \in \mathbb{V}$, a norma de x é denotada por $\|x\| = \sqrt{\langle x, x \rangle}$. A Desigualdade geral de Cauchy-Bunyakovskii-Schwarz é um importante resultado para espaços vetoriais com produto interno, pois relaciona o produto interno de dois vetores com o produto da norma desses vetores.

Teorema 2.2. (*Desigualdade de Cauchy-Bunyakovskii-Schwarz*) *Seja \mathbb{V} um espaço vetorial com produto interno. Para quaisquer $x, y \in \mathbb{V}$,*

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|, \quad (2.1)$$

Note que a igualdade se verifica se e somente se $y = \alpha x$, onde $\alpha = \frac{\langle x, y \rangle}{\|x\|^2}$.

Demonstração: Ver [28, Teorema 5.3.4]. □

2.1.2 Normas Matriciais

Em Álgebra Linear, uma norma matricial é uma norma definida para matrizes, segundo a definição a seguir.

Definição 2.3. [28, Definição 5.2.3] *Seja $\mathcal{M}(\mathbb{R})$ o conjunto das matrizes de ordens finitas. Uma norma matricial é uma função $\|\cdot\| : \mathcal{M}(\mathbb{R}) \rightarrow \mathbb{R}$ que satisfaz as seguintes propriedades,*

- i) $\|A\| \geq 0$ e $\|A\| = 0 \iff A = 0$;

- ii) $\|\alpha A\| = |\alpha| \|A\|$, para algum escalar α ;
- iii) $\|A + B\| \leq \|A\| + \|B\|$, para matrizes de mesma dimensão;
- iv) $\|AB\| \leq \|A\| \cdot \|B\|$, para matrizes que são compatíveis com a multiplicação.

Definição 2.4. [28, Definição 5.2.4] Dada uma norma vetorial definida em \mathbb{R}^p , onde $p = m \cdot n$, podemos induzir uma norma matricial no espaço vetorial das matrizes complexas de dimensão $m \times n$, definindo

$$\|A\| = \max_{\|x\|=1} \|Ax\|,$$

para $A \in \mathbb{R}^{m \times n}$ e $x \in \mathbb{R}^n$.

Utilizando a Definição 2.4, definimos um tipo específico de norma matricial de a qual será utilizada ao longo deste trabalho. Lembrando que dado $A \in \mathbb{R}^{m \times n}$ a matriz transposta é denotada por A^T .

Definição 2.5. [28, Definição 5.2.7] Seja $A \in \mathbb{R}^{m \times n}$, a *2-norma* matricial de A é a norma matricial induzida pela norma euclidiana, dada por

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \sqrt{\lambda_{\max}},$$

onde λ_{\max} é o maior valor de λ tal que $A^T A - \lambda I$ é singular, isto é, o maior autovalor de $A^T A$.

Em seguida, apresentamos algumas propriedades da 2-norma.

Propriedade 2.6. Além das propriedades de uma norma induzida, a 2-norma dispõe das seguintes propriedades especiais,

- i) $\|A\|_2 = \max_{\|x\|_2=1} \max_{\|y\|_2=1} |y^T Ax|$;
- ii) $\|A\|_2 = \|A^T\|_2$;
- iii) $\|A^T A\|_2 = \|A\|_2^2$;
- iv) $\left\| \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \right\|_2 = \max\{\|A\|_2, \|B\|_2\}$;

v) $\|U^T A V\|_2 = \|A\|$, quando $U U^T = I$ e $V^T V = I$.

Demonstração: Ver [28, Propriedades 5.2.9 - 5.2.13]. □

2.1.3 Positividade de uma matriz simétrica

A positividade de uma matriz simétrica pode ser definida como se segue.

Definição 2.7. [20, p.495-496] Seja $A \in \mathbb{R}^{n \times n}$ uma matriz simétrica de ordem n . Dizemos que A é:

- i) *semidefinida positiva*, e denota-se por $A \geq 0$, se $x^T A x \geq 0$, para todo $x \in \mathbb{R}^n$;
- ii) *definida positiva*, e denota-se por $A > 0$, se $x^T A x > 0$, para todo $x \in \mathbb{R}^n$, $x \neq 0$.

2.1.4 Decomposição de Matrizes

Considere uma matriz quadrada $A \in \mathbb{R}^{n \times n}$. A primeira decomposição a ser abordada é a decomposição LU, em que podemos reescrever a matriz A como

$$A = LU,$$

onde $L \in \mathbb{R}^{n \times n}$ é uma matriz triangular inferior com diagonal unitária e $U \in \mathbb{R}^{n \times n}$ é triangular superior. Dizemos que a submatriz $k \times k$ consistindo nas primeiras k linhas e colunas de A é chamada *k-ésima submatriz principal* de A . O teorema enunciado a seguir evidencia as condições para a existência e unicidade da decomposição LU.

Teorema 2.8. (*Decomposição LU*) *Seja $A \in \mathbb{R}^{n \times n}$ uma matriz cujas submatrizes principais são todas não singulares. Existe uma matriz triangular inferior unitária $L \in \mathbb{R}^{n \times n}$ e uma matriz triangular superior $U \in \mathbb{R}^{n \times n}$ tais que $A = LU$, e esta decomposição é única.*

Demonstração: Ver [33, Teorema 1.7.19]. □

Considerando o sistema linear $Ax = b$, e utilizando a decomposição LU da matriz A , é possível encontrar a solução para esse sistema apenas com os sistemas triangulares dados

por

$$Ly = b,$$

$$Ux = y.$$

Quando não é possível realizar a decomposição LU de A , como no caso em que as matrizes não singulares possuem pivôs nulos quando aplicamos o processo de Eliminação Gaussiana, podemos utilizar outra estratégia, chamada pivoteamento parcial. Assim, existe uma matriz de permutação P tal que $PA = LU$, tal fatoração é chamada de decomposição LU com Pivotação Parcial.

Teorema 2.9. (*Decomposição LU com Pivotação Parcial*) *Seja $A \in \mathbb{R}^{n \times n}$ uma matriz não singular. Então existem matrizes de permutação $P_1, P_2 \in \mathbb{R}^{n \times n}$ onde $P_1A = LU$ ou $AP_2 = LU$.*

Demonstração: Ver [10, Teorema 2.5] □

Novamente, considerando o sistema linear $Ax = b$, e utilizando a decomposição definida acima, fazendo $\bar{b} = Pb$ conseguimos resolver tal sistema de maneira consideravelmente simples, através dos seguintes sistemas triangulares

$$Ly = \bar{b},$$

$$Ux = y.$$

Por fim apresentamos a Decomposição de Cholesky.

Teorema 2.10. (*Decomposição de Cholesky*) *Seja $A \in \mathbb{R}^{n \times n}$ uma matriz definida positiva. Então A pode ser decomposta de maneira única como o produto $A = R^T R$, onde $R \in \mathbb{R}^{n \times n}$ é uma matriz triangular superior, cujas entradas da diagonal principal são todas positivas. A matriz R é chamada de Fator de Cholesky de A .*

Demonstração: Ver [33, Teorema 1.4.7]. □

Fazemos o uso da Decomposição de Cholesky nas implementações de cada um dos métodos. Entretanto, tais decomposições não serão necessárias para a compreensão dos resultados teóricos apresentados neste trabalho.

2.2 Otimização

A Otimização estuda problemas de maximizar ou minimizar uma função de várias variáveis dentro de um conjunto factível. Além da matemática, métodos de otimização podem ser aplicados em diversas outras áreas como engenharia, administração, logística, transporte, economia, biologia ou de outras ciências, como afirmam Nocedal e Wright em [29] e Martínez e Santos em [26].

Considere o seguinte problema de otimização

$$\min_{x \in \Omega} f(x) \tag{2.2}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é a função objetivo e $\Omega \subseteq \mathbb{R}^n$ é o conjunto factível.

Definição 2.11. [26, p.07]

- a) Dizemos que x_* é um *minimizador global* de (2.2) se $f(x_*) \leq f(x)$ para todo $x \in \Omega$. Neste caso, $f(x_*)$ é chamado mínimo de f em Ω .
- b) Dizemos que x_* é um *minimizador local* de (2.2) se existe $\epsilon > 0$ tal que $f(x_*) \leq f(x)$ para todo $x \in \Omega$ tal que $\|x - x_*\| \leq \epsilon$.

Também costuma-se dizer que x_* é *mínimo local estrito* de (2.2) se existe $\epsilon > 0$ tal que $f(x_*) < f(x)$ para todo $x \in \Omega$ tal que $0 < \|x - x_*\| \leq \epsilon$.

O próximo teorema é um importante resultado de otimização, uma vez que garante a existência de um minimizador global para o problema (2.2), satisfeitas determinadas condições acerca da função objetivo e seu domínio.

Teorema 2.12. (Bolzano-Weierstrass) *Se $\Omega \subseteq \mathbb{R}^n$ é um subconjunto compacto, e $f : \Omega \rightarrow \mathbb{R}$ uma função contínua. Então existe $x_* \in \Omega$ minimizador global do problema (2.2).*

Demonstração: Ver [26, Teorema 1.3.1] □

Definiremos a seguir uma hipótese que aparece com frequência quando estamos trabalhando com métodos de otimização.

Definição 2.13. [29, p.624] Seja $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$. A função f é dita *Lipschitz continua* em $N \subset D$ se e somente se existe uma constante $L > 0$ tal que

$$\|f(x) - f(y)\| \leq L \|x - y\|, \text{ para todos } x, y \in N.$$

2.2.1 Condições de otimalidade para problemas de minimização sem restrições

Analizamos nesta subsecção o caso em que o conjunto factível Ω coincide com \mathbb{R}^n . Neste caso, o problema (2.2) é chamado de *Problema de Minimização Irrestrita* e podemos reescrevê-lo como

$$\min_{x \in \mathbb{R}^n} f(x). \quad (2.3)$$

Note que uma função $f : \mathbb{R} \rightarrow \mathbb{R}$ é de classe C^1 quando todas suas derivadas parciais de primeira ordem existem e são contínuas, e é de classe C^2 quando todas as derivadas de segunda ordem existem e são contínuas.

Lema 2.14. *Seja $f : \mathbb{R} \rightarrow \mathbb{R}$:*

(i) *Se $f \in C^1$ e x_* é um minimizador local de f em \mathbb{R} , então $f'(x_*) = 0$.*

(ii) *Se $f \in C^2$ e x_* é um minimizador local de f em \mathbb{R} , então $f'(x_*) = 0$ e $f''(x_*) \geq 0$.*

Demonstração: Ver respectivamente, [19, Seção 9.7, Teorema 1 e Teorema 2]. □

Os próximos dois teoremas determinam condições necessárias de primeira e segunda ordem, para que um ponto x_* seja um minimizador local de f em (2.2).

Teorema 2.15. (*Condição Necessária de 1ª Ordem*) *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função de classe C^1 . Se x_* é minimizador local de f , então $\nabla f(x_*) = 0$.*

Demonstração: Consideremos a função $\phi : \mathbb{R} \rightarrow \mathbb{R}$ definida por $\phi(t) = f(x_* + td)$, onde d é uma direção arbitrária em \mathbb{R}^n . Como por hipótese x_* é um minimizador local de f , temos que $t = 0$ é um minimizador local de ϕ , logo pelo Lema 2.14 (i), temos que $\phi'(0) = 0$. Derivando ϕ , obtemos $\phi'(t) = \nabla f(x_* + td)^T d$. Substituindo $t = 0$ em $\phi'(t)$, resulta em $\phi'(0) = \nabla f(x_*)^T d$. Como escolhemos d de forma arbitrária, concluímos que $\nabla f(x_*) = 0$. □

Teorema 2.16. (Condição Necessária de 2ª Ordem) Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função de classe C^2 . Se x_* é minimizador local de f , então

$$(i) \quad \nabla f(x_*) = 0;$$

$$(ii) \quad \nabla^2 f(x_*) \geq 0 \text{ (semidefinida positiva).}$$

Demonstração: A demonstração do item (i) segue diretamente do Teorema 2.15. Para mostrar o item (ii), definamos $\phi : \mathbb{R} \rightarrow \mathbb{R}$ por $\phi(t) = f(x_* + td)$, onde d é uma direção arbitrária em \mathbb{R}^n . Calculando a segunda derivada de ϕ , obtemos $\phi''(t) = d^T \nabla^2 f(x_* + td)d$. Logo o Lema 2.14 (ii) fornece que $\phi''(0) = d^T \nabla^2 f(x_*)d \geq 0$. Portanto $\nabla^2 f(x_*) \geq 0$, pois $d \in \mathbb{R}^n$ foi escolhido de forma arbitrária. \square

Apresentamos também uma condição suficiente de segunda ordem para que um ponto x_* seja de fato um minimizador da função objetivo f .

Teorema 2.17. (Condição Suficiente de 2ª Ordem) Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é função de classe C^2 . Se $x_* \in \mathbb{R}^n$ é tal que $\nabla f(x_*) = 0$ e $\nabla^2 f(x_*) > 0$, então x_* é minimizador local estrito de f .

Demonstração: Consideremos a função $\vartheta : H \rightarrow \mathbb{R}$ definida por $\vartheta(h) = h^T \nabla^2 f(x_*)h$, onde $H = \{h \in \mathbb{R}^n : \|h\|_2 = 1\}$. A função ϑ é contínua em H , uma vez que H é um conjunto compacto (fechado e limitado). Logo o Teorema 2.12 garante que ϑ atinge um valor mínimo em H .

Se $a \in \mathbb{R}$ é o valor mínimo de ϑ em H . A hipótese que $\nabla^2 f(x_*) > 0$, garante que $\vartheta(h) = h^T \nabla^2 f(x_*)h \geq a > 0$ para todo $h \in H$. Seja $d \in \mathbb{R}^n$ uma direção arbitrária não-nula, note que $\frac{d}{\|d\|_2} \in H$, então

$$d^T \nabla^2 f(x_*)d \geq a \|d\|_2^2. \quad (2.4)$$

Desenvolvendo f em Série de Taylor ao redor de x_* , obtemos

$$f(x_* + d) - f(x_*) = \nabla f(x_*)^T d + \frac{1}{2} d^T \nabla^2 f(x_*)^T d + o(\|d\|^2). \quad (2.5)$$

A hipótese $\nabla f(x_*) = 0$ em (2.5) implica que

$$\begin{aligned} f(x_* + d) - f(x_*) &= \frac{1}{2} d^T \nabla^2 f(x_*)d + o(\|d\|_2^2) \\ &\geq \frac{a}{2} \|d\|_2^2 + o(\|d\|_2^2). \end{aligned} \quad (2.6)$$

Logo, para todo $d \in \mathbb{R}^n$ de modo que $\|d\|_2$ é suficientemente pequeno, o termo $f(x_* + d)$ define qual será o sinal para o primeiro membro da desigualdade. No entanto temos que $\frac{a}{2} \|d\|_2^2 > 0$.

Portanto, para $\|d\|_2$ suficientemente pequeno e não-nulo, ou seja $0 < \|d\|_2 < \epsilon$,

$$f(x_* + d) - f(x_*) > 0,$$

o que implica que $f(x_* + d) > f(x_*)$. E concluímos que x_* é minimizador local estrito de f .

□

2.2.2 Minimização Convexa

Fazendo referência ao caso de minimização convexa, definiremos f como uma função convexa e enunciaremos dois resultados.

Definição 2.18. [11, Definição 1.1.] Uma função continuamente diferenciável $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é chamada *convexa* se para todo $x, y \in \mathbb{R}^n$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle .$$

Proposição 2.19. Se $\nabla^2 f(x) \geq 0$ para todo $x \in \Omega$, então f é convexa em Ω .

Demonstração: Ver [11, Lema 1.5.]

□

Proposição 2.20. Se $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é convexa e $\nabla f(x) = 0$, então x^* é minimizador global de f em \mathbb{R}^n .

Demonstração: Ver [11, Lema 1.2.]

□

2.2.3 O Método de Região de Confiança

Nesta seção, vamos descrever conceitos introdutórios acerca do método de região de confiança na resolução de problemas de minimização irrestrita. Uma vez que o Método de Levenberg-Marquardt pode ser visto como um método de região de confiança, tal técnica é

empregada para desenvolver os métodos descritos no Capítulo 4, assim como para demonstrar a convergência global dos métodos abordados no Capítulo 6.

Considere o problema,

$$\min_{x \in \mathbb{R}^n} f(x) \quad (2.7)$$

onde a função objetivo $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função continuamente diferenciável.

O método de região de confiança gera uma sequência $(x_k)_{k \in \mathbb{N}}$ com a ajuda do modelo quadrático da função objetivo que aproxima localmente o problema original, temos então a seguinte classe de problemas de minimização,

$$\begin{aligned} \min m_k(d) \\ \text{s.a. } \|d\| \leq \Delta_k, \end{aligned} \quad (2.8)$$

onde $\Delta_k > 0$ é chamado raio de confiança, e $m_k(d)$ é uma aproximação da função f por um modelo quadrático ao redor de x_k , o qual é definido por

$$m_k(d) = f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T B_k d, \quad (2.9)$$

onde $B_k \in \mathbb{R}^{n \times n}$ é uma matriz simétrica. A matriz B_k pode ser tanto a Hessiana $\nabla^2 f(x_k)$, quando uma matriz simétrica qualquer, contanto que satisfaça $\|B_k\|_2 \leq l$, onde $l > 0$ e independe de $k \in \mathbb{N}$. Ao encontrar a solução de tal modelo, é confrontado a qualidade da solução e, em função disso, o raio de confiança é modificado.

Para exemplificar, temos a seguinte imagem inspirada em [5, p.42] que representa método de região de confiança tomando $B_k = \nabla^2 f(x_k)$.

Como afirmam Nocedal e Wright [29, p.66], se o minimizador obtido na resolução de (2.8) for aceito na k -ésima iteração, o raio da região de confiança é aumentado ou permanece inalterado, porém caso o minimizador não seja aceito, a região de confiança é reduzida e calcula-se outro minimizador. O tamanho do raio da região de confiança é muito importante, pois se for muito grande, o minimizador do modelo quadrático pode estar muito longe do minimizador da função objetivo na região escolhida, levando a uma nova tentativa, reduzindo o tamanho da região e recalculando o modelo. Se o tamanho da região de confiança for muito pequeno, o algoritmo pode calcular passos muito pequenos e então serão necessárias muitas iterações para que a solução do problema seja encontrada.

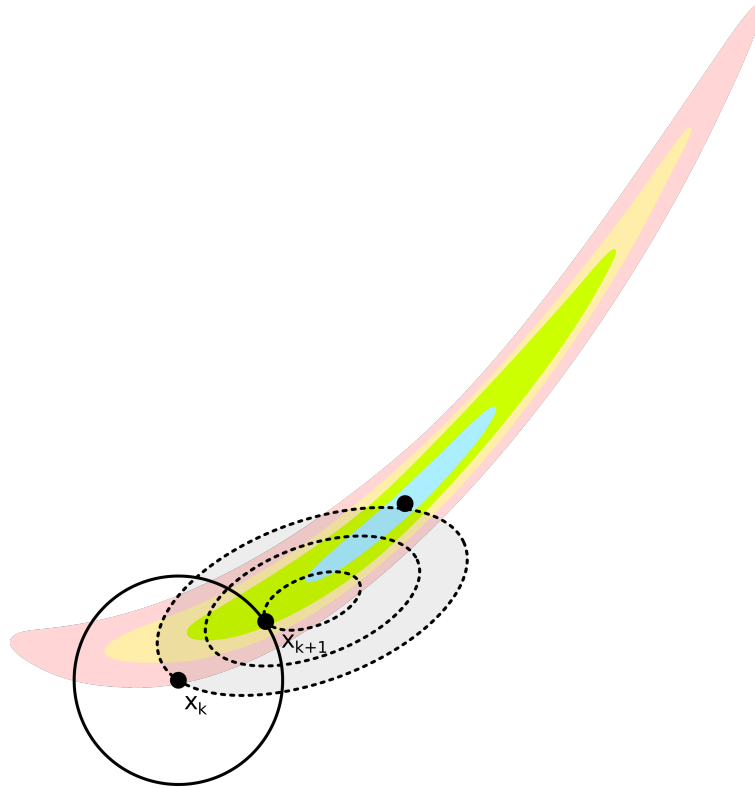


Figura 2.1: Exemplo de iteração do Método de Região de Confiança.

Desta forma, quando definimos o método de região de confiança, precisamos de uma estratégia para escolher um raio Δ_k em cada iteração. Para isso, consideremos a *Redução Real (Actual Reduction)* da função objetivo na iteração x_k

$$Ared_k = f(x_k) - f(x_k + d_k),$$

e a *Redução Predita (Predicted Reduction)* no modelo quadrático m_k

$$Pred_k = m_k(0) - m_k(d_k).$$

Assim, dada uma direção d_k , definimos a razão

$$\rho_k = \frac{f(x_k) - f(x_k + d_k)}{m_k(0) - m_k(d_k)}. \quad (2.10)$$

Observemos que, como obtemos a direção d_k minimizando o modelo m_k em uma região que inclui a direção $d = 0$, a redução prevista será sempre positiva, caso contrário (2.10) está mal definido. Isto se deve ao fato de que o gradiente em x_k é não nulo, logo existe uma direção de descida para algum d_k suficientemente pequeno.

Logo, quando ρ_k é negativo, temos que $f(x_k + d_k)$ é maior do que o valor $f(x_k)$, e então o passo deve ser rejeitado. Por outro lado, se ρ_k é próximo de 1, existe uma boa concordância entre o modelo m_k e a função f neste passo, é seguro expandir a região de confiança na próxima iteração. Se ρ_k é positivo, mas não próximo de 1, não alteramos a região de confiança, mas se ρ_k é próximo de zero ou negativo, diminuimos a região de confiança.

O próximo teorema e sua recíproca estabelecem a equivalência entre o problema de minimização e a solução de um sistema linear, o que facilita a obtenção da solução.

Teorema 2.21. *Sejam $B \in \mathbb{R}^{n \times n}$ uma matriz simétrica, $S \in \mathbb{R}^{n \times n}$, $g \in \mathbb{R}^n$ e $c \in \mathbb{R}$. Dado, $d_* \in \mathbb{R}^n$, com $\|Sd_*\| \leq \Delta$ e $\lambda_* \geq 0$ tais que as três condições são simultaneamente satisfeitas*

$$(B + \lambda_* S^T S)d_* = -g, \quad (\Delta - \|S \cdot d_*\|)\lambda_* \quad e \quad (B + \lambda_* S^T S) \geq 0,$$

então d_* é solução global do problema de região de confiança

$$\begin{aligned} \min m(d) &= c + g^T d + \frac{1}{2} d^T B d \\ \text{s.a } \|Sd\| &\leq \Delta \end{aligned}$$

Demonstração: Ver [5, Teorema 1.17]. □

Teorema 2.22. *Se d_* é solução global do problema*

$$\begin{aligned} \min m(d) &= c + g^T d + \frac{1}{2} d^T B d \\ \text{s.a } \|Sd\| &\leq \Delta \end{aligned}$$

então existe λ_* tal que

$$(B + \lambda_* S^T S)d_* = -g, \quad (\Delta - \|S \cdot d_*\|)\lambda_* \quad e \quad (B + \lambda_* S^T S) \geq 0,$$

Demonstração: Ver [5, Teorema 1.18]. □

Encontrar a solução do problema (2.8) nem sempre é uma tarefa simples, porém é possível encontrar uma solução aproximada e ainda sim obter a convergência do método.

Estratégias conhecidas usam o chamado *Ponto de Cauchy*. Para encontrar mais informações sobre esse assunto as referências [9] e [29] são recomendadas.

Em seguida, apresentamos um algoritmo para o Método de Região de Confiança, baseado em [29].

Algoritmo 1: MÉTODO DE REGIÃO DE CONFIANÇA

Entrada: x_0 , raio máximo $\bar{\Delta} > 0$, raio inicial $\Delta_0 \in (0, \bar{\Delta})$, $\eta \in [0, \frac{1}{4}]$ e $\varepsilon > 0$.

$k \leftarrow 0$;

Calcule $\nabla f(x_k)$ e B_k ;

repita

Encontre d_k solução de

$$\begin{aligned} \min_{d \in \mathbb{R}^n} m_k &= f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T B_k d \\ \text{s.a. } \|d\| &\leq \Delta_k \end{aligned}$$

Faça

$$\rho_k = \frac{f(x_k) - f(x_k + d_k)}{m_k(0) - m_k(d_k)};$$

Escolha Δ_{k+1} , definida por

$$\Delta_{k+1} = \begin{cases} \frac{1}{4}\Delta_k, & \text{se } \rho_k < \frac{1}{4}; \\ \Delta_k, & \text{se } \rho_k \in [\frac{1}{4}, \frac{3}{4}]; \\ \min(2\Delta_k, \hat{\Delta}), & \text{se } \rho_k > \frac{3}{4}. \end{cases}$$

Escolha x_{k+1} , tal que

$$x_{k+1} = \begin{cases} x_k + d_k & \text{se } \rho_k > \eta; \\ x_k & \text{caso contrário.} \end{cases}$$

Faça:

$$k \leftarrow k + 1.$$

até $\|\nabla f(x_k)\| \leq \varepsilon$

retorna x_k

PROBLEMA DE QUADRADOS MÍNIMOS

O problema de quadrados mínimos e uma proposta de solução para o mesmo foi inicialmente apresentado por Adrien-Marie Legendre (1752–1833) em 1805, em seu livro *Nouvelles méthodes pour la détermination des orbites des comètes*, como um modelo estatístico mais eficiente que os métodos conhecidos na época para combinar observações. Contudo, foi somente com o matemático alemão Carl Friedrich Gauss (1777–1855) que as teorias sobre técnicas de resolução ganharam consistência e popularizou-se na comunidade acadêmica [23, p.819].

Nas seções seguintes, nos dedicamos ao estudo introdutório do caso não linear para o problema de quadrados mínimos e técnicas para resolvê-lo, apresentando também o método de Gauss-Newton.

As principais referências deste capítulo são [5], [26] e [32].

3.1 A formulação do problema de quadrados mínimos

O problema de quadrados mínimos é um caso particular de um problema de otimização matemática, o qual procura determinar o melhor ajuste para um conjunto de dados tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados. Tais diferenças são funções suaves $r_i : \mathbb{R}^n \rightarrow \mathbb{R}$ com $1 \leq i \leq m$, denominadas de resíduos.

Os problemas de quadrados mínimos têm como função objetivo:

$$f(x) = \frac{1}{2} \sum_{i=1}^m r_i^2(x), \quad (3.1)$$

onde $m \geq n$.

Seja $R : \mathbb{R}^n \rightarrow \mathbb{R}^m$ um vetor residual, cujas entradas correspondem a cada resíduo r_i de (3.1). Assim

$$R(x) = \begin{bmatrix} r_1(x) \\ r_2(x) \\ \vdots \\ r_m(x) \end{bmatrix}.$$

Seguindo tal notação, podemos reescrever (3.1) como

$$f(x) = \frac{1}{2} \|R(x)\|_2^2. \quad (3.2)$$

Portanto, o problema de quadrados mínimos pode ser escrito como se segue

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|R(x)\|_2^2.$$

As derivadas parciais de primeira ordem dos resíduos podem ser expressas em termos da matriz Jacobiana de f , dada por

$$J(x) = \begin{bmatrix} \nabla r_1(x)^T \\ \nabla r_2(x)^T \\ \vdots \\ \nabla r_m(x)^T \end{bmatrix},$$

onde cada $\nabla r_i(x)$, $i = 1, 2, \dots, m$, é o vetor gradiente de r_i . Deste modo o gradiente e a Hessiana de f podem ser escritos como

$$\nabla f(x) = \sum_{i=1}^m r_i(x) \nabla r_i(x) = J(x)^T R(x), \quad (3.3)$$

$$\begin{aligned} \nabla^2 f(x) &= \sum_{i=1}^m \nabla r_i(x) \nabla r_i(x)^T + \sum_{i=1}^m r_i(x) \nabla^2 r_i(x), \\ &= J(x)^T J(x) + S(x), \end{aligned} \quad (3.4)$$

onde $S(x) = \sum_{i=1}^m r_i(x) \nabla^2 r_i(x)$.

Na maioria dos casos, obtemos facilmente o gradiente de (3.2), e portanto o primeiro termo de $\nabla^2 f(x)$, sem precisar calcular nenhuma segunda derivada de r_i . Segundo

Nocedal e Wright [29, p.246-247], essa é uma característica distinta dos problemas de quadrados mínimos, onde o termo $J(x)^T J(x)$ é mais importante que $S(x)$ em (3.4), seja porque os resíduos r_i se comportam como funções afim quando estão próximo da solução (ou seja, $\nabla^2 r_i$ são relativamente pequenos), ou porque os resíduos são pequenos.

3.2 O problema de quadrados mínimos não-linear

Quando trabalhamos com problemas reais, na maioria das vezes nos deparamos com modelos não lineares. Desta forma, estamos interessados em resolver o problema (3.2), quando

$$R(x) = 0, \quad (3.5)$$

é um sistema não-linear. Associado ao problema (3.5) podemos definir o seguinte problema de otimização

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|R(x)\|_2^2, \quad (3.6)$$

chamado *Problema de Quadrados Mínimos Não-Linear*. Quando o sistema (3.5) admite uma solução $x^* \in \mathbb{R}^n$, as soluções de (3.5) e (3.6) coincidem. Caso contrário, devemos buscar uma solução x^* que torne os resíduos $R(x)$ mínimos. Neste sentido, necessitamos de métodos iterativos capazes de resolver o problema (3.6). Dentre esses métodos, apresentaremos aqui o método de Gauss-Newton.

Primeiramente, relembremos que pela definição de f valem $\nabla f(x) = J(x)^T R(x)$ e $\nabla^2 f(x) = J(x)^T J(x) + S(x)$, onde $J(x)$ é a matriz jacobiana de f e $S(x) = \sum_{i=1}^m r_i(x) \nabla^2 r_i(x)$.

O Método de Gauss-Newton

Uma possível ideia de resolução do problema (3.6) seria o Método de Newton, porém neste caso haveria a necessidade de expressar a segunda derivada como dado de entrada. Para contornar este problema, podemos utilizar uma aproximação linear para $F(x)$, como segue

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

ao redor de x_k , onde $d = x - x_k$. Calculando a 2-norma, esperamos que $\|R(x)\|_2 \approx \|R(x_k) + J(x_k)d\|_2$, e assim podemos considerar o problema aproximado

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} \|R(x_k) + J(x_k)d\|_2^2. \quad (3.7)$$

Pela Condição de Minimização de Primeira Ordem, a direção de minimização deve satisfazer

$$\begin{aligned} \frac{\partial}{\partial d} \left(\frac{1}{2} \|R(x_k) + J(x_k)d\|_2^2 \right) &= 0, \\ \frac{\partial}{\partial d} \left(\frac{1}{2} (R(x_k)^T R(x_k) + 2d^T J(x_k)^T R(x_k) + d^T J(x_k)^T J(x_k)d) \right) &= 0, \\ J(x_k)^T R(x_k) + J(x_k)^T J(x_k)d &= 0, \\ J(x_k)^T J(x_k)d &= -J(x_k)^T R(x_k). \end{aligned} \quad (3.8)$$

A direção determinada por (3.8) é denominada de direção de Gauss-Newton, a qual denotaremos por d_k^{GN} . O teorema a seguir estabelece um resultado de convergência local para o Método de Gauss-Newton.

Teorema 3.1. *Seja $R : \mathbb{R}^n \rightarrow \mathbb{R}^m$ e seja $f(x) = \frac{1}{2}R(x)^T R(x)$ duas vezes diferenciável em um conjunto aberto e conexo D em \mathbb{R}^n . Suponhamos que $J(x)$ seja Lipschitziana em D com uma constante γ , que $\|J(x)\|_2 \leq \alpha$ para todo $x \in D$, e que existem $x_* \in D$ e $\lambda, \sigma \in \mathbb{R}$, $\sigma \geq 0$, tal que $J(x_*)^T R(x_*) = 0$, λ é o menor autovalor de $J(x_*)^T J(x_*)$, e*

$$\|(J(x) - J(x_k))^T R(x_*)\|_2 \leq \sigma \|x - x_*\|_2.$$

Para todo $x \in D$. Se $\sigma < \lambda$, então para todo $c \in (1, \frac{\lambda}{\sigma})$, existe $\epsilon > 0$ tal que para todo $x_0 \in B(x_, \epsilon) \equiv \{x \in \mathbb{R}^n : \|x - x_*\|_2 < \epsilon\}$, a sequência gerada pelo Método de*

Gauss-Newton está bem definida e converge para x_ , satisfazendo*

$$\|x_{k+1} - x_*\|_2 \leq \frac{c\sigma}{\lambda} \|x_k - x_*\|_2 + \frac{c\sigma\gamma}{2\lambda} \|x_k - x_*\|_2^2,$$

e

$$\|x_{k+1} - x_*\|_2 \leq \frac{c\sigma + \lambda}{2\lambda} \|x_k - x_*\|_2 < \|x_k - x_*\|_2^2.$$

Demonstração: Ver [12, Teorema 10.2.1]. □

Utilizando o Teorema 3.1, obtemos o resultado de convergência quadrática para o método, como enunciado no Teorema 3.2.

Teorema 3.2. *Sob as hipóteses do teorema anterior, se $R(x_*) = 0$, então existe $\epsilon > 0$ tal que para todo $x_0 \in B(x_*, \epsilon)$, a sequência gerada pelo Método de Gauss-Newton está bem definida e converge quadraticamente para x_* .*

Demonstração: Ver [12, Corolário 10.2.2]. □

Pelo Teorema 3.2, o método de Gauss-Newton apresenta convergência quadrática, no entanto tal método possui algumas desvantagens, por exemplo, não estando bem definido quando a matriz $J(x_k)^T J(x_k)$ é singular. Para contornar esse problema, a família Levenberg-Marquardt, soma a correção $\lambda_k I$ com $\lambda_k > 0$ na diagonal de $J(x_k)^T J(x_k)$, o que a torna não singular. Como podemos ver no próximo capítulo.

O MÉTODO DE LEVENBERG-MARQUARDT

O Método de Levenberg-Marquardt foi publicado inicialmente em 1944 pelo estatístico americano Kenneth Levenberg (1919–1971), sendo posteriormente redescoberto em 1963 pelo também estatístico Donald Marquardt (1929–1997), o qual aprimorou o método.

Neste capítulo apresentamos, na Seção 4.1, alguns conceitos teóricos a respeito do Método de Levenberg-Marquardt (LM). Nas Seções 4.2 e 4.3 descrevemos derivações do método LM propostas por Fan em [16] e [17], denominados Levenberg-Marquardt Modificado e o Método de Levenberg-Marquardt Modificado Acelerado. Na Seção 4.4, trabalharemos com o Método de Levenberg-Marquardt Modificado Acelerado Adaptado, proposto por Amini e Rostami em [2], pensado como uma forma de melhorar os métodos de Fan.

4.1 Levenberg-Marquardt

Como afirmado no capítulo anterior, estamos interessados em resolver o problema de quadrados mínimos não-linear

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|R(x)\|_2^2. \quad (4.1)$$

O Teorema 3.2 garante a convergência quadrática do método de Gauss-Newton, porém quando $J(x)^T J(x)$ não for definida positiva, tal método pode não estar definido. Para contornar este problema, o Método de Levenberg-Marquardt introduz um parâmetro $\lambda_k > 0$ na diagonal de $J(x)^T J(x)$, estabelecendo o seguinte sistema

$$\left(J(x)^T J(x) + \lambda_k I \right) d_k = -J(x)^T R(x), \quad (4.2)$$

onde λ_k é conhecido como parâmetro de *damping* ou parâmetro LM.

O Lema 4.1 nos mostra que a matriz $J(x)^T J(x) + \lambda_k I$ é sempre definida positiva, para todo $\lambda > 0$.

Lema 4.1. *Seja A uma matriz em $\mathbb{R}^{m \times n}$. Para todo $\lambda > 0$ a matriz $A^T A + \lambda I$ é definida positiva.*

Demonstração: Ver [2, Lema 3.1] □

Observe que, diferentemente do que acontece com o método Gauss-Newton, dado um $\lambda_k > 0$ ao adicionarmos $\lambda_k I$ na diagonal da matriz $J(x)^T J(x)$, obtemos a não singularidade exigida. Logo, a cada iteração o problema tem solução única e o método está bem definido.

Os próximos dois resultados garantem boas propriedades para a convergência do método de Levenberg-Marquardt.

Teorema 4.2. *Seja d a solução de (4.2) para um valor dado $\lambda > 0$. Então $\|d\|_2^2$ é uma função contínua decrescente na variável λ , tal que*

$$\lim_{\lambda \rightarrow \infty} \|d\|_2^2 = 0.$$

Demonstração: Ver [25, Teorema 2] □

Teorema 4.3. *Seja $\gamma(\lambda)$ o ângulo entre d (direção de Levenberg-Marquardt) e $-p$ (direção de máxima descida). Então $\gamma(\lambda)$ é uma função contínua monótona decrescente na variável λ tal que*

$$\lim_{\lambda \rightarrow \infty} \gamma(\lambda) = 0.$$

Além disso, d rotaciona na direção $-p$ quando $\lambda \rightarrow \infty$.

Demonstração: Ver [25, Teorema 3] □

Em seguida, apresentamos um algoritmo para o Método de Levenberg-Marquardt Clássico, baseado em [25, p. 438].

Algoritmo 2: LEVENBERG-MARQUARDT CLÁSSICO

Entrada: $R(x)$, $J(x)$, uma aproximação inicial $x_0 \in \mathbb{R}^n$, um valor inicial $\lambda_0 \in \mathbb{R}_+$, parâmetros de atualização $\kappa_1, \kappa_2 > 1$, uma tolerância $\varepsilon > 0$.

$k \leftarrow 0$;

repita

Obtenha a d_k solução de

$$(J(x_k)^T J(x_k) + \lambda_k I) d_k = -J(x_k)^T R(x_k);$$

Faça $x_n = x_k + d_k$;

se $R(x_n) < R(x_k)$ **então**

$$x_{k+1} = x_n;$$

$$\lambda_{k+1} = \frac{\lambda_k}{\kappa_1}.$$

senão

$$\lambda_{k+1} = \lambda_k \kappa_2.$$

fim se

Faça

$$k \leftarrow k + 1.$$

até $\|J(x_k)^T R(x_k)\| \leq \varepsilon$

retorna x_k

O método Levenberg-Marquardt tem a mesma convergência quadrática do Método de Gauss-Newton quando a jacobiana é Lipschitz contínua e não singular. Porém a condição de não singularidade é muito forte. As próximas seções apresentam variações do método Levenberg-Marquardt sob a condição de limite de erro local, que é mais fraca que a não singularidade.

4.2 Levenberg-Marquardt Modificado

Como visto anteriormente, para $R : \mathbb{R}^n \rightarrow \mathbb{R}^m$ uma função continuamente diferenciável, estamos interessados em resolver o sistema de equações não lineares

$$R(x) = 0, \quad (4.3)$$

Os métodos que trabalharemos nas próximas seções deste capítulo fazem uso de uma condição mais fraca que a não singularidade. Condição essa chamada limite de erro local definida como segue.

Definição 4.4. [34, Definição 1.1] Sejam $X^* \neq \emptyset$ o conjunto de soluções para (4.3) e $N \subset \mathbb{R}^n$ tal que $X^* \cap N \neq \emptyset$. Dizemos que $\|R(x)\|_2$ fornece um *limite de erro local* em N para o sistema não linear (4.3) se existe uma constante positiva c tal que

$$c \cdot \text{dist}(x, X^*) \leq \|R(x)\|_2 \quad \forall x \in N.$$

Yamashita e Fukushima mostraram em [34] que o método de Levenberg-Marquardt preserva convergência quadrática, quando o parâmetro *damping* escolhido for $\lambda_k = \|R(x_k)\|_2^2$. No mesmo ano, Fan e Yuan publicaram um artigo [13] mostrando que o método ainda preserva a mesma taxa de convergência quando o parâmetro *damping* é dado por $\lambda_k = \|R(x_k)\|_2^\delta$, com $\delta \in [1, 2]$, sob as mesmas condições.

Mais tarde, ainda sob a condição de limite de erro local, Fan propôs o Método de Levenberg-Marquardt Modificado (ver [16, p. 448]). O método desenvolvido por Fan, assim como o método Levenberg-Marquardt, resolve o seguinte sistema de equações lineares para obter o passo d_k ,

$$\begin{aligned} \left(J(x_k)^T J(x_k) + \lambda_k I \right) d_k &= -J(x_k)^T R(x_k) \\ \text{com } \lambda_k &= \mu_k \|R(x_k)\|_2^\delta, \quad \delta \in [1, 2] \text{ e } \mu_k > 0, \end{aligned} \quad (4.4)$$

E depois resolve outro sistema de equações lineares para obter o passo aproximado \hat{d}_k ,

$$\left(J(x_k)^T J(x_k) + \lambda_k I \right) d_k = -J(x_k)^T R(y_k) \quad \text{com } y_k = x_k + d_k. \quad (4.5)$$

Assim, a cada iteração o passo teste é definido como

$$s_k = d_k + \hat{d}_k. \quad (4.6)$$

Note que para diminuir o custo de calcular duas jacobianas $J(y_k)$ e $J(x_k)$ em cada iteração, é utilizado em (4.5) $\left(J(x_k)^T J(x_k) + \lambda_k I\right)$ ao invés de $\left(J(y_k)^T J(y_k) + \lambda_k I\right)$.

A técnica de região de confiança é utilizada para obter a convergência global deste método. Como já vimos, a redução predita é necessária para aplicarmos o método de região de confiança e deve ser sempre não negativa, porém no caso em que estamos trabalhando $\|R(x_k)\|_2^2 - \left\|R(x_k) + J(x_k)(d_k + \hat{d}_k)\right\|_2^2$ pode não possuir essa propriedade. Faz-se então necessário o cálculo de uma nova redução predita, tal como faz Fan em [16, p. 449-450].

Cálculo da nova redução Predita

Seja

$$\psi(x) = \|R(x)\|_2^2, \quad (4.7)$$

a função objetivo para resolver (4.3).

Definimos a redução real de $\psi(x)$ na k -ésima iteração como

$$Ared_k = \|R(x_k)\|_2^2 - \left\|R(x_k + d_k + \hat{d}_k)\right\|_2^2. \quad (4.8)$$

Para definir a nova redução predita, note que d_k é um minimizador do problema de minimização convexa

$$\min_{d \in \mathbb{R}^n} \|R(x_k) + J(x_k)d\|_2^2 + \lambda \|d\|_2^2.$$

Então, d_k é também uma solução do seguinte problema de região de confiança

$$\begin{aligned} & \min_{d \in \mathbb{R}^n} \|R(x_k) + J(x_k)d\|_2^2 \\ & s.a \|d\| \leq \Delta_{k,1} \end{aligned}$$

onde,

$$\Delta_{k,1} = \|d_k\|_2 = \left\| -\left(J(x_k)^T J(x_k) + \lambda I\right)^{-1} J(x_k)^T R(x_k) \right\|_2.$$

Pelo resultado de Powell [30] sabemos que

$$\|F(x_k)\|_2^2 - \|R(x_k) + J(x_k)d_k\|_2^2 \geq \|J(x_k)^T R(x_k)\|_2 \min \left\{ \|d_k\|_2, \frac{\|J(x_k)^T R(x_k)\|_2}{\|J(x_k)^T J(x_k)\|_2} \right\}. \quad (4.9)$$

A direção \hat{d}_k não é somente minimizador do problema

$$\min_{d \in \mathbb{R}^n} \|R(y_k) + J(x_k)d\|_2^2 + \lambda \|d\|_2^2,$$

mas também é solução do problema de região de confiança

$$\begin{aligned} & \min_{d \in \mathbb{R}^n} \|R(y_k) + J(x_k)d\|_2^2 \\ & \text{s.a. } \|d\|_2 \leq \Delta_{k,2} \end{aligned}$$

onde,

$$\Delta_{k,2} = \left\| \hat{d}_k \right\|_2 = \left\| - \left(J(x_k)^T J(x_k) + \lambda I \right)^{-1} J(x_k)^T R(y_k) \right\|_2.$$

Logo, temos novamente por [30] que

$$\|R(y_k)\|_2^2 - \left\| R(y_k) + J(x_k)\hat{d}_k \right\|_2^2 \geq \|J(x_k)^T R(y_k)\|_2 \min \left\{ \left\| \hat{d}_k \right\|_2, \frac{\|J(x_k)^T R(y_k)\|_2}{\|J(x_k)^T J(x_k)\|_2} \right\}. \quad (4.10)$$

Pelas desigualdades (4.9) e (4.10) é possível definir a nova redução predita como sendo

$$Pred_k = \|R(x_k)\|_2^2 - \|R(x_k) + J(x_k)d_k\|_2^2 + \|R(y_k)\|_2^2 - \left\| R(y_k) + J(x_k)\hat{d}_k \right\|_2^2, \quad (4.11)$$

satisfazendo

$$\begin{aligned} Pred_k \geq & \|J(x_k)^T R(x_k)\|_2 \min \left\{ \|d_k\|_2, \frac{\|J(x_k)^T R(x_k)\|_2}{\|J(x_k)^T J(x_k)\|_2} \right\} \\ & + \|J(x_k)^T R(y_k)\|_2 \min \left\{ \left\| \hat{d}_k \right\|_2, \frac{\|J(x_k)^T R(y_k)\|_2}{\|J(x_k)^T J(x_k)\|_2} \right\}. \end{aligned} \quad (4.12)$$

Apresentamos agora o algoritmo Levenberg-Marquardt Modificado, que foi adaptado de [18, p.450].

Algoritmo 3: LEVENBERG-MARQUARDT MODIFICADO

Entrada: $R(x)$, $J(x)$, uma aproximação inicial $x_0 \in \mathbb{R}^n$, $\mu_1 > M > 0$, $0 < q_1 \leq q_2 \leq q_3$,

$1 \leq \delta \leq 2$ e uma tolerância $\varepsilon > 0$.

$k \leftarrow 0$;

repita

Calcule $\lambda_k = \mu_k \|R(x_k)\|_2^\delta$.

Obtenha d_k resolvendo

$$(J(x_k)^T J(x_k) + \lambda_k I) d_k = -J(x_k)^T R(x_k);$$

Faça $y_k = x_k + d_k$.

Obtenha \hat{d}_k calculando

$$(J(x_k)^T J(x_k) + \lambda_k I) d_k = -J(x_k)^T R(y_k);$$

Faça

$$s_k = d_k + \hat{d}_k.$$

Calcule $\rho_k = \frac{Ared_k}{Pred_k}$, de acordo com (4.8) e (4.11).

Escolha x_{k+1} dado por

$$x_{k+1} = \begin{cases} x_k + s_k & \text{se } \rho_k \geq q_0; \\ x_k & \text{caso contrário.} \end{cases}$$

E escolha μ_{k+1} como

$$\mu_{k+1} = \begin{cases} 4\mu_k, & \text{se } \rho_k < q_1; \\ \mu_k, & \text{se } \rho_k \in [q_1, q_2]; \\ \max \left\{ \frac{\mu_k}{4}, M \right\}, & \text{se } \rho_k > q_2. \end{cases}$$

Faça $k \leftarrow k + 1$

até $\|J(x_k)^T R(x_k)\| \leq \varepsilon$

retorna x_k

O próximo Teorema 4.5 garante a convergência global do método Levenberg-Marquardt Modificado.

Teorema 4.5. *Seja $R(x)$ continuamente diferenciável, e ambas $R(x)$ e sua Jacobiana $J(x)$ são Lipschitz contínuas. Então o algoritmo termina em um número finito de iterações ou satisfaz*

$$\lim_{k \rightarrow \infty} \|J(x_k)R(x_k)\|_2 = 0.$$

Demonstração: Ver [18, Teorema 2.3]. □

4.3 Levenberg-Marquardt Modificado Acelerado

Na seção anterior foi apresentado o Método de Levenberg-Marquardt Modificado que calcula não apenas a direção d_k pelo sistema linear (4.4), mas também calcula a direção LM aproximada \hat{d}_k pelo sistema linear (4.5). Assim, o passo teste é definido por

$$s_k = d_k + \hat{d}_k.$$

Nesta seção, abordamos o método Levenberg-Marquardt Modificado Acelerado, que aplica uma busca linear para o passo LM aproximado, ou seja a direção teste é agora calculada por

$$s_k = d_k + \alpha_k \hat{d}_k. \tag{4.13}$$

Como d_k é uma direção de descida, é esperada que haja uma redução de $\psi(x)$ em y_k . Portanto Fan em [17] realiza uma busca linear em y_k ao longo de \hat{d}_k resolvendo o problema

$$\min_{\alpha > 0} \left\| R(y_k + \alpha \hat{d}_k) \right\|_2^2,$$

o qual pode ser aproximado por

$$\min_{\alpha > 0} \left\| R(y_k) + \alpha J(y_k) \hat{d}_k \right\|_2^2.$$

Como não calculamos $J(y_k)$ e sim $J(x_k)$ em (4.5), o problema pode novamente ser aproximado por

$$\min_{\alpha > 0} \left\| R(y_k) + \alpha J(x_k) \hat{d}_k \right\|_2^2. \tag{4.14}$$

O problema (4.14) é equivalente a

$$\max_{\alpha > 0} \|R(y_k)\|_2^2 - \left\| R(y_k) + \alpha J(x_k) \hat{d}_k \right\|_2^2,$$

onde a função objetivo pode ser reescrita como

$$\varphi(\alpha) = -\hat{d}_k^T J(x_k)^T J(x_k) \hat{d}_k \alpha^2 + 2\hat{d}_k^T (J(x_k)^T J(x_k) + \lambda I) \hat{d}_k \alpha, \quad (4.15)$$

que é uma função quadrática de α , a qual atinge seu máximo em

$$\tilde{\alpha}_k = \frac{\hat{d}_k^T (J(x_k)^T J(x_k) + \lambda_k I) \hat{d}_k}{\hat{d}_k^T J(x_k)^T J(x_k) \hat{d}_k} = 1 + \frac{\lambda_k \hat{d}_k^T \hat{d}_k}{\hat{d}_k^T J(x_k)^T J(x_k) \hat{d}_k} > 1. \quad (4.16)$$

Como $\tilde{\alpha}_k$ pode ser muito grande quando $J(x_k) \hat{d}_k \neq 0$, estabelecemos um limite superior para o tamanho do passo α_k e definimos

$$\alpha_k = \min\{\tilde{\alpha}_k, \hat{\alpha}\}, \quad (4.17)$$

onde $\hat{\alpha} > 1$ é uma constante positiva. As ideias expressas são formalizadas no Algoritmo 4.

Algoritmo 4: LEVENBERG-MARQUARDT MODIFICADO ACELERADO

Entrada: $R(x)$, $J(x)$, uma aproximação inicial $x_0 \in \mathbb{R}^n$, $\mu_1 > M > 0$, $0 < q_1 \leq q_2 \leq q_3$,
 $1 \leq \delta \leq 2$, $\hat{\alpha} > 1$ e uma tolerância $\varepsilon > 0$.

$k \leftarrow 0$;

repita

Calcule $\lambda_k = \mu_k \|R(x_k)\|_2^\delta$.

Obtenha d_k resolvendo

$$(J(x_k)^T J(x_k) + \lambda k I) d_k = -J(x_k)^T R(x_k);$$

Faça $y_k = x_k + d_k$.

Obtenha \hat{d}_k calculando

$$(J(x_k)^T J(x_k) + \lambda k I) d_k = -J(x_k)^T R(y_k);$$

Calcule α_k por (4.17) e faça

$$s_k = d_k + \alpha_k \hat{d}_k$$

Calcule $\rho_k = \frac{Ared_k}{Pred_k}$, como em (4.8) e (4.11).

Escolha x_{k+1} definido por

$$x_{k+1} = \begin{cases} x_k + s_k & \text{se } \rho_k \geq q_0; \\ x_k & \text{caso contrário.} \end{cases}$$

E escolha μ_{k+1} como

$$\mu_{k+1} = \begin{cases} 4\mu_k, & \text{se } \rho_k < q_1; \\ \mu_k, & \text{se } \rho_k \in [q_1, q_2]; \\ \max \left\{ \frac{\mu_k}{4}, M \right\}, & \text{se } \rho_k > q_2. \end{cases}$$

Faça $k \leftarrow k + 1$

até $\|J(x_k)^T R(x_k)\| \leq \varepsilon$

retorna x_k

O Teorema 4.6 garante a convergência global do Método de Levenberg-Marquardt Modificado Acelerado.

Teorema 4.6. *Seja $R(x)$ continuamente diferenciável, ambas $R(x)$ e sua Jacobiana $J(x)$ Lipschitz contínuas. Então o algoritmo termina em um número finito de iterações ou satisfaz*

$$\lim_{k \rightarrow \infty} \|J(x_k)^T R(x_k)\|_2 = 0.$$

Demonstração: Ver [17, Teorema 2.3]. □

4.4 Levenberg-Marquardt Modificado Acelerado Adaptado

Nos métodos propostos por Fan, o parâmetro *damping* é calculado por $\lambda_k = \mu_k \|R(x_k)\|_2^\delta$ com $\delta \in [1, 2]$. Como afirmam Amini e Rostami [2], quando a sequência está longe do conjunto de soluções e $\delta \geq 1$ temos que $\lambda_k = \mu_k \|R(x_k)\|_2^\delta$ pode ser muito grande, diminuindo o passo e impedindo que as iterações se movam rapidamente para o conjunto de soluções.

Para superar essa desvantagem foi introduzido no método Levenberg-Marquardt Modificado Acelerado uma adaptação para δ_k da seguinte maneira:

$$\delta_k = \begin{cases} \frac{1}{\|R(x_k)\|_2}, & \text{se } \|R(x_k)\|_2 \geq 1; \\ 1 + \frac{1}{k} & \text{caso contrário.} \end{cases} \quad (4.18)$$

No Algoritmo 5, expomos com detalhes a formalização das ideias apresentadas em [2,p.343] [2, p.343].

Algoritmo 5: LEVENBERG-MARQUARDT MODIFICADO ACELERADO ADAPTADO

Entrada: $R(x)$, $J(x)$, uma aproximação inicial $x_0 \in \mathbb{R}^n$, $\mu_1 > M > 0$, $0 < q_1 \leq q_2 \leq q_3$, $\hat{\alpha} > 1$ e uma tolerância $\varepsilon > 0$.

$k \leftarrow 0$;

repita

Calcule $\lambda_k = \mu_k \|R(x_k)\|_2^\delta$, onde δ é escolhido por (4.18).

Obtenha d_k resolvendo

$$(J(x_k)^T J(x_k) + \lambda k I) d_k = -J(x_k)^T R(x_k);$$

Faça $y_k = x_k + d_k$.

Obtenha \hat{d}_k calculando

$$(J(x_k)^T J(x_k) + \lambda k I) d_k = -J(x_k)^T R(y_k);$$

Calcule α_k por (4.17) e faça

$$s_k = d_k + \alpha_k \hat{d}_k$$

Calcule $\rho_k = \frac{Ared_k}{Pred_k}$, com $Ared_k$ e $Pred_k$ calculados respectivamente por (4.8) e (4.11).

Escolha x_{k+1}

$$x_{k+1} = \begin{cases} x_k + s_k & \text{se } \rho_k \geq q_0; \\ x_k & \text{caso contrário.} \end{cases}$$

E escolha μ_{k+1} como

$$\mu_{k+1} = \begin{cases} 4\mu_k, & \text{se } \rho_k < q_1; \\ \mu_k, & \text{se } \rho_k \in [q_1, q_2]; \\ \max\left\{\frac{\mu_k}{4}, M\right\}, & \text{se } \rho_k > q_2. \end{cases}$$

Faça $k \leftarrow k + 1$

até $\|J(x_k)^T R(x_k)\| \leq \varepsilon$

retorna x_k

Assim como nos métodos anteriores, expomos o resultado de convergência para o Algoritmo 5.

Teorema 4.7. *Seja $R(x)$ continuamente diferenciável, ambas $R(x)$ e sua Jacobiana $J(x)$ Lipschitz contínuas. Então o algoritmo termina em um número finito de iterações ou satisfaz*

$$\lim_{k \rightarrow \infty} \|J(x_k)^T R(x_k)\|_2 = 0.$$

Demonstração: Ver [2, Teorema 3.1].

□

REVISÃO LITERÁRIA DO PROBLEMA LOVO

O problema de Otimização de Menor Valor Ordenado (*Low Order-Value Optimization problem* - LOVO) foi inicialmente proposto por Andreani et al [4], em 2005. Um ano depois, Yano [35] defendeu sua tese de doutorado no tema. Essas as principais referências deste capítulo.

Na Seção 5.1, abordamos as definições iniciais dos problemas do tipo LOVO. Na Seção 5.2, são discutidas as condições de otimalidade do problema e é apresentado um algoritmo LOVO com convergência a pontos fracamente críticos. Por fim, na Seção 5.3, descrevemos o Método de Gauss-Newton para problemas LOVO e Método Levenberg-Marquardt com Busca Linear para problemas LOVO.

5.1 Otimização de Menor Valor Ordenado

Dadas r funções reais $F_i : \mathbb{R}^n \rightarrow \mathbb{R}$ com $1 \leq i \leq r$, e um inteiro $p \in \{1, 2, \dots, r\}$, o problema LOVO consiste em minimizar a soma das funções que tomam os p menores valores.

Para cada $x \in \mathbb{R}^n$, considere uma ordenação dos valores

$$F_{i_1(x)}(x) \leq F_{i_2(x)}(x) \leq \dots \leq F_{i_p(x)}(x) \leq \dots \leq F_{i_r(x)}(x). \quad (5.1)$$

Note que a ordem $\{i_1(x), \dots, i_r(x)\}$ depende de x e $\{i_1(x), \dots, i_r(x)\} \subset \{1, \dots, r\}$. Denota-

mos por $S_p : \mathbb{R}^n \rightarrow \mathbb{R}$ a função de Menor Valor Ordenado

$$S_p(x) = \sum_{j=1}^p F_{i_j(x)}(x), \quad \forall x \in \mathbb{R}^n.$$

Definimos então o problema LOVO da seguinte maneira,

$$\min_{x \in \mathbb{R}^n} S_p(x). \quad (5.2)$$

Como analisado em [35], se as funções F_i , $i = 1, \dots, r$, são contínuas então a função S_p também é contínua. Porém, mesmo que as funções F_i sejam diferenciáveis, S_p pode não ser diferenciável na maioria dos casos.

Há ainda outro modo de definir o problema LOVO. Seja $m = \frac{r!}{p!(r-p)!}$, considere a coleção $C = \{C_1, \dots, C_m\}$, de todos os subconjuntos formados por p elementos do conjunto $\{1, 2, \dots, r\}$. Assim, para todo $i = 1, \dots, m$ e $x \in \mathbb{R}^n$, podemos definir

$$f_i(x) = \sum_{j \in C_i} F_j(x).$$

E finalmente definir

$$f_{min}(x) = \min\{f_1(x), \dots, f_m(x)\}.$$

Portanto, $f_{min}(x) = S_p(x)$ para todo $x \in \mathbb{R}^n$. Assim, podemos também definir o problema LOVO como

$$\min_{x \in \mathbb{R}^n} f_{min}(x). \quad (5.3)$$

Observe que nesta nova formulação, a não diferenciabilidade da função fica mais evidente, o que, para fins teóricos, torna as demonstrações mais claras.

Uma generalização do problema de quadrados mínimos via problemas LOVO

Sejam y_1, y_2, \dots, y_r observações sobre os cenários t_1, t_2, \dots, t_r . Denotando $\Phi(x, t)$ como uma função de ajuste, definimos o erro quadrático obtido pela i -ésima observação por $F_i(x) = (\Phi(x, t_i) - y_i)^2$. Logo, o problema de quadrados mínimos pode ser representado por

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^r F_i(x). \quad (5.4)$$

Note que este problema é equivalente ao problema LOVO considerando $p = r$. Considerando que $r - p$ valores foram obtidos com erros de observação, desejamos eliminar sua influência do problema (5.4). Logo, podemos reescrever o problema (5.4) como

$$\min_{x \in \mathbb{R}^n} S_p(x),$$

ou como apresentado anteriormente,

$$\min_{x \in \mathbb{R}^n} f_{min}(x).$$

Dessa forma, o problema LOVO generaliza o problema de quadrados mínimos não-linear.

Para ilustrar nossa abordagem, consideremos o seguinte exemplo retirado de [8, p.2]. Dada a função de ajuste $\Phi(x, t) = t(\sin(x) + \cos(x))$, e o conjunto de dados

\mathbf{x}	\mathbf{y}
- 0.5	0.119447
0.0	0.3
0.5	0.203551
0.75	0.423998

Tabela 5.1: Dados para o exemplo

Como podemos ver temos $r = 4$, sejam as F_i dadas por

$$\begin{aligned} F_1(x) &= (\Phi(-0.5, t_1) - 0.119447)^2, \\ F_2(x) &= (\Phi(0.0, t_2) - 0.3)^2, \\ F_3(x) &= (\Phi(0.5, t_3) - 0.203551)^2, \\ F_4(x) &= (\Phi(0.75, t_4) - 0.423998)^2. \end{aligned} \tag{5.5}$$

Considerando $p = 3$, há 4 possíveis subconjuntos de $\{1, 2, 3, 4\}$ com 3 elemento, a saber,

$$C_1 = \{1, 2, 3\}, C_2 = \{1, 2, 4\}, C_3 = \{1, 3, 4\} \text{ e } C_4 = \{2, 3, 4\}. \tag{5.6}$$

Logo, associado a cada conjunto C_i , $i = 1, \dots, 4$, as funções seguem

$$\begin{aligned} f_1(x) &= F_1(x) + F_2(x) + F_3(x), \\ f_2(x) &= F_1(x) + F_2(x) + F_4(x), \\ f_3(x) &= F_1(x) + F_3(x) + F_4(x), \\ f_4(x) &= F_2(x) + F_3(x) + F_4(x). \end{aligned} \tag{5.7}$$

Consequentemente,

$$f_{\min}(x) = \min\{f_1(x), f_2(x), f_3(x), f_4(x)\} = S_3(x) \quad (5.8)$$

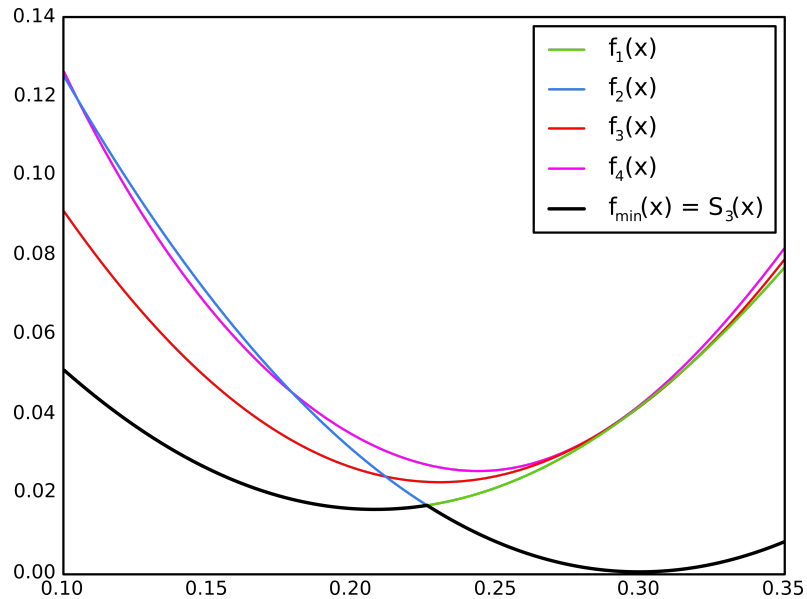


Figura 5.1: Gráfico da função LOVO.

Note que cada f_i é diferenciável, no entanto f_{\min} não é diferenciável em um ponto no intervalo $[0.20, 0.25]$.

5.2 Otimalidade do problema LOVO

Conforme [35] denotamos o conjunto de índices das funções f_i que minimizam a função f_{\min} em x , por

$$I_{\min}(x) = \{i \in \{1, \dots, m\} \mid f_i(x) = f_{\min}(x)\}. \quad (5.9)$$

O seguinte lema mostra que um minimizador global x_* de (5.3) é, necessariamente, um minimizador global de $f_i(x)$ para todo $i \in I_{\min}(x_*)$.

Lema 5.1. *Sejam $A \subset \mathbb{R}^n$ e $x_* \in A$. Se x_* é um minimizador global de $f_{\min}(x)$ sujeito a $x \in A$, então x_* é minimizador de $f_i(x)$ sujeito a $x \in A$ para todo $i \in I_{\min}(x_*)$. Em particular, tomando $A = \mathbb{R}^n$, se x_* é minimizador global de (5.3) então x_* é um minimizador global de $f_i(x)$ para todo $i \in I_{\min}(x_*)$.*

Demonstração: Ver [35, Lema 4.1]. □

A mesma propriedade vale para minimizadores locais como vemos no teorema seguir.

Teorema 5.2. *Seja $x_* \in \mathbb{R}^n$ um minimizador local de (5.3). Para todo $i \in I_{min}(x_*)$, x_* é minimizador local de $f_i(x)$ sujeito a $x \in \mathbb{R}^n$.*

Demonstração: Ver [35, Teorema 4.2] □

Enunciaremos agora uma importante definição no contexto da convergência para problemas LOVO.

Definição 5.3. [4, p.6] Seja $x_* \in \mathbb{R}^n$. Dizemos que:

- a) x_* é *fortemente crítico* se, para todo $i \in I_{min}(x_*)$, x_* é ponto estacionário de f_i .
- b) x_* é *fracamente crítico* se existe $i \in I_{min}(x_*)$ tal que x_* é ponto estacionário de f_i .

Nossa abordagem ao problema LOVO e ao subproblema de quadrados mínimos não linear, resume-se aos algoritmos com convergência para pontos fracamente críticos, devido a complexidade teórica relacionada a convergência para pontos fortemente críticos. De modo prático, a convergência para pontos fracamente críticos é suficiente para a solução de uma ampla gama de problemas.

Convergência para pontos fracamente críticos

Como afirmamos anteriormente, não há garantia de S_p ser diferenciável, ou seja, podem haver pontos nos quais $\nabla f_{min}(x)$ não exista. Para superar este problema, sob determinadas situações uma saída simples é escolher $i \in I_{min}(x)$ sempre que $\nabla f_{min}(x)$ é exigido pelo algoritmo e fazer

$$\nabla f_{min}(x) \leftarrow \nabla f_i(x).$$

No entanto, como afirmam Andreani et al [4, p. 06], para muitos problemas não diferenciáveis, aplicar essa estratégia pode ser desastrosa. Entretanto, como evidenciado em [35, p. 44], as consequências são menos expressivas no caso de problemas LOVO e,

necessariamente, temos a convergência para pontos fracamente críticos (lembramos que para o problema LOVO Irrestrito, um ponto fracamente crítico é um ponto tal que $\nabla f_i(x) = 0$, para algum $i \in I_{min}(x)$).

5.3 Métodos para resolver problemas LOVO no contexto de ajuste

Nesta seção, serão investigadas duas adaptações clássicas para resolver LOVO no contexto de ajuste. Tais algoritmos empregam a técnica de busca linear, em contraposição a teoria de região de confiança, à qual empregamos nos demais métodos vistos no Capítulo 6. Estes métodos serão utilizados nos testes numéricos do Capítulo 7.

LOVO Gauss-Newton

O Método de Gauss-Newton aqui descrito foi proposto por Andreani et al. [3, Algoritmo 1], para problemas de Otimização de Valor Ordenado (OVO), do qual o problema LOVO é um caso particular. Como feito em [32, p. 54], iremos empregar apenas no uso deste algoritmo para LOVO.

Considere a seguinte notação, dadas r funções reais $F_i : \mathbb{R}^n \rightarrow \mathbb{R}$, de classe $C^1(\mathbb{R}^n)$ e $p \in \{1, \dots, r\}$ um número inteiro, temos o problema LOVO

$$\min_{x \in \mathbb{R}^n} f_{min}(x), \quad (5.10)$$

com

$$f_{min}(x) = \min \left\{ \sum_{j \in C_i} F_j(x), i = 1, \dots, m \right\} \quad (5.11)$$

onde $m = \frac{r!}{p!(r-p)!}$ e cada C_i é um subconjunto distinto de cardinalidade p do conjunto $\{1, \dots, r\}$. A matriz $J_{C_i}(x) \in \mathbb{R}^{p \times n}$ é a Jacobiana relativa a função vetorial $F_{C_i} : \mathbb{R}^n \rightarrow \mathbb{R}^p$

dada por

$$F_{C_i}(x) = \begin{bmatrix} F_{j_1}(x) \\ F_{j_2}(x) \\ \vdots \\ F_{j_p}(x) \end{bmatrix}, \text{ onde } \{j_1, j_2, \dots, j_p\} \in C_i.$$

Fixado $x_k \in \mathbb{R}^n$ com $i \in I_{\min}(x_k)$, o gradiente de f_i pode ser escrito como $\nabla f_i(x_k) = J_{C_i}(x_k)^T F_{C_i}(x_k)$. Logo, a direção de Gauss-Newton é dada por $J_{C_i}(x_k)^T J_{C_i}(x_k) d_k = -J_{C_i}(x_k)^T F_{C_i}(x_k)$.

Note que os pontos limites das sequências geradas por este algoritmo são pontos fracamente críticos de $f_{\min}(x)$ [3, p.4]. Mais ainda, a convergência deste algoritmo é superlinear, como pode ser encontrado em [3, p.3]. O seguinte algoritmo foi proposto em [3, Algoritmo 1], note que tal algoritmo utiliza busca linear para a determinação do passo t .

Algoritmo 6: LOVO GAUSS-NEWTON

Entrada: $F(x)$, $J(x)$, $\gamma \in (0, 1)$, $\lambda > 0$, $\varepsilon > 0$ e uma aproximação inicial $x_0 \in \mathbb{R}^n$,

$k \leftarrow 0$;

repita

Escolha $i \in I_{\min}(x_k)$;

Caso a matriz $J_{C_i}(x_k)^T J_{C_i}(x_k)$ seja definida positiva, obtenha d_k solução de:

$$J_{C_i}(x_k)^T J_{C_i}(x_k) d_k = -J_{C_i}(x_k)^T F_{C_i}(x_k)$$

Caso contrário, obtenha d_k resolvendo:

$$(J_{C_i}(x_k)^T J_{C_i}(x_k) + \lambda I) d_k = -J_{C_i}(x_k)^T F_{C_i}(x_k)$$

Faça $t = 1$;

enquanto $f_{\min}(x_k + t d_k) > f_{\min}(x_k) + \gamma t (J_{C_i}(x_k)^T F_{C_i}(x_k))^T d_k$ **faça**

Escolha novo $t \in [0.1t, 0.9t]$.

fim enquanto

Faça $x_{k+1} \leftarrow x_k + t d_k$ e $k \leftarrow k + 1$

até $\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \leq \varepsilon$

retorna x_k

LOVO Levenberg-Marquardt

O Método Levenberg-Marquardt com Busca Linear para problemas LOVO irredutíveis e com convergência para pontos fracamente críticos foi proposto como resultado da dissertação de mestrado por Schwertner em [32, p. 58]. O Algoritmo 7, resume as ideias utilizadas em [32]. Note ainda, que o termo busca linear, refere-se a determinação do passo t .

Algoritmo 7: LOVO LEVENBERG-MARQUARDT

Entrada: $F(x)$, $J(x)$, uma aproximação inicial $x_0 \in \mathbb{R}^n$, uma constante $\gamma \in (0, 1)$ e uma tolerância $\varepsilon > 0$.

$k \leftarrow 0$;

repita

Escolha $i \in I_{min}(x_k)$;

Calcule λ_k ;

Obtenha d_k solução de:

$$(J_{C_i}(x_k)^T J_{C_i}(x_k) + \lambda_k I) d_k = -J_{C_i}(x_k)^T F_{C_i}(x_k); \quad (5.12)$$

Faça $t = 1$;

enquanto $f_i(x_k + td_k) > f_i(x_k) + \gamma t \nabla f_i(x_k)^T d_k$ **faça**

Escolha novo $t \in [0.1t, 0.9t]$.

fim enquanto

Faça:

$$x_{k+1} \leftarrow x_k + td_k;$$

$$k \leftarrow k + 1;$$

até $\|\nabla f_i(x_k)\|_2 \leq \varepsilon$

retorna x_k

Neste trabalho, foram apresentados e testados diversos parâmetros de *damping* distintos, sendo $\lambda_k = \frac{2\sqrt{f(x_k)}}{3k}$ o parâmetro que obteve melhor desempenho.

LEVENBERG-MARQUARDT COM PASSOS MÚLTIPLOS PARA LOVO

Destinamos este capítulo para as principais contribuições do presente trabalho. Em resumo, apresentamos uma variação dos métodos Levenberg-Marquardt Modificado, Levenberg-Marquardt Modificado Acelerado e Levenberg-Marquardt Modificado Acelerado Adaptado que abordados no Capítulo 4, no contexto dos problemas do tipo LOVO. Assim como resultados que garantem a convergência desses métodos.

6.1 Conceitos iniciais

Como vimos no Capítulo 5, consideremos r funções $F_i : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe $C^1(\mathbb{R}^n)$, e um inteiro $p \in \{1, 2, \dots, r\}$. No contexto de funções de ajuste, podemos interpretar $F_i(x)$ como o desvio do modelo em relação ao dados observados. Definimos então

$$F_i(x) = \Phi(x, t_i) - y_i,$$

onde $\Phi(x, t_i)$ é um modelo escolhido previamente.

Num contexto mais geral, estamos interessados em resolver o problema de encontrar $x \in \mathbb{R}^n$ e um subconjunto \mathcal{S} de \mathbb{N} com p elementos tais que

$$F_i(x) = 0, \text{ para todo } i \in \mathcal{S}. \tag{6.1}$$

Usaremos a minimização de uma função do tipo LOVO conveniente para resolver o problema proposto. Para isso, vamos definir

$$R_i(x) = \frac{1}{2}(F_i(x))^2.$$

Dado um $x \in \mathbb{R}^n$, ordenamos os valores de R_i da seguinte forma,

$$R_{i_1(x)}(x) \leq R_{i_2(x)}(x) \leq \cdots \leq R_{i_r(x)}(x),$$

tomando as p menores funções, geramos a função LOVO,

$$S_p(x) = \sum_{k=1}^p R_{i_k(x)}(x). \quad (6.2)$$

Ao encontrar o minimizador global de

$$\min_{x \in \mathbb{R}^n} S_p(x), \quad (6.3)$$

estaremos resolvendo o problema (6.1), desde que exista uma solução.

Como afirmado anteriormente, trabalhar com a ordenação pode tornar a notação complicada, então podemos definir esse problema utilizando a notação combinatorial. Neste sentido, definamos $C = \{C_1, C_2, \dots, C_m\}$ como sendo o subconjunto de todas as combinações de p elementos tomados no conjunto $\{1, 2, \dots, r\}$. Assim, a função (6.2) e o problema (6.3) podem ser reescrito da seguinte forma,

$$f_i(x) = \sum_{j \in C_i} R_j(x), i \in \{1, \dots, n\},$$

e

$$f_{min}(x) = \min\{f_1(x), \dots, f_n(x)\},$$

e, portanto,

$$f_{min}(x) = S_p(x).$$

Continuaremos a fixar notações. De modo análogo ao realizado em (5.9), definimos o conjunto de índices das funções f_i que minimizam a função f_{min} em x , como

$$I_{min}(x) = \{i \in \{1, \dots, n\} : f_{min}(x) = f_i(x)\}. \quad (6.4)$$

Fixando $x_k \in \mathbb{R}^n$ e $i \in I_{min}(x_k)$, como em (3.3), o gradiente de f_i pode ser escrita como $\nabla f_i(x) = \sum_{j \in C_i} \nabla R_j(x) = J_{C_i}(x)^T F_{C_i}(x)$.

6.2 Método de Levenberg-Marquardt Modificado para problemas LOVO

O algoritmo desta seção é uma de nossas contribuições para este trabalho e corresponde a uma adaptação do método de Levenberg-Marquardt Modificado apresentado no Capítulo 4, no contexto do problema LOVO irrestrito.

Antes de apresentarmos o algoritmo, definamos a *Redução Real (Actual Reduction)* da função objetivo para este problema por

$$Ared_k = \|F_{C_i}(x_k)\|_2^2 - \left\| F_{C_i}(x_k + d_k + \hat{d}_k) \right\|_2^2. \quad (6.5)$$

Assim, a nova *Redução Predita (Predicted Reduction)* pode ser definida como:

$$Pred_k = \|F_{C_i}(x_k)\|_2^2 - \|F_{C_i}(x_k) + J_{C_i}(x_k)d_k\|_2^2 + \|F_{C_i}(y_k)\|_2^2 - \left\| F_{C_i}(y_k) + J_{C_i}(x_k)\hat{d}_k \right\|_2^2, \quad (6.6)$$

a qual, como visto em (4.12), satisfaz

$$\begin{aligned} Pred_k \geq & \|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \min \left\{ \|d_k\|_2, \frac{\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2}{\|J_{C_i}(x_k)^T J_{C_i}(x_k)\|_2} \right\} \\ & + \|J_{C_i}(x_k)^T F_{C_i}(y_k)\|_2 \min \left\{ \|\hat{d}_k\|_2, \frac{\|J_{C_i}(x_k)^T F_{C_i}(y_k)\|_2}{\|J_{C_i}(x_k)^T J_{C_i}(x_k)\|_2} \right\}. \end{aligned} \quad (6.7)$$

O Algoritmo 8 é uma adaptação para resolução de problemas do tipo LOVO do Algoritmo de Levenberg-Marquardt Modificado encontrado em [16].

Algoritmo 8: LOVO LEVENBERG-MARQUARDT MODIFICADO

Entrada: $F(x)$, $J(x)$, uma aproximação inicial $x_0 \in \mathbb{R}^n$, $\mu_1 > M > 0$, $0 < q_1 \leq q_2 \leq q_3$,

$1 \leq \delta \leq 2$ e uma tolerância $\varepsilon > 0$.

$k \leftarrow 0$;

repita

Escolha $i \in I_{min}(x_k)$.

Calcule $\lambda_k = \mu_k \|F_{C_i}(x_k)\|_2^\delta$.

Obtenha d_k resolvendo:

$$(J_{C_i}(x_k)^T J_{C_i}(x_k) + \lambda_k I) d_k = -J_{C_i}(x_k)^T F_{C_i}(x_k); \quad (6.8)$$

Faça $y_k = x_k + d_k$;

Obtenha \hat{d}_k resolvendo

$$(J_{C_i}(x_k)^T J_{C_i}(x_k) + \lambda_k I) \hat{d}_k = -J_{C_i}(x_k)^T F_{C_i}(y_k); \quad (6.9)$$

Faça

$$s_k = d_k + \hat{d}_k.$$

Calcule $\rho_k = \frac{Ared_k}{Pred_k}$, de acordo com (6.5) e (6.6).

Escolha x_{k+1} , onde

$$x_{k+1} = \begin{cases} x_k + s_k & \text{se } \rho_k \geq q_1; \\ x_k & \text{caso contrário.} \end{cases} \quad (6.10)$$

E escolha μ_{k+1} , como

$$\mu_{k+1} = \begin{cases} 4\mu_k, & \text{se } \rho_k < q_2; \\ \mu_k, & \text{se } \rho_k \in [q_2, q_3]; \\ \max\left\{\frac{\mu_k}{4}, M\right\}, & \text{se } \rho_k > q_3. \end{cases} \quad (6.11)$$

Faça $k \leftarrow k + 1$

até $\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \leq \varepsilon$

retorna x_k

Os próximos dois lemas são resultados clássicos, vistos regularmente nos cursos de Análise no \mathbb{R}^n , que utilizaremos na demonstração do resultado de convergência global do método.

Lema 6.1. *Dado U um subconjunto aberto de \mathbb{R}^n , seja $F : U \rightarrow \mathbb{R}^n$ é diferenciável em cada ponto do segmento de reta $(a, a + v)$ e tal que sua restrição ao segmento $[a, a + v] \subset U$ seja contínua. Se $\|J(x)\|_2 \leq P$ para todo $x \in (a, a + v)$ então $\|F(a + v) - F(a)\|_2 \leq P \cdot \|v\|_2$.*

Demonstração: Ver [24, Desigualdade do Valor Médio] □

Lema 6.2. *Seja $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ uma função continuamente diferenciável em um conjunto aberto e convexo $D \in \mathbb{R}^n$. Para quaisquer $x, x + d \in D$,*

$$F(x + d) - F(x) = \int_0^1 J(x + td)d dt.$$

Demonstração: Ver [11, Lema 4.1.2] □

Para o estudo da convergência global do Algoritmo 8, é necessária a seguinte hipótese.

Hipótese 6.3. *Seja $(x_k)_{k \in \mathbb{N}}$ uma sequência gerada pelo Algoritmo 8, e seja $\mathcal{K} = \{k_0, k_1, \dots\}$ é o conjunto de índices tais que $i \in I_{\min}(x_k) \forall k \in \mathcal{K}$, tal que $\lim_{k \in \mathcal{K}} x_k = x_*$. Suponha que $F_{C_i}(x)$ é continuamente diferenciável e ambos $F_{C_i}(x)$ e sua Jacobiana $J_{C_i}(x)$ são Lipschitz contínuas, isto é, existem constantes L_1 e L_2 tais que*

$$\|J_{C_i}(y) - J_{C_i}(x)\| \leq L_1 \|y - x\|, \quad \forall x, y \in \mathbb{R}^n \text{ e} \quad (6.12)$$

$$\|F_{C_i}(y) - F_{C_i}(x)\| \leq L_2 \|y - x\|, \quad \forall x, y \in \mathbb{R}^n. \quad (6.13)$$

Temos ainda que,

$$\|F_{C_i}(y) - F_{C_i}(x) - J_{C_i}(y)(y - x)\| \leq L_1 \|y - x\|^2, \quad \forall x, y \in \mathbb{R}^n. \quad (6.14)$$

Teorema 6.4. *Sob as condições da Hipótese (6.3), temos que x_* é fracamente crítico, $i \in I_{\min}(x_*)$ e*

$$\lim_{k \in \mathcal{K}} \|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 = 0 \quad (6.15)$$

Demonstração: Seja $(x_k)_{k \in \mathbb{N}}$ uma sequência gerada pelo Algoritmo 8, defina o conjunto de índices $\mathcal{K} = \{k_0, k_1, \dots\}$ tais que $i = v(k)$, $\forall k \in \mathcal{K}$ e $\lim_{k \in \mathcal{K}} x_k = x_*$. Observe que o índice i e o conjunto \mathcal{K} existem uma vez que o conjunto de índices $\{1, 2, \dots, m\}$ é finito.

Primeiramente mostraremos que $i \in I_{\min}(x_*)$. Assim como afirmam Andreani et al [3, Teorema 3.2] e Schwertner [32, Teorema 6.1] considere,

$$f_i(x) = \sum_{j \in C_i} R_j(x) = \frac{1}{2} \sum_{j \in C_i} (F_j(x))^2,$$

pela hipótese (6.3), as funções F_j são continuamente diferenciáveis, logo $f_i \in C^1$. A continuidade de $f_i(x)$ garante que $\lim_{k \in \mathcal{K}} f_i(x_k) = f_i(x_*)$. Pela definição do conjunto $I_{\min}(x)$, $i \in I_{\min}(x_k)$, para todo $k \in \mathcal{K}$, logo, $f_i(x_k) \leq f_s(x_k)$, para todo $s \in \{1, 2, \dots, m\}$, aplicando os limites em ambos os lados da desigualdade anterior, concluímos que $f_i(x_*) \leq f_s(x_*)$, para todo $s \in \{1, 2, \dots, m\}$. Portanto, $i \in I_{\min}(x_*)$.

Mostraremos agora que

$$\lim_{k \in \mathcal{K}} \|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 = 0.$$

De fato, suponhamos por contradição que existe $\tau > 0$ e infinitos $k \in \mathcal{K}$ tais que

$$\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \geq \tau. \quad (6.16)$$

Sejam K_1 e K_2 subconjuntos de \mathcal{K} definidos a seguir

$$K_1 = \{k \in \mathcal{K} \mid \|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \geq \tau\},$$

$$K_2 = \{k \in \mathcal{K} \mid \|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \geq \frac{\tau}{2} \text{ e } x_{k+1} \neq x_k\}.$$

Note que K_1 é um subconjunto infinito de \mathcal{K} , por construção. Separaremos agora a demonstração em dois casos distintos para K_2 .

Caso I: K_2 é finito.

Pela definição de K_2 o subconjunto

$$K_3 = \{k \in \mathcal{K} \mid \|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \geq \tau \text{ e } x_{k+1} \neq x_k\},$$

é também finito.

Se \bar{k} é o maior índice de K_3 , temos então que $x_{k+1} = x_k$ vale para todo $k \in \{k > \bar{k} \mid k \in K_1\}$. Considere agora, K_4 o conjunto de de índices de K_1 que não pertencem a K_3 ,

$$K_4 = \{k > \bar{k} \mid \|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \geq \tau \text{ e } x_{k+1} = x_k\}.$$

Note que se $k + 1 \in K_4$, temos que $\|J_{C_i}(x_{k+1})^T F_{C_i}(x_{k+1})\|_2 \geq \tau$ e $x_{k+2} = x_{k+1}$. Por outro lado, caso $x_{k+2} \neq x_{k+1}$ teríamos que $k + 1 \in K_3$, o que contradiz o fato de que \bar{k} é o maior índice de K_3 , portanto $k + 1 \in K_4$. Por indução, temos que $\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \geq \tau$ e $x_{k+1} = x_k$ são válidos para todos os $k > \bar{k}$.

Como visto no parágrafo anterior, $x_{k+1} = x_k$, para todo $k > \bar{k}$, logo segue de (6.10), que $\rho_k < q_1$ para todo $k > \bar{k}$. Em particular $\rho_k < q_2$, pois $q_1 < q_2$. Logo, por (6.11) e pela definição de λ_k ,

$$\mu_k \rightarrow +\infty \text{ e } \lambda_k \rightarrow +\infty. \quad (6.17)$$

Lembremos que d_k é definido por

$$(J_{C_i}(x_k)^T J_{C_i}(x_k) + \lambda_k I) d_k = -J_{C_i}(x_k)^T F_{C_i}(x_k) \quad (6.18)$$

Como, neste caso, $x_{k+1} = x_k$, para todo $k > \bar{k}$ e $\lambda_k \rightarrow +\infty$, segue da expressão (6.18) que

$$d_k \rightarrow 0. \quad (6.19)$$

Para simplificar os próximos cálculos utilizaremos a seguinte notação

$$B_{C_i}(x_k) = \left(J_{C_i}(x_k)^T J_{C_i}(x_k) + \lambda_k I \right)^{-1}$$

Segue de (6.14), (6.17) e da definição de \hat{d}_k que,

$$\begin{aligned}
\|\hat{d}_k\|_2 &= \|B_{C_i}(x_k)J_{C_i}(x_k)^T F_{C_i}(y_k)\|_2 \\
&\leq \|B_{C_i}(x_k)J_{C_i}(x_k)^T (F_{C_i}(y_k) - F_{C_i}(x_k) - J_{C_i}(x_k)d_k)\|_2 \\
&\quad + \|B_{C_i}(x_k)J_{C_i}(x_k)^T (F_{C_i}(x_k) + J_{C_i}(x_k)d_k)\|_2 \\
&\leq \|B_{C_i}(x_k)J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \\
&\quad + \|B_{C_i}(x_k)J_{C_i}(x_k)^T J_{C_i}(x_k)d_k\|_2 \\
&\quad + \|B_{C_i}(x_k)J_{C_i}(x_k)^T (F_{C_i}(y_k) - F_{C_i}(x_k) - J_{C_i}(x_k)d_k)\|_2 \\
&\leq \|d_k\|_2 + \|B_{C_i}(x_k)J_{C_i}(x_k)^T J_{C_i}(x_k)\|_2 \|d_k\|_2 \\
&\quad + \|B_{C_i}(x_k)J_{C_i}(x_k)^T\|_2 \|F_{C_i}(y_k) - F_{C_i}(x_k) - J_{C_i}(x_k)d_k\|_2 \\
&\leq \|d_k\|_2 + \|B_{C_i}(x_k)J_{C_i}(x_k)^T J_{C_i}(x_k)\|_2 \|d_k\|_2 \\
&\quad + L_1 \|d_k\|_2^2 \|B_{C_i}(x_k)J_{C_i}(x_k)^T\|_2.
\end{aligned} \tag{6.20}$$

Note que por (6.17),

$$\|B_{C_i}(x_k)J_{C_i}(x_k)^T J_{C_i}(x_k)\|_2 \rightarrow 0, \tag{6.21}$$

então para k suficientemente grande temos que (6.21) é menor ou igual a 1.

Do mesmo modo obtemos que

$$\|B_{C_i}(x_k)J_{C_i}(x_k)^T\|_2 \leq 1. \tag{6.22}$$

Logo, por (6.20), (6.21) e (6.22),

$$\begin{aligned}
\|\hat{d}_k\|_2 &\leq \|d_k\|_2 + \|d_k\|_2 + L_1 \|d_k\|_2^2 = \|d_k\|_2 (2 + L_1 \|d_k\|_2) \\
&\leq \bar{c} \|d_k\|_2,
\end{aligned} \tag{6.23}$$

vale para todo k suficientemente grande, onde $\bar{c} \in \mathbb{R}^+$ é uma constante. Portanto,

$$\|s_k\|_2 = \|d_k + \hat{d}_k\|_2 \leq (1 + \bar{c}) \|d_k\|_2. \tag{6.24}$$

Além disso, de (6.5), (6.6), (6.7), temos

$$\begin{aligned}
|\rho_k - 1| &= \left| \frac{Ared_k - Pred_k}{Pred_k} \right| \\
&\leq \left| \frac{\|F_{C_i}(x_k + s_k)\|_2^2 - \|F_{C_i}(y_k) + J_{C_i}(x_k)\hat{d}_k\|_2^2}{Pred_k} \right| \\
&\quad + \left| \frac{\|F_{C_i}(y_k)\|_2^2 - \|F_{C_i}(x_k) + J_{C_i}(x_k)d_k\|_2^2}{Pred_k} \right|.
\end{aligned} \tag{6.25}$$

Note que em (6.25)

$$\begin{aligned}
&\|F_{C_i}(x_k + s_k)\|_2^2 - \|F_{C_i}(y_k) + J_{C_i}(x_k)\hat{d}_k\|_2^2 = \\
&\underbrace{\left(\|F_{C_i}(x_k + s_k)\|_2 - \|F_{C_i}(y_k) + J_{C_i}(x_k)\hat{d}_k\|_2 \right)}_A \underbrace{\left(\|F_{C_i}(x_k + s_k)\|_2 + \|F_{C_i}(y_k) + J_{C_i}(x_k)\hat{d}_k\|_2 \right)}_B.
\end{aligned}$$

Agora, pelo Lema 6.2 e por (6.12),

$$\begin{aligned}
A &\leq \left\| F_{C_i}(x_k + s_k) - F_{C_i}(y_k) - J_{C_i}(x_k)\hat{d}_k \right\|_2 \\
&= \left\| F_{C_i}(x_k + s_k) - F_{C_i}(x_k) - \int_0^1 J_{C_i}(x_k + td_k)d_k dt - J_{C_i}(x_k)\hat{d}_k \right\|_2 \\
&= \left\| F_{C_i}(x_k + s_k) - F_{C_i}(x_k) - J_{C_i}(x_k)\hat{d}_k - J_{C_i}(x_k)d_k - \int_0^1 \left(J_{C_i}(x_k + td_k) - J_{C_i}(x_k) \right) d_k dt \right\|_2 \\
&= \left\| F_{C_i}(x_k + s_k) - F_{C_i}(x_k) - J_{C_i}(x_k)s_k - \int_0^1 \left(J_{C_i}(x_k + td_k) - J_{C_i}(x_k) \right) d_k dt \right\|_2 \\
&= \left\| \int_0^1 \left(J_{C_i}(x_k + ts_k) - J_{C_i}(x_k) \right) s_k dt - \int_0^1 \left(J_{C_i}(x_k + td_k) - J_{C_i}(x_k) \right) d_k dt \right\|_2 \\
&\leq \frac{L_1}{2} \|s_k\|_2^2 + \frac{L_1}{2} \|d_k\|_2^2,
\end{aligned}$$

e também,

$$\begin{aligned}
B &= \left\| F_{C_i}(x_k) + J_{C_i}(x_k)s_k + \int_0^1 \left(J_{C_i}(x_k + ts_k) - J_{C_i}(x_k) \right) s_k dt \right\|_2 \\
&\quad + \left\| F_{C_i}(x_k) + J_{C_i}(x_k)s_k + \int_0^1 \left(J_{C_i}(x_k + td_k) - J_{C_i}(x_k) \right) d_k dt \right\|_2 \\
&\leq \|F_{C_i}(x_k) + J_{C_i}(x_k)s_k\|_2 + \left\| \int_0^1 \left(J_{C_i}(x_k + ts_k) - J_{C_i}(x_k) \right) s_k dt \right\|_2 \\
&\quad + \|F_{C_i}(x_k) + J_{C_i}(x_k)s_k\|_2 + \left\| \int_0^1 \left(J_{C_i}(x_k + td_k) - J_{C_i}(x_k) \right) d_k dt \right\|_2 \\
&\leq 2 \|F_{C_i}(x_k) + J_{C_i}(x_k)s_k\|_2 + \frac{L_1}{2} \|s_k\|_2^2 + \frac{L_1}{2} \|d_k\|_2^2.
\end{aligned}$$

Lembrando que $x_{k+1} = x_k$, para todo $k > \bar{k}$, e como $\|d_k\|_2 \rightarrow 0$ e $\|s_k\|_2 \rightarrow 0$.

Temos então que,

$$\begin{aligned}
& \|F_{C_i}(x_k + s_k)\|_2^2 - \left\| F_{C_i}(y_k) + J_{C_i}(x_k)\hat{d}_k \right\|_2^2 \\
& \leq \left(\frac{L_1}{2} \|s_k\|_2^2 + \frac{L_1}{2} \|d_k\|_2^2 \right) \left(2 \|F_{C_i}(x_k) + J_{C_i}(x_k)s_k\|_2 + \frac{L_1}{2} \|s_k\|_2^2 + \frac{L_1}{2} \|d_k\|_2^2 \right) \\
& = \|F_{C_i}(x_k) + J_{C_i}(x_k)s_k\|_2 L_1 \|s_k\|_2^2 + \frac{L_1^2}{4} \|s_k\|_2^4 + \frac{L_1^2}{4} \|s_k\|_2^2 \|d_k\|_2^2 \\
& \quad + \|F_{C_i}(x_k) + J_{C_i}(x_k)s_k\|_2 L_1 \|d_k\|_2^2 + \frac{L_1^2}{4} \|s_k\|_2^2 \|d_k\|_2^2 + \frac{L_1^2}{4} \|d_k\|_2^4 \\
& = \|F_{C_i}(x_k) + J_{C_i}(x_k)s_k\|_2 L_1 (\|d_k\|_2^2 + \|s_k\|_2^2) + \frac{L_1^2}{4} (\|s_k\|_2^4 + \|d_k\|_2^4) + \frac{L_1^2}{2} \|s_k\|_2^2 \|d_k\|_2^2 \\
& \rightarrow 0.
\end{aligned} \tag{6.26}$$

Novamente, pelo Lema 6.2 e por (6.12),

$$\begin{aligned}
& \|F_{C_i}(y_k)\|_2^2 - \|F_{C_i}(x_k) + J_{C_i}(x_k)d_k\|_2^2 = \\
& \underbrace{\left(\|F_{C_i}(y_k)\|_2 - \|F_{C_i}(x_k) + J_{C_i}(x_k)d_k\|_2 \right)}_C \underbrace{\left(\|F_{C_i}(y_k)\|_2 + \|F_{C_i}(x_k) + J_{C_i}(x_k)d_k\|_2 \right)}_D
\end{aligned}$$

Consequentemente,

$$\begin{aligned}
C & \leq \|F_{C_i}(y_k) - F_{C_i}(x_k) - J_{C_i}(x_k)d_k\|_2 \\
& = \left\| \int_0^1 J_{C_i}(x_k + td_k)d_k dt - J_{C_i}(x_k)d_k \right\|_2 \\
& = \left\| \int_0^1 \left(J_{C_i}(x_k + td_k) - J_{C_i}(x_k) \right) d_k dt \right\|_2 \\
& \leq \frac{L_1}{2} \|d_k\|_2^2,
\end{aligned}$$

e também,

$$\begin{aligned}
D & \leq \|F_{C_i}(y_k) + F_{C_i}(x_k) + J_{C_i}(x_k)d_k\|_2 \\
& = \left\| F_{C_i}(x_k) + \int_0^1 J_{C_i}(x_k + td_k)d_k dt + F_{C_i}(x_k) + J_{C_i}(x_k)d_k \right\|_2 \\
& = \left\| 2F_{C_i}(x_k) + 2J_{C_i}(x_k)d_k + \int_0^1 \left(J_{C_i}(x_k + td_k) - J_{C_i}(x_k) \right) d_k dt \right\|_2 \\
& \leq 2 \|F_{C_i}(x_k)\|_2 + 2 \|J_{C_i}(x_k)d_k\|_2 + \frac{L_1}{2} \|d_k\|_2^2.
\end{aligned}$$

Logo, como em (6.26) temos

$$\begin{aligned}
& \|F_{C_i}(y_k)\|_2^2 - \|F_{C_i}(x_k) + J_{C_i}(x_k)d_k\|_2^2 \\
&= \left(\frac{L_1}{2} \|d_k\|_2^2\right) \left(2 \|F_{C_i}(x_k)\|_2 + 2 \|J_{C_i}(x_k)d_k\|_2 + \frac{L_1}{2} \|d_k\|_2^2\right) \\
&\leq L_1 \|d_k\|_2^2 \|F_{C_i}(x_k)\|_2 + L_1 \|J_{C_i}(x_k)\|_2 \|d_k\|_2^3 + \frac{L_1^2}{4} \|d_k\|_2^4 \\
&\rightarrow 0.
\end{aligned} \tag{6.27}$$

Portanto, por (6.19), (6.24), (6.26) e (6.27),

$$\begin{aligned}
& |\rho_k - 1| \\
&\leq \left| \frac{L_1 \|F_{C_i}(x_k) + J_{C_i}(x_k)s_k\|_2 \left(\|d_k\|_2^2 + \|s_k\|_2^2\right) + \frac{L_1^2}{4} \left(\|s_k\|_2^4 + \|d_k\|_2^4\right) + \frac{L_1^2}{2} \|s_k\|_2^2 \|d_k\|_2^2}{Pred_k} \right| \\
&+ \left| \frac{L_1 \|d_k\|_2^2 \|F_{C_i}(x_k)\|_2 + L_1 \|J_{C_i}(x_k)\|_2 \|d_k\|_2^3 + \frac{L_1^2}{4} \|d_k\|_2^4}{Pred_k} \right|
\end{aligned}$$

onde,

$$\begin{aligned}
Pred_k &\geq \|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \min \left\{ \|d_k\|_2, \frac{\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2}{\|J_{C_i}(x_k)^T J_{C_i}(x_k)\|_2} \right\} \\
&+ \|J_{C_i}(x_k)^T F_{C_i}(y_k)\|_2 \min \left\{ \|\hat{d}_k\|_2, \frac{\|J_{C_i}(x_k)^T F_{C_i}(y_k)\|_2}{\|J_{C_i}(x_k)^T J_{C_i}(x_k)\|_2} \right\}.
\end{aligned}$$

Temos então que

$$\begin{aligned}
& |\rho_k - 1| \\
&\leq \left| \frac{L_1 \|F_{C_i}(x_k) + J_{C_i}(x_k)s_k\|_2 \left(\|d_k\|_2^2 + \|s_k\|_2^2\right) + \frac{L_1^2}{4} \left(\|s_k\|_2^4 + \|d_k\|_2^4\right) + \frac{L_1^2}{2} \|s_k\|_2^2 \|d_k\|_2^2}{\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \min \left\{ \|d_k\|_2, \frac{\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2}{\|J_{C_i}(x_k)^T J_{C_i}(x_k)\|_2} \right\}} \right| \\
&+ \left| \frac{L_1 \|d_k\|_2^2 \|F_{C_i}(x_k)\|_2 + L_1 \|J_{C_i}(x_k)\|_2 \|d_k\|_2^3 + \frac{L_1^2}{4} \|d_k\|_2^4}{\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \min \left\{ \|d_k\|_2, \frac{\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2}{\|J_{C_i}(x_k)^T J_{C_i}(x_k)\|_2} \right\}} \right|
\end{aligned}$$

Segue da hipótese de contradição e da desigualdade do valor médio que,

$$\begin{aligned}
& |\rho_k - 1| \\
& \leq \left| \frac{L_1 \|F_{C_i}(x_k) + J_{C_i}(x_k)s_k\|_2 \left(\|d_k\|_2^2 + \|s_k\|_2^2 \right) + \frac{L_1^2}{4} \left(\|s_k\|_2^4 + \|d_k\|_2^4 \right) + \frac{L_1^2}{2} \|s_k\|_2^2 \|d_k\|_2^2}{\tau \min \left\{ \|d_k\|_2, \frac{\tau}{L_2^2} \right\}} \right| \\
& + \left| \frac{L_1 \|d_k\|_2^2 \|F_{C_i}(x_k)\|_2 + L_1 \|J_{C_i}(x_k)\|_2 \|d_k\|_2^3 + \frac{L_1^2}{4} \|d_k\|_2^4}{\tau \min \left\{ \|d_k\|_2, \frac{\tau}{L_2^2} \right\}} \right| \\
& \rightarrow 0.
\end{aligned}$$

Concluimos que $\rho_k \rightarrow 1$. Tendo em vista a regra de atualização para μ_k , sabemos que existe uma constante positiva $\bar{M} > M$, tal que $\mu_k < \bar{M}$ vale para todo k suficientemente grande, o que contradiz (6.17). Assim a suposição (6.16) não é verdadeira quando K_2 é finito.

Caso II : K_2 é infinito.

Segue da hipótese de contradição (6.16), de (6.7), (6.13) e da desigualdade do valor médio que,

$$\begin{aligned}
\|F_{C_i}(x_1)\|_2^2 & \geq \sum_{k \in K_2} \|F_{C_i}(x_k)\|_2^2 - \|F_{C_i}(x_{k+1})\|_2^2 \geq \sum_{k \in K_2} q_1 \text{Pred}_k \\
& \geq \sum_{k \in K_2} q_1 \left(\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \min \left\{ \|d_k\|_2, \frac{\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2}{\|J_{C_i}(x_k)^T J_{C_i}(x_k)\|_2} \right\} \right) \\
& + \|J_{C_i}(x_k)^T F_{C_i}(y_k)\|_2 \min \left\{ \|\hat{d}_k\|_2, \frac{\|J_{C_i}(x_k)^T F_{C_i}(y_k)\|_2}{\|J_{C_i}(x_k)^T J_{C_i}(x_k)\|_2} \right\} \\
& \geq \sum_{k \in K_2} q_1 \frac{\tau}{2} \min \left\{ \|d_k\|_2, \frac{\tau}{2L_2^2} \right\},
\end{aligned} \tag{6.28}$$

o que implica que

$$\lim_{k \rightarrow \infty, k \in K_2} d_k = 0. \tag{6.29}$$

Pela definição de d_k temos que

$$\lambda_k = \frac{\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 + \|J_{C_i}(x_k)^T J_{C_i}(x_k) d_k\|_2}{\|d_k\|_2},$$

e, novamente por (6.16), (6.13) e pela desigualdade do valor médio temos

$$\lambda_k \geq \frac{\tau}{2 \|d_k\|_2} + L_2^2,$$

consequentemente,

$$\lambda_k \rightarrow \infty, \quad k \in K_2. \quad (6.30)$$

De modo similar ao que foi feito em (6.20 - 6.23), temos que existe uma constante $\tilde{c} \in \mathbb{R}^+$ tal que

$$\|\hat{d}_k\|_2 \leq \tilde{c} \|d_k\|_2, \quad (6.31)$$

vale para todo k suficientemente grande. Logo temos que,

$$\|s_k\|_2 = \|d_k + \hat{d}_k\|_2 \leq (1 + \tilde{c}) \|d_k\|_2.$$

Assim obtemos de (6.28) que

$$\sum_{k \in K_2} \|s_k\|_2 = \sum_{k \in K_2} \|d_k + \hat{d}_k\|_2 \leq \sum_{k \in K_2} (1 + \tilde{c}) \|d_k\|_2 \leq \infty.$$

Mais ainda, como destacado em [16, p. 453], temos por (6.12) e (6.13) que

$$\sum_{k \in K_2} \left| \|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 - \|J_{C_i}(x_{k+1})^T F_{C_i}(x_{k+1})\|_2 \right| < +\infty.$$

Dado que (6.16) vale para infinitos k , então existe um \hat{k} suficientemente grande tal que

$$\|J_{C_i}(x_{\hat{k}})^T F_{C_i}(x_{\hat{k}})\|_2 \geq \tau, \quad (6.32)$$

e

$$\sum_{k \in K_2, k \geq \hat{k}} \left| \|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 - \|J_{C_i}(x_{k+1})^T F_{C_i}(x_{k+1})\|_2 \right| < \frac{\tau}{2}.$$

Se $\hat{k} + 1 \in K_2$, então $\|J_{C_i}(x_{\hat{k}})^T F_{C_i}(x_{\hat{k}})\|_2 \geq \frac{\tau}{2}$. E se $\hat{k} + 1 \notin K_2$, temos duas opções $x_{\hat{k}+2} = x_{\hat{k}+1}$ ou $\|J_{C_i}(x_{\hat{k}+1})^T F_{C_i}(x_{\hat{k}+1})\|_2 \geq \tau$. O segundo caso não é verdadeiro, pois caso contrário (6.32) não seria válido. Consequentemente $x_{\hat{k}+2} = x_{\hat{k}+1}$, então

$$\|J_{C_i}(x_{\hat{k}+2})^T F_{C_i}(x_{\hat{k}+2})\|_2 = \|J_{C_i}(x_{\hat{k}+1})^T F_{C_i}(x_{\hat{k}+1})\|_2 \geq \frac{\tau}{2}. \quad (6.33)$$

Da mesma forma, se $\hat{k} + 2 \in K_2$, então $\|J_{C_i}(x_{\hat{k}})^T F_{C_i}(x_{\hat{k}})\|_2 \geq \frac{\tau}{2}$. E se $\hat{k} + 2 \notin K_2$, há duas opções $x_{\hat{k}+3} = x_{\hat{k}+2}$ ou $\|J_{C_i}(x_{\hat{k}+2})^T F_{C_i}(x_{\hat{k}+2})\|_2 \geq \tau$. O segundo caso não é verdadeiro, pois (6.33) não seria válido. Consequentemente $x_{\hat{k}+3} = x_{\hat{k}+2}$, então

$$\|J_{C_i}(x_{\hat{k}+3})^T F_{C_i}(x_{\hat{k}+3})\|_2 = \|J_{C_i}(x_{\hat{k}+2})^T F_{C_i}(x_{\hat{k}+2})\|_2 \geq \frac{\tau}{2}.$$

Por indução vemos que $\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \geq \frac{\tau}{2}$ vale para todo $k \geq \hat{k}$. Deduzimos de (6.28 - 6.31) que

$$d_k \rightarrow 0 \quad e \quad \hat{d}_k \rightarrow 0.$$

E segue da definição de d_k que,

$$\mu_k \rightarrow +\infty. \tag{6.34}$$

Em seguida, fazendo a mesma análise de (6.25 - 6.2), concluímos que $\rho_k \rightarrow 1$. Logo, existe uma constante positiva $\tilde{M} > M$ tal que $\mu_k < \tilde{M}$ para todo k suficientemente grande, o que contradiz (6.34). Assim a suposição (6.16) não é verdadeira quando K_2 é infinito. Portanto, (6.15) é válida. \square

6.3 O Método de Levenberg-Marquardt Modificado Acelerado para problemas LOVO

O algoritmo a seguir corresponde a uma adaptação do método de Levenberg-Marquardt Modificado Acelerado exposto no Capítulo 4, no contexto do problema LOVO irrestrito.

Aqui, como na subseção 4.3, a direção teste é calculada por $s_k = d_k + \alpha \hat{d}_k$, onde α_k é definido por

$$\alpha_k = \min\{\tilde{\alpha}_k, \hat{\alpha}\}, \tag{6.35}$$

em que $\hat{\alpha}$ é uma constante positiva, e $\tilde{\alpha}_k = 1 + \frac{\lambda_k \hat{d}_k^T \hat{d}_k}{\hat{d}_k^T J_{C_i}(x_k)^T J_{C_i}(x_k) \hat{d}_k} > 1$.

A convergência do método Levenberg-Marquardt Modificado Acelerado para problemas LOVO segue de forma análoga ao Teorema 6.7.

Algoritmo 9: LOVO LEVENBERG-MARQUARDT MODIFICADO ACELERADO

Entrada: $F(x)$, $J(x)$, uma aproximação inicial $x_0 \in \mathbb{R}^n$, $\mu_1 > M > 0$, $0 < q_1 \leq q_2 \leq q_3$,

$1 \leq \delta \leq 2$, $\hat{\alpha}$ e uma tolerância $\varepsilon > 0$.

$k \leftarrow 0$;

repita

Escolha $i \in I_{min}(x_k)$.

Calcule $\lambda_k = \mu_k \|F_{C_i}(x_k)\|_2^\delta$.

Obtenha d_k resolvendo:

$$(J_{C_i}(x_k)^T J_{C_i}(x_k) + \lambda_k I) d_k = -J_{C_i}(x_k)^T F_{C_i}(x_k);$$

Faça $y_k = x_k + d_k$;

Obtenha \hat{d}_k resolvendo

$$(J_{C_i}(x_k)^T J_{C_i}(x_k) + \lambda_k I) \hat{d}_k = -J_{C_i}(x_k)^T F_{C_i}(y_k);$$

Calcule α_k por (6.35) e faça

$$s_k = d_k + \alpha \hat{d}_k$$

Calcule $\rho_k = \frac{Ared_k}{Pred_k}$, como em (6.5) e (6.6).

Escolha x_{k+1} , onde

$$x_{k+1} = \begin{cases} x_k + s_k & \text{se } \rho_k \geq q_1; \\ x_k & \text{caso contrário.} \end{cases}$$

E escolha μ_{k+1} , como

$$\mu_{k+1} = \begin{cases} 4\mu_k, & \text{se } \rho_k < q_2; \\ \mu_k, & \text{se } \rho_k \in [q_2, q_3]; \\ \max \left\{ \frac{\mu_k}{4}, M \right\}, & \text{se } \rho_k > q_3. \end{cases}$$

Faça $k \leftarrow k + 1$

até $\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \leq \varepsilon$

retorna x_k

6.4 O Método de Levenberg-Marquardt Modificado Acelerado Adaptado para problemas LOVO

Assim como os dois algoritmos anteriores, este também é uma das nossas contribuições e corresponde a uma adaptação do método de Levenberg-Marquardt Modificado Acelerado Adaptado apresentado no Capítulo 4, no contexto do problema LOVO irrestrito. Neste método, o parâmetro de *damping* é calculado por $\lambda_k = \mu_k \|F_{C_i}(x_k)\|_2^\delta$, onde

$$\delta_k = \begin{cases} \frac{1}{\|F_{C_i}(x_k)\|_2}, & \text{se } \|F_{C_i}(x_k)\|_2 \geq 1; \\ 1 + \frac{1}{k} & \text{caso contrário.} \end{cases} \quad (6.36)$$

A convergência global do algoritmo segue pelo Teorema 6.5.

Teorema 6.5. *Seja $(x_k)_{k \in \mathbb{N}}$ uma sequência gerada pelo Algoritmo 10 e seja $\mathcal{K} = \{k_0, k_1, \dots\}$ é o conjunto de índices tais que $i \in I_{\min}(x_k)$, tal que $\lim_{k \in \mathcal{K}} x_k = x_*$. Suponha que $F_{C_i}(x)$ é continuamente diferenciável, e ambos $F_{C_i}(x)$ e sua Jacobiana $J_{C_i}(x)$ são Lipschitz contínuas. Então, x_* é fracamente crítico, $i \in I_{\min}(x_*)$ e*

$$\lim_{k \in \mathcal{K}} \|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 = 0.$$

Demonstração: Se $\|F_{C_i}(x_k)\|_2 \geq 1$ então $\delta_k = \frac{1}{\|F_{C_i}(x_k)\|_2} \in (0, 1]$ e se $\|F_{C_i}(x_k)\|_2 \leq 1$ então $\delta_k = 1 + \frac{1}{k} \in (1, 2]$. Portanto para quaisquer k , $\delta_k \in (0, 2]$, e a prova segue como no Teorema 6.7. \square

Algoritmo 10: LOVO LEVENBERG-MARQUARDT MODIFICADO ACELERADO

ADAPTADO

Entrada: $F(x)$, $J(x)$, uma aproximação inicial $x_0 \in \mathbb{R}^n$, $\mu_1 > M > 0$, $0 < q_1 \leq q_2 \leq q_3$, $\hat{\alpha}$ e uma tolerância $\varepsilon > 0$.

$k \leftarrow 0$;

repita

Escolha $i \in I_{min}(x_k)$.

Calcule $\lambda_k = \mu_k \|F_{C_i}(x_k)\|_2^\delta$. Onde δ é escolhido por (6.36),

Obtenha d_k resolvendo:

$$(J_{C_i}(x_k)^T J_{C_i}(x_k) + \lambda_k I) d_k = -J_{C_i}(x_k)^T F_{C_i}(x_k);$$

Faça $y_k = x_k + d_k$;

Obtenha \hat{d}_k resolvendo

$$(J_{C_i}(x_k)^T J_{C_i}(x_k) + \lambda_k I) \hat{d}_k = -J_{C_i}(x_k)^T F_{C_i}(y_k);$$

Calcule α_k por (4.17) e faça $s_k = d_k + \alpha \hat{d}_k$

Calcule $\rho_k = \frac{Ared_k}{Pred_k}$.

Escolha x_{k+1} , onde

$$x_{k+1} = \begin{cases} x_k + s_k & \text{se } \rho_k \geq q_1; \\ x_k & \text{caso contrário.} \end{cases}$$

E escolha μ_{k+1} , como

$$\mu_{k+1} = \begin{cases} 4\mu_k, & \text{se } \rho_k < q_2; \\ \mu_k, & \text{se } \rho_k \in [q_2, q_3]; \\ \max\left\{\frac{\mu_k}{4}, M\right\}, & \text{se } \rho_k > q_3. \end{cases}$$

Faça $k \leftarrow k + 1$

até $\|J_{C_i}(x_k)^T F_{C_i}(x_k)\|_2 \leq \varepsilon$

retorna x_k

EXPERIMENTOS NUMÉRICOS

Neste capítulo comparamos o desempenho numérico de cinco algoritmos para problemas LOVO. Os métodos LOVO Gauss-Newton, introduzido por Andreani et al [3, Algoritmo 1] e LOVO Levenberg-Marquardt, devido a Schwertner [32, p. 58], utilizam busca linear para determinar os novos iterandos. Os outros três algoritmos são as adaptações dos Métodos Levenberg-Marquardt Modificado, Levenberg-Marquardt Modificado Acelerado e Levenberg-Marquardt Acelerado Adaptado no contexto dos problemas LOVO, que abordamos no Capítulo 6, os quais utilizam técnicas de região de confiança.

Na realização dos testes usamos um conjunto de seis funções modelo estabelecidas por Schwertner em [32, p. 58] que, por não haver na literatura um conjunto de testes para problemas LOVO, elaborou tal família que encontra-se explicitada na Tabela (7.2). Cada teste foi realizado com três conjuntos de pontos com 100, 1000 e 5000 pontos, respectivamente, a uma taxa de confiança de 90%. Como critério de parada tomamos a norma do gradiente da função objetivo menor que uma tolerância, $\varepsilon = 10^{-4}$, e também o número máximo de 500 iterações.

Na Tabela 7.1 encontram-se especificados as siglas para cada método e os parâmetros utilizados para os testes realizados, tal escolha foi realizada de acordo com [16], [17] e [2], onde utilizados os valores com melhores resultados.

Métodos	Siglas	Parâmetros
Gauss-Newton	GN	$\lambda = 0.1$ e $\gamma = 10^{-4}$
Levenberg-Marquardt	LM	$\gamma = 10^{-2}$
Levenberg-Marquardt Modificado	LMM	$\mu_k = 0.5$, $\delta = 1.0$, $M = 10^{-6}$, $q_1 = 0.0001$, $q_2 = 0.25$ e $q_3 = 0.75$
Levenberg-Marquardt Modificado Acelerado	LMMA	$\mu_k = 0.5$, $\delta = 1.0$, $M = 10^{-6}$, $\alpha_{max} = 5.0$ $q_1 = 0.0001$, $q_2 = 0.25$ e $q_3 = 0.75$
Levenberg-Marquardt Modificado Acelerado Adaptado	LMMAAd	$\mu_k = 0.5$, $M = 10^{-6}$, $\alpha_{max} = 5.0$, $q_1 = 0.0001$, $q_2 = 0.25$ e $q_3 = 0.75$

Tabela 7.1: Parâmetros utilizados.

Os algoritmos foram implementados na linguagem *Julia*, versão 1.0.4. Os arquivos contendo os dados, estão disponibilizados no repositório aberto:

`github.com/ediduran/Levenberg-Marquardt_com_passos_multiplos_para_LOVO.git`

7.1 Análise dos Resultados

7.1.1 Funções testadas

A Tabela 7.2 descreve as informações essenciais sobre as funções modelo adotadas para gerar os dados empregados nos testes, para mais informações acerca da construção da família de funções ver [32].

Modelo	Expressão Matemática	Parâmetros adotados	Intervalo dos dados
Polinomial de Grau 1	$r(x) = ax + b$	$a = -3.2531, b = 15.2347$	$[-10.0, 10.0]$ $[-50.0, 50.0]$
Polinomial de Grau 3	$r(x) = ax^3 + bx^2 + cx + d$	$a = 1.12548, b = 2.53168$ $c = 3.14724, d = 0.589134$	$[-10.0, 10.0]$ $[-50.0, 50.0]$
Exponencial	$r(x) = ae^{bx+c} + d$	$a = 5.3127, b = -0.252$ $c = -0.755, d = -25.64$	$[-14.0, -6.0]$ $[-30.0, 10.0]$
Senoidal(Tipo 1)	$r(x) = a \sin(bx + c) + d$	$a = 40.5367, b = 2.345$ $c = -5.234, d = 24.12$	$[-0.5, 0.5]$ $[-2.5, 2.5]$
Senoidal(Tipo 2)	$r(x) = a \sin(bx) + c \cos(dx) + e$	$a = 13.535, b = -11.86$ $c = 12.8239, d = -6.30077$ $e = 4.82982$	$[-0.5, 0.5]$ $[-2.5, 2.5]$
Logístico	$r(x) = \frac{a}{1 + e^{bx+c}}$	$a = -10.5772, b = -4.52081$ $c = 19.6434$	$[2.0, 6.0]$ $[-6.0, 14.0]$

Tabela 7.2: Modelos utilizados para os testes.

7.1.2 Análises

A seguir analisamos os resultados dos testes numéricos realizados, cujos resultados são descritos na Subseção 7.1.3. Para facilitar a compreensão e deixar o texto menos cansativo usamos as seguintes abreviações “pts.”, “it.” e “a.f.”, para pontos, iterações e número de avaliação de função, respectivamente.

Polinomial de Grau 1

Analisando os dados apresentados nas Tabelas 7.3, 7.4 e 7.5 e comparando-os com os parâmetros adotados nas Tabela 7.2, podemos concluir que todos os métodos obtiveram sucesso na determinação dos possíveis parâmetros para o modelo.

Os métodos com o menor número de iterações foram GN e LMMAAd (100 pts.: 3 it.; 1000 pts.: 3 it.; 5000 pts.: 3 it.), no cenário com 1000 pontos temos que o métodos LMMA também resolveu o problema com 3 iterações. O método GN também foi o que conseguiu o menor número de avaliação de funções (100 pts.: 4 a.f.; 1000 pts.: 4 a.f.; 5000 pts.: 4 a.f.).

Note que o método LMM destaca-se por obter o menor tempo na resolução do problema LOVO, para os conjuntos de 100, e o método LMMA também destaca-se por obter o menor tempo na resolução do problema LOVO, para os conjuntos de 1000 pontos.

Polinomial de Grau 3

Pelos dados observados nas Tabelas 7.6 e 7.7, vemos que os métodos GN e LMMAAd obtiveram o menor número de iterações (100 pts.: 3 it.; 1000 pts.: 3 it.). Para 5000 pontos, o método de melhor número de iterações foi GN com 3 iterações e o método LMMAAd como 4 iterações. Considerando o menor número de avaliações de funções como critério de classificação em primeiro lugar ficou GN com (100 pts.: 4 a.f.; 1000 pts.: 3 a.f.; 5000 pts.: 4 a.f.) e em segundo lugar temos LM (100 pts.: 6 a.f.; 1000 pts.: 5 a.f.; 5000 pts.: 7 a.f.). Para 100 e 1000 pontos o método LMM utilizou um menor tempo para a resolução do problema, e para 5000 pontos o método com menor tempo foi GN.

Quando testamos o problema para 5000 pts., o método LMMA excedeu o número máximo de iterações. No entanto analisando o teste ponto a ponto, nota-se que o termo da quinta iteração já estava muito próximo de uma possível solução ($\|J_{C_i}(x_5)^T F_{C_i}(x_5)\|_2 = 0.000349$). Contudo o algoritmo não consegue reduzir essa norma para um valor menor que a tolerância $\epsilon = 10^{-4}$, performando diversas iterações até atingir o número máximo de iterações e obtendo o valor da norma do gradiente igual a 0.000331.

Exponencial

No teste realizado com 100 pts., GN resolveu o problema com o menor número de iterações e o menor número de avaliação de função (11 it.; 25 a.f.), enquanto o segundo melhor em relação ao número de iterações e ao número de avaliação de função foi o LMMA (17 it.; 30 a.f.). Para esse teste o método LMM resolveu o problema LOVO com o melhor tempo.

Pela Tabela 7.10, vemos que o método com o menor número de iterações e menor número de avaliação de função (9 it.; 12 a.f.) foi neste caso o LM. Para 5000 os métodos GN e LM resolveram o problema com apenas 9 iterações, onde o segundo resolveu em menos tempo.

Senoidal (Tipo 01)

Para o conjunto de 100 pts. dado pela Tabela 7.12, o método com melhores resultados nos testes foi o LMMA (8 it.; 17 a.f.), e o método que resolveu em um menor tempo foi o LMM. Ressaltamos que o segundo melhor desempenho em relação ao número de iterações e ao número de avaliação de função foi o LM (18 it.; 19 a.f.). Porém, observando os dados relativos a função objetivo na última iteração, nota-se que os métodos GN e LMM encontraram um ponto fracamente crítico que não é um minimizante global para o problema LOVO.

Já, quando o teste foi realizado para 1000 pontos LM fez o menor número de iterações, avaliação de função (12 it.; 14 a.f.) e também o menor tempo. Neste caso o segundo melhor método foi o LMM como 21 iterações, 41 avaliação de função, e também em segundo lugar considerando o tempo de resolução.

Observando a Tabela 7.14, vemos que apenas os métodos GN e LM resolveram o problema. Enquanto os métodos LMM, LMMA e LMMAAd excederam o número máximo de 500 iterações. Fazendo uma análise mais profunda, notamos que, para três casos a norma do gradiente oscila até um certo ponto e depois decresce vagarosamente, com os valores finais para a norma do gradiente sendo 0.000655 para LM, 0.000433 para LMM e 0.000244 para LMMAAd, isto nos dá indícios de que tais métodos estariam indo para uma solução contudo numa taxa muito pequena.

Senoidal (Tipo 02)

Observando os dados das Tabelas 7.15 e 7.16, vemos que em ambos os casos apenas os métodos GN e LMMAAd convergiram para uma solução. O melhor deles foi o LMMAAd como menor número de iterações e menor número de avaliação de função (7 it.; 14 a.f.). Os demais métodos não obtiveram sucesso. Analisando as iterações, concluímos que os métodos apresentaram a mesma dificuldade do problema anterior, em que a norma do gradiente passa a convergir apenas em um certo ponto, e quando o método começa a convergir, essa convergência é muito lenta. Note também que entre os métodos GN e LMMAAd, apenas o segundo convergiu para os parâmetros adotados na Tabela 7.2, observando o valor da função objetivo para GN, vemos que seu valor é muito grande, concluímos então que o método não encontrou

um minimizante global para o problema.

Já pela Tabela 7.17, todos os métodos convergiram, onde o melhor deles em respeito ao tempo, número de iterações e número de avaliação de função foi o GN (21 it.; 24 a.f.). ressaltando que o segundo método com melhor resultado foi o LMM (28 it.; 52 a.f.). No entanto, considerando os parâmetros da Tabela 7.2, vemos que nenhum dos métodos convergiu para os parâmetros originais. Analisando os valores da função objetivo na última iteração, nota-se que tais valores são relativamente grandes, nos levando a acreditar que os métodos encontraram um ponto fracamente crítico que não é um minimizante global.

Logístico

Como pode ser visto nas Tabelas 7.18, 7.19 e 7.20, concluímos que nos três cenários possíveis (100, 1000 e 5000 pts.) todos os métodos obtiveram sucesso na determinação dos possíveis parâmetros para o modelo.

Para o teste com 100 pontos, o método LMMA teve o menor número de iterações (11 it.) e o método LM o menor número de avaliação de função (21 it.), o melhor tempo foi obtido por LMMA. Com 1000 pontos, os melhores resultados foram do LM (10 it.; 13 a.f.), e do LMM (13 it.; 25 a.v.), onde o segundo resolveu o problema em menor tempo. Testando 5000 pontos, os métodos LMM, LMMA e LMMAad têm o menor número de iterações (10 it.), e o menor tempo foi do LMM.

7.1.3 Tabelas com Resultados

As próximas tabelas apresentadas são referentes a cada teste realizado. Para simplificar a escrita, representamos Solução, Número de Iterações, Número de Avaliações de Função, Tempo em segundos e Função Objetivo pelas abreviações “Sol.”, “n° It.”, “n° a.f.”, “t(s)” e “f.o.” respectivamente. Para indicar que o método excedeu o número máximo de iterações (500 iterações), utilizamos a sigla ITMAX.

Método	100 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	[-3.2531, 15.2347]	3	4	0.079723	0.0
LM	[-3.2531, 15.2347]	5	6	0.076312	3.018776e-16
LMM	[-3.2531, 15.2347]	4	9	0.075440	0.0
LMMA	[-3.2531, 15.2347]	4	9	0.076001	0.0
LMMAAd	[-3.2531, 15.2347]	3	7	0.081612	0.0

Tabela 7.3: Polinomial de Grau 1 para 100 pontos, com ponto inicial $x_0 = (0.0, 0.0)$.

Método	1000 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	[-3.2531, 15.2347]	3	4	0.084586	0.0
LM	[-3.2531, 15.2347]	5	6	0.091813	5.998498e-27
LMM	[-3.2531, 15.2347]	4	9	0.080746	0.0
LMMA	[-3.2531, 15.2347]	3	7	0.078853	2.745946e-26
LMMAAd	[-3.2531, 15.2347]	3	7	0.089551	0.0

Tabela 7.4: Polinomial de Grau 1 para 1000 pontos, com ponto inicial $x_0 = (0.0, 0.0)$.

Método	5000 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	[-3.2531, 15.2347]	3	4	0.122803	8.529795e-26
LM	[-3.2531, 15.2347]	5	6	0.157496	1.123744e-23
LMM	[-3.2531, 15.2347]	5	11	0.160947	1.974992e-26
LMMA	[-3.2531, 15.2347]	4	9	0.147372	0.0
LMMAAd	[-3.2531, 15.2347]	3	7	0.135554	1.974992e-26

Tabela 7.5: Polinomial de Grau 1 para 5000 pontos, com ponto inicial $x_0 = (0.0, 0.0)$.

Método	100 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	[1.12548, 2.53168, 3.14724, 0.589134]	3	4	0.086423	5.332099e-25
LM	[1.12548, 2.53168, 3.14724, 0.589134]	5	6	0.100103	1.073908e-20
LMM	[1.12548, 2.53168, 3.14724, 0.589134]	5	9	0.084860	1.906562e-23
LMMA	[1.12548, 2.53168, 3.14724, 0.589134]	4	9	0.087789	9.328323e-25
LMMAAd	[1.12548, 2.53168, 3.14724, 0.589134]	3	7	0.100816	8.820960e-25

Tabela 7.6: Polinomial de Grau 3 para 100 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0, 0.0)$.

Método	1000 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	[1.12548, 2.53168, 3.14724, 0.589134]	3	4	0.096011	4.531105e-24
LM	[1.12548, 2.53168, 3.14724, 0.589134]	4	5	0.102194	9.360708e-12
LMM	[1.12548, 2.53168, 3.14724, 0.589134]	4	9	0.088986	5.300664e-24
LMMA	[1.12548, 2.53168, 3.14724, 0.589134]	5	11	0.100161	6.524068e-24
LMMAAd	[1.12548, 2.53168, 3.14724, 0.589134]	3	7	0.104986	8.122230e-24

Tabela 7.7: Polinomial de Grau 3 para 1000 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0, 0.0)$.

Método	5000 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	[1.12548, 2.53168, 3.14724, 0.589134]	3	4	0.156102	1.618838e-19
LM	[1.12548, 2.53168, 3.14724, 0.589134]	6	7	0.169835	1.220360e-19
LMM	[1.12548, 2.53168, 3.14724, 0.589134]	5	11	0.171665	1.049801e-19
LMMA	ITMAX Norma do gradiente: 0.000331	-	-	0.549122	2.849703e-19
LMMAAd	[1.12548, 2.53168, 3.14724, 0.589134]	4	9	0.186553	1.065129e-19

Tabela 7.8: Polinomial de Grau 3 para 5000 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0, 0.0)$.

Método	100 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	[2.79552, -0.252, -0.112917, -25.64]	11	25	0.087087	1.674362e-11
LM	[2.96159, -0.252, -0.170627, -25.64]	27	121	0.097959	6.661141e-27
LMM	[3.1623, -0.252, -0.236198, -25.64]	19	34	0.085341	8.796214e-20
LMMA	[3.42227, -0.252, -0.315204, -25.64]	17	30	0.087570	9.100031e-20
LMMAAd	[3.45667, -0.252, -0.325206, -25.64]	26	43	0.090978	1.212710e-22

Tabela 7.9: Exponencial para 100 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0, 0.0)$.

Método	1000 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	[0.827048, -0.252, 1.10499, -25.64]	11	17	0.093939	9.370822e-13
LM	[3.40436, -0.252, -0.309956, -25.64]	9	12	0.104007	7.799695e-22
LMM	[3.42392, -0.252, -0.315685, -25.64]	20	35	0.104633	8.200998e-26
LMMA	[3.4882, -0.252, -0.334285, -25.64]	22	37	0.105498	1.406691e-19
LMMAAd	[3.4188, -0.252, -0.31419, -25.64]	26	41	0.107428	1.610648e-18

Tabela 7.10: Exponencial para 1000 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0, 0.0)$.

Método	5000 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	[67.5363, -0.252, -3.29757, -25.64]	9	12	0.208101	2.678761e-21
LM	[27.934, -0.252, -2.41474, -25.64]	9	14	0.204384	1.034315e-21
LMM	[53.6582, -0.252, -3.06754, -25.64]	27	44	0.361357	4.876733e-23
LMMA	[65.6955, -0.252, -3.26993, -25.64]	26	42	0.374139	9.427877e-22
LMMAAd	[83.6198, -0.252, -3.51118, -25.64]	35	53	0.361861	8.753382e-22

Tabela 7.11: Exponencial para 5000 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0, 0.0)$.

Método	100 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	[-4.62229, 13.961, 5.98103, 54.6642]	108	219	0.089443	8879.636564
LM	[40.5367, -2.345, 8.37559, 24.12]	18	19	0.105522	3.221781e-13
LMM	[4.62228, -13.961, 6.58534, 54.6642]	69	125	0.088487	8879.636564
LMMA	[40.5367, 2.345, 13.6156, 24.12]	8	17	0.092313	7.251093e-16
LMMAAd	[40.5367, 2.345, -11.5172, 24.12]	29	51	0.098613	2.541653e-23

Tabela 7.12: Senoidal Tipo 01 para 100 pontos, com ponto inicial $x_0 = (1.0, 1.0, 1.0, 1.0)$.

Método	1000 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	[-1.46229, 42.3382, -1.78101, 54.8688]	136	279	0.220228	95766.197154
LM	[40.5367, 2.345, 7.33237, 24.12]	12	14	0.110281	1.869608e-14
LMM	[40.5367, 2.345, 7.33237, 24.12]	21	41	0.115905	8.576493e-20
LMMA	[40.5367, -2.345, 8.37559, 24.12]	69	122	0.133716	8.008835e-19
LMMAAd	[-2.30418, 28.3555, -7.3397, 54.8764]	70	127	0.135265	94359.130425

Tabela 7.13: Senoidal Tipo 01 para 1000 pontos, com ponto inicial $x_0 = (1.0, 1.0, 1.0, 1.0)$.

Método	5000 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	[-5.78448, 12.3957, -6.58737, 16.644]	25	29	0.405611	2.891014e6
LM	[21.5109, 4.2051, 4.16853, 13.533]	49	276	0.678183	2.082790e6
LMM	ITMAX Norma do gradiente: 0.000655	-	-	0.753710	2.940816e6
LMMA	ITMAX Norma do gradiente: 0.000433	-	-	0.842518	2.927187e6
LMMAAd	ITMAX Norma do gradiente: 0.000244	-	-	0.451682	2.833132e6

Tabela 7.14: Senoidal Tipo 01 para 5000 pontos, com ponto inicial $x_0 = (1.0, 1.0, 1.0, 1.0)$.

Método	100 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	$[-6.31897, -20.1616, 12.8251, 4.43783, -2.18673]$	24	29	0.086566	6180.921943
LM	ITMAX Norma do gradiente: 0.031021	-	-	0.120686	5344.791096
LMM	ITMAX Norma do gradiente: 0.030207	-	-	0.116982	5343.885592
LMMA	ITMAX Norma do gradiente: 0.036290	-	-	0.120895	5343.883235
LMMAAd	$[13.535, -11.86, 12.8239, 6.30077, 4.82982]$	7	14	0.089935	2.991876e-18

Tabela 7.15: Senoidal Tipo 02 para 100 pontos, com ponto inicial $x_0 = (5.0, 5.0, 5.0, 5.0, 5.0)$.

Método	1000 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	$[-6.31897, -20.1616, 12.8251, 4.43783, -2.18673]$	24	29	0.092778	6180.921943
LM	ITMAX Norma do gradiente: 0.031021	-	-	0.121699	5344.791096
LMM	ITMAX Norma do gradiente: 0.030207	-	-	0.121191	5343.885592
LMMA	ITMAX Norma do gradiente: 0.036290	-	-	0.122296	5343.883235
LMMAAd	$[13.535, -11.86, 12.8239, 6.30077, 4.82982]$	7	14	0.092165	2.991876e-18

Tabela 7.16: Senoidal Tipo 02 para 1000 pontos, com ponto inicial $x_0 = (5.0, 5.0, 5.0, 5.0, 5.0)$.

Método	5000 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	$[3.15606, 5.12259, 10.6147, 6.20736, 2.79645]$	21	24	0.406421	337100.601659
LM	$[3.15606, 5.12259, 10.6147, 6.20736, 2.79645]$	36	367	0.587235	337100.601650
LMM	$[3.15606, 5.12259, 10.6147, 6.20736, 2.79645]$	28	52	0.409223	337100.601650
LMMA	$[3.15606, 5.12259, 10.6147, 6.20736, 2.79645]$	29	54	0.416836	337100.601650
LMMAAd	$[3.15606, 5.12259, 10.6147, 6.20736, 2.79645]$	31	53	0.461618	337100.601650

Tabela 7.17: Senoidal Tipo 02 para 5000 pontos, com ponto inicial $x_0 = (5.0, 5.0, 5.0, 5.0, 5.0)$.

Método	100 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	$[-10.5772, -4.52081, 19.6434]$	18	81	0.089792	5.693397e-18
LM	$[-10.5772, -4.52081, 19.6434]$	13	21	0.103252	2.711367e-11
LMM	$[-10.5772, -4.52081, 19.6434]$	14	27	0.088883	2.545605e-24
LMMA	$[-10.5772, -4.52081, 19.6434]$	11	22	0.088545	6.495063e-13
LMMAAd	$[-10.5772, -4.52081, 19.6434]$	17	30	0.100088	7.434560e-11

Tabela 7.18: Logístico para 100 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0)$.

Método	1000 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	$[-10.5772, -4.52081, 19.6434]$	16	54	0.102370	2.705574e-24
LM	$[-10.5772, -4.52081, 19.6434]$	10	13	0.110454	8.167556e-21
LMM	$[-10.5772, -4.52081, 19.6434]$	13	25	0.106702	0.0
LMMA	$[-10.5772, -4.52081, 19.6434]$	14	26	0.106730	1.495200e-12
LMMAAd	$[-10.5772, -4.52081, 19.6434]$	21	36	0.116307	1.414200e-20

Tabela 7.19: Logístico para 1000 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0)$.

Método	5000 pontos				
	Sol.	n° It.	n° a.f.	t(s)	f.o.
GN	$[-10.5772, -4.52081, 19.6434]$	13	16	0.252878	1.130004e-25
LM	$[-10.5772, -4.52081, 19.6434]$	13	15	0.252234	2.881284e-17
LMM	$[-10.5772, -4.52081, 19.6434]$	10	21	0.290513	2.435614e-27
LMMA	$[-10.5772, -4.52081, 19.6434]$	10	21	0.230281	1.026780e-27
LMMAAd	$[-10.5772, -4.52081, 19.6434]$	10	21	0.229097	1.063983e-26

Tabela 7.20: Logístico para 5000 pontos, com ponto inicial $x_0 = (0.0, 0.0, 0.0)$.

CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Nesta dissertação, trabalhamos com métodos para resolução de problemas de otimização de menor valor ordenado. Dada a importância prática de problemas LOVO em diversas áreas, o estudo de métodos capazes de resolver o problema em questão de forma eficiente é um tema relevante dentro da área de otimização não linear.

Especificamente, consideramos três métodos de passos múltiplos baseados no método de Levenberg-Marquardt e introduzimos estes métodos no contexto de problemas LOVO. Dessa forma, nossa proposta resultou em três novos algoritmos, para os quais estabelecemos resultados de convergência à pontos fracamente críticos, sendo esta, nossa principal contribuição neste trabalho. Para validar os novos métodos realizamos implementações e testes numéricos com outros métodos presentes na literatura.

Considerando o estudo numérico exposto, notamos que de modo geral, os métodos apresentados em [3] e [32] foram melhores do ponto de vista da avaliação de função e do número de iterações. Neste sentido, uma análise superficial, sugere que o parâmetro de *damping* não está sendo atualizado de forma adequada as adaptações. Assim, uma sugestão para um trabalho futuro seria o estudo de novas formas de atualização do parâmetro de *damping*, almejando uma melhoria na robustez e eficiência destes métodos. Outras sugestões incluem o estudo da ordem de convergência destes métodos e o estabelecimento de condições para a convergência global à pontos fortemente críticos.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Allaire, G., Kaber, S. M. (2008), *Numerical linear algebra* (Vol. 55), New York: Springer.
- [2] Amini, K., Rostami, F. (2015). A modified two steps Levenberg–Marquardt method for nonlinear equations. *Journal of Computational and Applied Mathematics*, 288, 341-350.
- [3] Andreani, R., Cesar, G., Cesar-Jr, R. M., Martínez, J. M., e Silva, P. J. S. (2007). Efficient curve detection using a Gauss-Newton method with applications in agriculture., *Proc. 1st International Workshop on Computer Vision Applications for Developing Regions in Conjunction with ICCV, 2007-CVDR-ICCV07*.
- [4] Andreani, R., Martínez, J. M., Martínez, L., e Yano, F. S. (2005). Low order-value optimization and applications (Technical report MCDO 051013). Campinas: IMECC, Universidade Estadual de Campinas.
- [5] Benatti, K. A. (2017), O Método de Levenberg-Marquardt para o Problema de Quadrados Mínimos não Linear, Dissertação de Mestrado, Universidade Federal do Paraná, Curitiba, PR.
- [6] Birgin, E. G., Bueno, L. F., Krejić, N., e Martínez, J. M. (2011). Low order-value approach for solving VaR-constrained optimization problems. *Journal of Global Optimization*, 51(4), 715-742.
- [7] Cabral, I., Perdigão, C., e Saiago, C. (2012), *Álgebra Linear: Teoria, Exercícios resolvidos e Exercícios propostos com soluções*, 3ª edição, Lisboa: Escolar Editora.
- [8] Castelani et al., (2019). RAFF.jl: Robust Algebraic Fitting Function in Julia. *Journal of Open Source Software*, 4(39), 1385, <https://doi.org/10.21105/joss.01385>

-
- [9] Conn, A. R., Gould, N. I., e Toint, P. L. (2000). Trust region methods (Vol. 1). SIAM.
- [10] Demmel, J. W. (1997), *Applied numerical linear algebra* (Vol. 56), Philadelphia: SIAM.
- [11] Delfino, A. R. (2010). Um método ótimo para otimização convexa irrestrita.
- [12] Dennis, J. E., Schnabel R. B. (1996), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Vol. 16), Philadelphia: SIAM.
- [13] Fan, J. Y., Yuan, Y. X. (2001). On the convergence of a new Levenberg-Marquardt method. *Technical Report, AMSS, Chinese Academy of Sciences*, 1-11.
- [14] Fan, J. Y. (2003). A modified Levenberg-Marquardt algorithm for singular system of nonlinear equations. *Journal of Computational Mathematics*, 625-636.
- [15] Fan, J. Y., Yuan, Y. X. (2005). On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption. *Computing*, 74(1), 23-39.
- [16] Fan, J. (2012). The modified Levenberg-Marquardt method for nonlinear equations with cubic convergence. *Mathematics of Computation*, 81(277), 447-466.
- [17] Fan, J. (2014). Accelerating the modified Levenberg-Marquardt method for nonlinear equations. *Mathematics of Computation*, 83(287), 1173-1187.
- [18] Friedlander, A. (2012), *Elementos de programação não-linear*, Campinas: UNICAMP.
- [19] Guidorizzi, H. L. (2001), *Um curso de cálculo* (Vol. 1), Rio de Janeiro: LTC.
- [20] Intriligator, M. D. (2002), *Mathematical Optimization and Economic Theory*, Philadelphia: Siam.
- [21] Legendre, A. M. (1805). Nouvelles méthodes pour la détermination des orbites des comètes. F. Didot.
- [22] Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares, *Quarterly of Applied Mathematics*, 2(2), 164-168.
- [23] Katz, V. J. (2008). The History of Mathematics: An Introduction. Pearson.

- [24] Lima, E. L. (2015). Curso de análise, vol. 2, Projeto Euclides. Rio de Janeiro: IMPA.
- [25] Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431-441.
- [26] Martínez, J. M., e Santos, S. A. (1995). Métodos computacionais de otimização. *Colóquio Brasileiro de Matemática*, Apostilas, 20.
- [27] Martínez, J. M. (2009). Order-value optimization and new applications. In 6th International Congress on Industrial and Applied Mathematics Zürich, Switzerland, 16-20 July 2007 (pp. 279-296).
- [28] Meyer, C. D. (2000). *Matrix analysis and applied linear algebra* (Vol. 71), Philadelphia: Siam.
- [29] Nocedal, J., Wright, S. J. (2006), *Numerical Optimization*, 2 edition, New York: Springer.
- [30] Powell, M. J. D. (1975). Convergence properties of a class of minimization algorithms. In *Nonlinear programming 2* (pp. 1-27). Academic Press.
- [31] Santos, R. N. (2007), *Uma breve história do desenvolvimento das teorias dos determinantes e das matrizes*, São Paulo: IME, USP (<http://milanesa.ime.usp.br/imath/files/1/43.pdf>).
- [32] Schwertner, A. E. (2019), O Método de Levenberg-Marquardt para problemas de Otimização de Menor Valor Ordenado, Dissertação de Mestrado, Universidade Estadual de Maringá, Maringá, PR.
- [33] Watkins, D. S. (2010), *Fundamentals of matrix computations* (Vol. 64), New York: John Wiley e Sons.
- [34] Yamashita, N., e Fukushima, M. (2001). On the rate of convergence of the Levenberg-Marquardt method. *Topics in Numerical Analysis*, p.239-249, Vienna: Springer.
- [35] Yano, F. S. (2006), Otimização da menor soma de valores ordenados, Tese de Doutorado, Universidade Estadual de Campinas, Campinas, SP.