UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS REITOR JOÃO DAVID FERREIRA LIMA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA PURA E APLICADA

Jonatan Ismael Eisermann

Um Método de Região de Confiança Escalada Secante para Sistemas Quadrados de Equações Não Lineares Sujeitos a Restrições de Caixa

Florianópolis
2021

Jonatan Ismael Eisermann

Um Método de Região de Confiança Escalada Secante para Sistemas Quadrados de Equações Não Lineares Sujeitos a Restrições de Caixa

Dissertação submetida ao Programa de Pós-Graduação em Matemática Pura e Aplicada da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Matemática.
Orientador: Prof. Dr. Juliano de Bem Francisco.

Florianópolis
2021

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Eisermann, Jonatan Ismael
Um Método de Região de Confiança Escalada Secante para Sistemas Quadrados de Equações Não Lineares Sujeitos a Restrições de Caixa / Jonatan Ismael Eisermann ; orientador, Juliano de Bem Francisco, 2021.
101 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências Físicas e Matemáticas, Programa de Pós-Graduação em Matemática Pura e Aplicada, Florianópolis, 2021.

Inclui referências.

1. Matemática Pura e Aplicada. 2. Sistemas Quadrados de Equações Não Lineares. 3. Restrições de Caixa. 4. Região de Confiança Afim-Escala. 5. Método Quase-Newton. I. Francisco, Juliano de Bem. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Matemática Pura e Aplicada. III. Título.

Jonatan Ismael Eisermann

Um Método de Região de Confiança Escalada Secante para Sistemas Quadrados de Equações Não Lineares Sujeitos a Restrições de Caixa

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Dr. Douglas Soares Gonçalves
Universidade Federal de Santa Catarina

Prof. Dr. Fermín Sinforiano Viloche Bazán
Universidade Federal de Santa Catarina

Prof. Dr. Lucas Garcia Pedroso
Universidade Federal do Paraná

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Mestre em Matemática.

ICP-Edu
Documento assinado digitalmente
Daniel Goncalves
Data: 14/06/2021 07:29:19-0300
CPF: 017.022.459-77
Verifique as assinaturas em https://v.ufsc.br

Prof. Dr. Daniel Gonçalves
Coordenador do Programa

ICP-Edu
Documento assinado digitalmente
Juliano de Bem Francisco
Data: 14/06/2021 08:19:48-0300
CPF: 019.470.379-78
Verifique as assinaturas em https://v.ufsc.br

Prof. Dr. Juliano de Bem Francisco.
Orientador

Florianópolis, 2021.

Dedico este trabalho aos meus pais, Ademir e Janete, e à minha querida avó Celeste.

AGRADECIMENTOS

A Deus, que em sua infinita bondade nos ofereceu a vida e, junto dela, a oportunidade de nos aprimorarmos a cada dia, de questionar realidades, refletir e encorajar-se para assumir novas possibilidades; por alimentar sonhos de um mundo mais justo, igualitário e fraterno em mim e em muitas outras pessoas.

Aos meus pais, Ademir e Janete, por não medirem esforços na criação e educação dos seus quatro filhos; à minha avó, Celeste que sempre esteve do meu lado me apoiando e me incentivando na realização de meus sonhos; aos meus irmãos, Junior Henrique, Jean Ricardo e João Felipe, por toda ajuda prestada.

Aos meus amigos que compartilharam e continuam compartilhando comigo tantos momentos de aprendizado, companheirismo, amor, alegrias, superação, união. Em especial àqueles que me acompanharam de perto nos dois anos de mestrado e deram todo apoio necessário: Maritza C. A. Brito, Diego Martins, Matheus W. S. Carvalho, Guilherme S. de Godoy e Marisvaldo S. Lima. Também, à Helena Günther, Francieli Triches e Everton Boos pelas várias ajudas prestadas e conselhos dados.

A todos(as) professores(as) que fizeram e fazem parte de minha trajetória de vida, por toda educação que me constituiu um ser humano melhor. De maneira especial, ao meu orientador, prof. Juliano de Bem Francisco, pelos auxílios prestados durante a realização desta importante etapa de minha formação acadêmica.

À Universidade Federal de Santa Catarina (UFSC), por me acolher tão bem e proporcionar uma formação de qualidade a mim e a tantos outros estudantes. À Elisa Amaral e à Érica Flores pelas ajudas prestadas na secretaria do Programa de Pós-Graduação em Matemática Pura e Aplicada da UFSC.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Por fim, a todas pessoas que lutaram, lutam ou que ainda lutarão pelo crescimento da ciência no país, alicerçado na expansão, democratização e qualificação do ensino superior brasileiro.

“Não preciso saber a tua idade,
nem onde mora ou com o que trabalha.
Quero saber da sua relação com as estrelas,
o quanto de cura tem no teu sorriso
e se há amor na tua fala.”
(Autor desconhecido).

RESUMO

Apresentamos um método iterativo para resolução de sistemas quadrados de equações não lineares sujeito a restrições de caixa. Este método, nomeado de STRQN (*Scaled Trust-Region Quasi-Newton*), combina ideias das clássicas regiões de confiança Quase-Newton para equações não lineares irrestritas e da abordagem afim-escala para problemas de otimização restritos. O método gera somente iterações viáveis e lida com os limites da caixa implicitamente, quando estes existem; para os casos em que não existem limites inferior e superior nas variáveis, o método é reduzido a um método de região de confiança padrão para problemas irrestritos. Propriedades locais e globais de convergência rápida para o método STRQN foram obtidas. O desempenho numérico do método foi analisado sobre noventa problemas-teste, e comparado, posteriormente, com o desempenho de dois métodos clássicos para sistemas de equações não lineares sujeitos a restrições de caixa. Esta comparação evidenciou altos índices de robustez e eficiência do método STRQN, principalmente ao utilizar da tradicional atualização secante Simétrica de Posto Um (SR1).

Palavras-chave: Equações Não Lineares. Restrições de Caixa. Afim-Escala. Região de Confiança. Quase-Newton.

ABSTRACT

We present an iterative method for solving square systems of nonlinear equations subject to box constraints. This method, named STRQN (*Scaled Trust-Region Quasi-Newton*), combines ideas from the classic trust-region Quasi-Newton method for unconstrained nonlinear equations and the affine-scale approach for constrained optimization problems. The method generates only feasible iterates and handles the bounds implicitly, when these exist; for cases where there are no upper and lower limits on the variables, the method becomes a standard trust-region method for unconstrained problems. Global and local fast convergence properties for the STRQN method were obtained. The numerical performance of the method was analyzed on ninety test-problems, and compared, subsequently, with the performance of two classic methods for systems of nonlinear equations subject to box constraints. This comparison showed high levels of robustness and efficiency of the STRQN method, mainly when using the traditional Symmetric Rank One (SR1) secant update.

Keywords: Nonlinear Equations. Box constraints. Affine-Scale. Trust-Region. Quasi-Newton.

LISTA DE FIGURAS

Figura 1 – Primeiras iterações do método de Newton para a função $F(x) = x^2$, partindo do ponto inicial $x_0 = 2$. 20
Figura 2 – Aproximação (local) da função não linear $F(x) = x^3 - x^2 - 1$ pelo modelo linear $M_{k+1}(x) = 4x - 5$. 23
Figura 3 – Passo Dogleg. 44
Figura 4 – Possibilidade de passo maior para uma região de confiança elíptica em um problema com restrição de caixa. 51
Figura 5 – Comparação entre F e f para a função $F(x) = x^3 - x^2 + 1$ 54
Figura 6 – Comportamento de λ em uma caixa em $\mathbb{R}^3$. 56
Figura 7 – Perfil de desempenho baseado no tempo de execução dos métodos STRQN-BFGS, STRQN-SR1 e STRQN-Broyden. 76
Figura 8 – Perfil de desempenho baseado no tempo de execução dos métodos STRQN-SR1, STRN e PLM. 86

LISTA DE TABELAS

Tabela 1 – Problemas-teste. 72
Tabela 2 – Desempenho dos métodos STRQN-SR1, STRN e PLM. 83
Tabela 3 – Desempenho dos métodos STRQN-BFGS, STRQN-SR1 e STRQN-Broyden. 97
Tabela 4 – Falhas dos métodos STRQN. 100

SUMÁRIO

1	**INTRODUÇÃO**	**12**
2	**SISTEMAS QUADRADOS DE EQUAÇÕES NÃO LINEARES**	**18**
2.1	MÉTODOS CLÁSSICOS	18
2.1.1	**Método de Newton**	**19**
2.1.2	**Métodos Quase-Newton**	**22**
2.2	ANÁLISE DE CONVERGÊNCIA	27
2.2.1	**Convergência do Método de Newton**	**30**
2.2.2	**Convergência dos Métodos Quase-Newton**	**32**
2.3	MÉTODO DE REGIÃO DE CONFIANÇA	38
2.3.1	**O ponto de Cauchy**	**42**
2.3.2	**O método Dogleg**	**43**
2.3.3	**Escalamento Afim**	**47**
3	**MÉTODO QUASE-NEWTON PARA RESOLUÇÃO DE SISTEMAS QUADRADOS DE EQUAÇÕES NÃO LINEARES COM RESTRIÇÕES DE CAIXA**	**53**
3.1	DESCRIÇÃO DO MÉTODO	53
3.2	RESULTADOS DE CONVERGÊNCIA	60
4	**RESULTADOS NUMÉRICOS**	**71**
4.1	IMPLEMENTAÇÃO COMPUTACIONAL DO MÉTODO STRQN	73
4.2	COMPARAÇÃO ENTRE MÉTODOS STRQN	75
4.3	COMPARAÇÃO ENTRE OS MÉTODOS STRQN-SR1, STRN E PLM	77
5	**CONCLUSÃO**	**89**
	REFERÊNCIAS	**91**
	APÊNDICE A – PERFIS DE DESEMPENHO	**95**
	APÊNDICE B – TABELA DE DESEMPENHO DOS MÉTODOS STRQN	**97**
	APÊNDICE C – TABELA DE FALHAS DOS MÉTODOS STRQN	**100**

1 INTRODUÇÃO

A resolução de sistemas não lineares com restrições canalizadas nas variáveis das equações envolvidas, também denominadas restrições de caixa, configuram um importante campo de estudos da Otimização. Problemas deste tipo ocorrem em modelos matemáticos de situações reais em que, por algum critério físico ou de outra natureza, a solução deve estar canalizada. Estas questões motivam o estudo, desenvolvimento e constante aprimoramento de métodos específicos que resolvam tais problemas de maneira robusta e eficiente.

Consideremos, então, uma caixa n-dimensional $\Omega \subseteq \mathbb{R}^n$, definida por

$$\Omega = \left\{ x \in \mathbb{R}^n \;\middle|\; l \leq x \leq u \right\},$$

onde l e u são vetores com n coordenadas que denotam, respectivamente, os limites inferior e superior da caixa. Suas componentes são da forma $l_i \in \mathbb{R} \cup \{-\infty\}$, $u_i \in \mathbb{R} \cup \{\infty\}$ e $l_i < u_i$, para $i = 1, ..., n$. Logo, Ω sempre é um conjunto não vazio.

Matematicamente, um sistema quadrado de equações não lineares sujeito a restrições de caixa pode ser descrito a partir de uma função vetorial $F : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^n$, onde cada componente de F é uma equação do sistema. Assim, a tarefa de encontrar uma solução viável para o referido problema será resumida em identificar algum $x \in \Omega$ que satisfaça

$$F(x) = 0, \tag{1}$$

isto é, identificar uma raiz de F em Ω. Como hipótese, assumimos neste estudo que F é continuamente diferenciável em um conjunto aberto que contém Ω.

Métodos computacionais para o problema correspondente irrestrito, onde $\Omega = \mathbb{R}^n$, foram discutidos em muitas obras clássicas da Otimização [8, 18, 27, 29]. Um enfoque específico na resolução de sistemas não lineares quadrados irrestritos a partir de métodos Quase-Newton foi dado, mais recentemente, nos estudos de Yuan e Wei [35] e de Zeng e Fu [36]. Em ambos, métodos secantes foram combinados com uma estratégia de região de confiança na qual um novo subproblema foi proposto a partir de uma nova maneira de definir o raio da região de confiança, considerada, a partir de resultados numéricos, mais eficaz que o método tradicional. A convergência (local) superlinear foi obtida sob uma condição conhecida como *error bound*, que é considerada mais fraca do que a suposição de não singularidade da matriz Jacobiana de F em uma solução do problema [34]. Contudo, assim como os algoritmos já existentes para a referida classe de problemas, para a convergência global se fez necessária essa hipótese de não singularidade.

Por outro lado, quando $\Omega \neq \mathbb{R}^n$, métodos globalmente convergentes para o problema irrestrito $F(x) = 0$ podem ser inadequados para o propósito da resolução de (1), tendo em vista serem propensos à identificação de soluções que não pertençam a Ω. Esta falha pode afetar

adversamente o desempenho de algoritmos numéricos, levando-nos a considerar estratégias e métodos específicos que busquem garantir a identificação de raízes de F sem desprezar o conjunto viável em questão.

Uma maneira conceitualmente simples de resolver sistemas não lineares restritos gerais é transformar (1) em um problema de mínimos quadrados não linear. Com este procedimento, além do aproveitamento da estrutura do problema, as restrições de caixa poderão ser incorporadas na função objetivo. Desta forma, passamos a considerar o problema equivalente definido por

$$\min_{x \in \Omega} f(x), \tag{2}$$

onde $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ é definida por $f(x) = \frac{1}{2}\|F(x)\|^2$.

Neste contexto, o alcance de um método globalmente convergente para a resolução de (2) é, em geral, efetivado pela combinação de um método local com uma estratégia de busca linear ou de região de confiança [27]. Martínez e Santos [25] destacam que estas costumam ser mais eficazes que aquelas, quando tratamos de métodos para sistemas de equações não lineares. Mais ainda, quando lidamos com sistemas sujeitos a restrições de caixa, as regiões de confiança euclidianas tendem ser menos eficazes que as regiões de confiança elípticas baseadas na abordagem afim-escala introduzida por Coleman e Li [6].

A estratégia Coleman-Li, também conhecida como afim-escala, tem como principal característica a construção de uma matriz de escalamento, diagonal e definida positiva, cujos elementos são determinados pela proximidade da iteração atual aos limites da caixa — denominada matriz afim-escala. Na prática, as regiões de confiança afim-escala são capazes de lidar com as restrições de caixa de maneira sensata, já que quando um ponto atual está distante de uma solução e próximo de um limite a referida estratégia tende forçar um grande passo no interior da região viável — um recurso que geralmente é necessário para métodos de otimização eficientes e que, neste caso, é proporcionado pelo alongamento do eixo maior da elipse que representa a região de confiança.

Nas últimas décadas foram desenvolvidos vários estudos teóricos e computacionais envolvendo a proposição de métodos globalmente convergentes para a resolução de (2) por meio de estratégias de regiões de confiança afim-escala [1–3, 12, 13, 15, 21, 23, 26, 33, 38]. Um destes foi o método STRN (*Scaled Trust-Region Newton*) desenvolvido por Bellavia, Macconi e Morini [1] para a resolução de sistemas quadrados de equações não lineares sujeitos a restrições de caixa. Este método é baseado no método clássico de regiões de confiança com passos Newton para sistemas não lineares e seu funcionamento consiste, basicamente, em minimizar, a cada iteração, um modelo quadrático Gauss-Newton sobre uma região de confiança elíptica cuja forma depende do posicionamento do iterado atual em relação aos limites da caixa. Deste modo, a única restrição do subproblema de região de confiança é uma restrição de norma 2, cuja solução não requer o cálculo de derivadas de segunda ordem. Sob condições adequadas, incluindo a não

singularidade de matriz Jacobiana, as autoras provaram que o método é globalmente convergente e a taxa de convergência para uma solução no interior de Ω é quadrática.

O método STRN foi generalizado, posteriormente, para sistemas indeterminados de equações não lineares sujeitos a restrições de caixa no trabalho de Francisco, Krejic e Martínez [12], onde a convergência quadrática para raízes de F no interior de Ω se manteve sob a condição de posto completo da matriz Jacobiana de F. Mais tarde, Kanzow e Petra [16] e Zhu [38] propuseram métodos globais com passos Levenberg-Marquardt para problemas semidiferenciáveis. Em particular, o artigo [16] está focado em problemas indeterminados que surgem de uma reformulação adequada de problemas de complementaridade mista; a convergência global é mostrada e a convergência local para raízes de F em $Int(\Omega)$ é provada sob uma suposição de *error bound*. Já em [38], uma técnica de busca linear interior não monótona é proposta a fim de garantir a convergência global; sob uma condição de *error bound*, a convergência (local) superlinear é alcançada para soluções interiores do sistema F.

Em [23, 26] métodos de região de confiança foram propostos para o problema (2), independentemente das dimensões de F. Esses procedimentos diferem no modelo quadrático usado durante as iterações: um método emprega um modelo Gauss-Newton enquanto o outro é baseado em um modelo Gauss-Newton regularizado, sendo definido como um método de Levenberg-Marquardt. Tais métodos oferecem propriedades de convergência global combinadas com convergência local potencialmente rápida para soluções do sistema. Em particular, a análise de convergência local é realizada sob a suposição de posto completo da matriz Jacobiana de F, J, para o método Gauss-Newton, e sob uma suposição de *error bound* para o método Levenberg-Marquardt [17]. Do ponto de vista teórico, a condição de *error bound* é considerada mais fraca que a condição de posto completo [30]. Morini e Porcelli [26], ao comparar algoritmos de região de confiança com modelos Gauss-Newton e Levenberg-Marquardt em sistemas não lineares sujeitos a restrições nas variáveis, concluíram que, em geral, aqueles são mais eficientes e robustos do que estes. Embora ambos tenham a vantagem de não exigir as derivadas segundas da função objetivo, o cálculo da matriz Jacobiana é exigido a cada iteração, podendo configurar, do ponto de vista computacional, um procedimento altamente custoso se considerarmos problemas de grandes dimensões em que a Jacobiana é de difícil obtenção.

Apesar do tema estar sendo discutido há mais de duas décadas, desconhecemos, até o momento, estudos sobre métodos de região de confiança baseados na abordagem afim-escala que usufruam de passos Quase-Newton. Tal fato, possivelmente, é justificado pelas dificuldades impostas pela abordagem afim-escala ao exigir o gradiente da função f, que por sua vez requer a matriz Jacobiana de F. Estes operadores são utilizados como base de decisão da direção a ser percorrida pelo método no interior da caixa, o que significa que as aproximações oferecidas pelos métodos Quase-Newton para a Jacobiana de F podem tornar o método suscetível a erros, principalmente ao trabalhar em regiões próximas dos limites da caixa.

Marini, Morini e Porcelli [24], cientes das dificuldades de relacionar métodos Quase-Newton com a abordagem afim-escala, propuseram uma estratégia de globalização que não requer o gradiente de f e nem impõe uma condição de decréscimo suficiente para f. O esquema em questão incorpora uma busca linear não monótona livre de derivadas a um método Quase-Newton baseado na projeção da função objetivo em um conjunto de restrição convexa. Os resultados reportados mostraram que a utilização de matrizes de atualização Quase-Newton no método não o tornaram suficientemente eficiente e robusto quando comparado com métodos com abordagem afim-escala; em contrapartida, a utilização da matriz Jacobiana obtida por Diferenças Finitas (metodologia Newton) no método mostrou-se mais vantajosa e competitiva.

Essas observações nos motivam a apresentar um método de região de confiança escalada baseado na abordagem afim-escala, que evite computar a matriz Jacobiana a cada iteração e que supere, na medida do possível, os empecilhos envolvidos numa combinação com uma perspectiva Quase-Newton. Adotamos a definição de métodos Quase-Newton dada por Dennis e Schnabel [8], a fim de indicar um procedimento em que, dado um iterado corrente x_k, o passo corrente p é obtido a partir da resolução do sistema linear

$$B_k p = -F(x_k),$$

onde $B_k \in \mathbb{R}^{n\times n}$ denota a matriz que aproxima a Jacobiana de F no ponto x_k. Desta forma, o modelo quadrático a ser considerado na estratégia de região de confiança será o modelo Quase-Newton dado por

$$\begin{aligned} m_k^{QN}(p) &= \tfrac{1}{2}\|F(x_k) + B_k p\|^2, \\ &= \tfrac{1}{2}\|F(x_k)\|^2 + F(x_k)^T B_k p + \tfrac{1}{2} p^T B_k^T B_k p. \end{aligned} \tag{3}$$

Com a definição do passo Quase-Newton e do modelo quadrático a ser utilizado, estamos aptos a descrever brevemente o funcionamento do algoritmo a ser proposto neste estudo.

Dada a iteração interna atual x_k, a região de confiança escalada com raio $\Delta > 0$ é definida e o passo Quase-Newton é calculado. Se este passo estiver dentro da região de confiança, será tomado como solução do subproblema de minimização do modelo (3); caso contrário será calculado e tomado como solução o passo Dogleg [27]. A redução prevista do modelo quadrático e a redução real da função objetivo proporcionadas pelo passo escolhido são calculadas. Se a redução real for grande o suficiente quando comparada à redução prevista, um múltiplo viável deste passo é então calculado e somado ao iterado atual x_k, gerando um ponto de teste. Se a redução real gerada pelo passo viável for suficientemente grande em relação à redução real do passo (inteiro), o ponto de teste será aceito e uma nova iteração será iniciada. Caso contrário, o raio da região de confiança será reduzido e repetir-se-ão os procedimentos para identificar um passo que satisfaça as condições de redução real da função objetivo. A filosofia do método

consiste em usar a direção Quase-Newton o mais frequente possível, aproveitando suas boas propriedades de convergência local [8].

Para a resolução do subproblema da região de confiança, buscaremos conciliar as estratégias de definição do raio de região de confiança utilizada por Yuan e Wei [35] e por Zeng e Fu [36]. Aqueles, propuseram um raio de região de confiança dependente da variável $\|F(x_k)\|$, sendo definido a cada iteração por $\Delta_k = c^t\|F(x_k)\|$, onde $c \in (0,1)$ e t é um número inteiro não negativo; já estas, definiram o raio simplesmente por $\Delta_k = c^t$. Em ambas técnicas, quando o método exige que seja reduzido o tamanho da região de confiança aumentamos o valor de t em uma unidade.

O raio de região de confiança utilizado por Yuan e Wei [35] mostra-se vantajoso, principalmente, pelos resultados teóricos que proporciona e por possibilitar passos longos quando o método usufruir de um bom modelo da função e estiver distante de uma solução. Em contrapartida, ao chegar próximo de uma solução o raio pode acabar ficando muito pequeno, a ponto de retardar a convergência do método com passos muito curtos. Este fenômeno não costuma ocorrer com o raio proposto por Zeng e Fu [36], já que não depende de $\|F(x_k)\|$, e isto caracteriza uma das principais vantagens de sua utilização; por outro lado, mesmo que o modelo represente a função de forma satisfatória, não é possibilitada a expansão do raio de região de confiança para além da unidade. Tais aspectos, nos motivam a utilizar, a cada iteração do nosso método, o raio de região de confiança definido por $\Delta_k = c^t\eta_k$, em que $c \in (0,1)$, t é um número inteiro não negativo e $\eta_k = \min\{10^2, \max\{1, \|F(x_k)\|\}\}$.

Experimentos numéricos em um grande número de problemas-teste usados frequentemente na literatura serão realizados neste estudo, a fim de comparar o desempenho de nosso método (Quase-Newton) com um método Gauss-Newton e outro Levenberg-Marquardt. Para aquele, utilizaremos o método STRN, citado anteriormente; já para este, utilizaremos o método PLM (*Projected Levenberg-Marquardt*), desenvolvido por Kanzow, Yamashita e Fukushima [17]. O método PLM é aplicado na resolução de sistemas não lineares sujeitos a restrições convexas. Um primeiro método introduzido em [17] exige a resolução de subproblemas bastante difíceis (mesmo para o caso de restrições de caixa), mas o segundo método, que é o que usaremos para comparações, tem uma implementação consideravelmente simples. Neste método um passo Levenberg-Marquardt é calculado em cada iteração e, em seguida, projetado no conjunto viável. Se a norma do sistema no ponto projetado for suficientemente pequena com relação à norma do sistema no ponto atual, o ponto Levenberg-Marquardt projetado é tomado como nova iteração. Se isto não acontecer e o ponto Levenberg-Marquardt projetado gerar uma direção de descida, essa direção é usada para uma busca linear. Caso contrário, o método usa um procedimento de gradiente projetado.

A fim de facilitar a compreensão dos conhecimentos teórico-práticos que estamos propondo, além da Introdução e das Considerações Finais, o presente estudo está organizado em

três partes. No Capítulo 2 serão descritos os métodos clássicos de Newton e Quase-Newton para sistemas de equações não lineares, dando destaque a sua fundamentação e aos resultados relativos à convergência dos referidos métodos. Considerando que estes métodos convergem apenas localmente, destinaremos uma seção na qual abordaremos os métodos de região de confiança como forma de globalizar a convergência. No Capítulo 3 apresentaremos um método de região de confiança escalada Quase-Newton para sistemas quadrados de equações não lineares sujeitos a restrições de caixa, a ser denominado STRQN (*Scaled Trust-Region Quasi-Newton*). Os respectivos resultados de convergência serão demonstrados, dando ênfase à convergência superlinear da sequência gerada pelo método a soluções do sistema no interior da caixa. Por fim, no Capítulo 4 apresentaremos os resultados numéricos decorrentes da utilização do método proposto em uma classe de sistemas não lineares quadrados sujeitos a restrições de caixa, e compararemos estes dados com o desempenho dos métodos STRN e PLM.

Em relação à notação matemática utilizada neste trabalho, o subscrito k será sempre usado como índice para uma sequência. Assim, teremos as seguintes equivalências entre aplicações: $F_k \equiv F(x_k)$, $D_k \equiv D(x_k)$, $J_k \equiv J(x_k)$, $\nabla f_k \equiv \nabla f(x_k)$. Em todos os casos, $\|\cdot\|$ denota a norma vetorial Euclidiana ou sua respectiva norma matricial induzida. Para uma caixa Ω denotaremos o conjunto de seus pontos interiores por $Int(\Omega)$, e, para todo $z \in \mathbb{R}^n$, denotaremos $P_\Omega(z)$ a projeção Euclidiana de z em Ω. Uma bola euclidiana centrada em x e com raio δ será denotada por $\mathscr{B}(x,\delta)$. Os operadores Gradiente e Hessiana de uma função f serão denotados, respectivamente, por ∇f e $\nabla^2 f$. A matriz Identidade de ordem n será denotada por I.

2 SISTEMAS QUADRADOS DE EQUAÇÕES NÃO LINEARES

A identificação de pontos cujas coordenadas satisfaçam uma série de relações matemáticas consiste em um dos grandes desafios dos problemas de Programação Não Linear. Quando essas relações assumem a forma de n igualdades, em que n denota o número de variáveis envolvidas, e pelo menos uma delas tem grau maior que um, dizemos que o problema consiste em resolver um sistema quadrado de equações não lineares. Escrevemos este problema, matematicamente, como

$$F(x) = 0, \tag{4}$$

onde $x \in \mathbb{R}^n$ e $F : \mathbb{R}^n \to \mathbb{R}^n$ é uma função vetorial da forma

$$F(x) = \begin{pmatrix} F_1(x) \\ F_2(x) \\ \vdots \\ F_n(x) \end{pmatrix},$$

em que $F_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, n$.

Para fins de análise teórica, assumiremos neste estudo que F é de classe C^1, isto é, que cada uma de suas componentes — as F_i's — é continuamente diferenciável em $\mathbb{R}^n$. Neste contexto, denotaremos a matriz Jacobiana de F por J, onde J é dada por

$$J(x) = \begin{pmatrix} \nabla F_1^T(x) \\ \nabla F_2^T(x) \\ \vdots \\ \nabla F_n^T(x) \end{pmatrix},$$

para $x \in \mathbb{R}^n$.

A partir das definições supracitadas, iniciaremos as discussões do capítulo destacando fundamentos e propriedades de convergência dos métodos clássicos para resolução de sistemas quadrados de equações não lineares: o método de Newton e os métodos Quase-Newton.

2.1 MÉTODOS CLÁSSICOS

Os métodos de Newton e Quase-Newton constituem, atualmente, os principais meios de resolução de sistemas de equações não lineares. Embora sejam métodos com propriedades boas de convergência, quando levamos em conta sua utilização para a resolução de (4), eles não são classificados como métodos globais, mas sim como métodos locais.

Um *método local* é caracterizado por ser um processo iterativo na qual a garantia de convergência da sequência gerada a uma solução do problema depende da aproximação inicial

estar suficientemente próxima de uma solução particular. Conforme destacado por Martínez e Santos [25], felizmente, em muitos casos práticos, o domínio de convergência de métodos locais é grande. No entanto, quando a estimativa inicial da solução é excessivamente ruim, os métodos locais devem ser modificados a fim de melhorar suas propriedades de convergência global.

Em contrapartida, um *método global* é aquele que, independentemente da aproximação inicial escolhida, garante que pelo menos um ponto-limite da sequência gerada pelo método é uma solução ou, pelo menos, um ponto estacionário de um problema de otimização. Em geral, métodos globais são modificações de métodos locais que preservam as propriedades de convergência local do algoritmo original.

Estas definições iniciais nos motivam a discutir a estruturação e as propriedades de convergência local do método de Newton e dos Quase-Newton. Antes, porém, faz-se necessário apresentar um variante multidimensional do teorema de Taylor, que oferecerá embasamento aos métodos a serem explorados nesta seção.

Teorema 2.1.1. *Suponha que $F : \mathbb{R}^n \to \mathbb{R}^n$ é continuamente diferenciável em um conjunto aberto e convexo Λ, e que x e $x+p$ são vetores em Λ. Então,*

$$F(x+p) = F(x) + \int_0^1 J(x+tp)p \, \mathrm{d}t. \tag{5}$$

Demonstração. Disponível em Dennis e Schnabel [8]. □

Este resultado permitirá a construção de um modelo que aproxime a função objetivo. Os detalhes serão aprofundados na subseção a seguir.

2.1.1 Método de Newton

O método de Newton é um dos métodos numéricos mais populares para a resolução de (4) [8]. Sua essência na resolução de equações não lineares consiste em considerar, a cada iteração, um modelo linear da função objetivo, envolvendo o valor da função e de suas derivadas de primeira ordem. Este modelo é baseado na aproximação de Taylor de primeira ordem (5), numa vizinhança do iterado atual x_k, sendo definido por

$$M_k(x) = F_k + J_k(x - x_k). \tag{6}$$

A partir da definição do modelo (6) podemos concluir que

$$M_k(x_k + p) = F_k + J_k p.$$

Consideramos, portanto, que o termo $J_k p$ aproxima a integral contida em (5), o que faz valer $M_k(x_k + p) \approx F(x_k + p)$.

Sabemos que a resolução de um sistema não linear quadrado irrestrito requer identificar as raízes de F em $\mathbb{R}^n$, e que esta é uma tarefa complexa de ser efetivada diretamente. Pensando nisto, o método de Newton propõe que consideremos um meio de resolução mais simples: procurar, a cada iteração, a raiz de $M_k(x_k+p)$, a fim de aproximar-se cada vez mais da solução desejada.

Desta forma, dado o iterado x_k, consideramos o ponto seguinte x_{k+1} como uma solução de $M_k(x_k+p)=0$. Portanto, se J_k é não singular, $M_k(x_k+p)=0$ tem solução única e o método terá que resolver, nesta iteração, o sistema linear

$$J_k p_k^N = -F_k, \tag{7}$$

e tomar $x_{k+1} = x_k + p_k^N$. Evidentemente, se J_k for não singular,

$$p_k^N = -J_k^{-1} F_k.$$

Em seguida calculamos F_{k+1} e J_{k+1}, e repetimos o mesmo procedimento para a iteração $k+1$. Este processo iterativo se repete até que seja identificada uma raiz de F, a ser denotada por x^*.

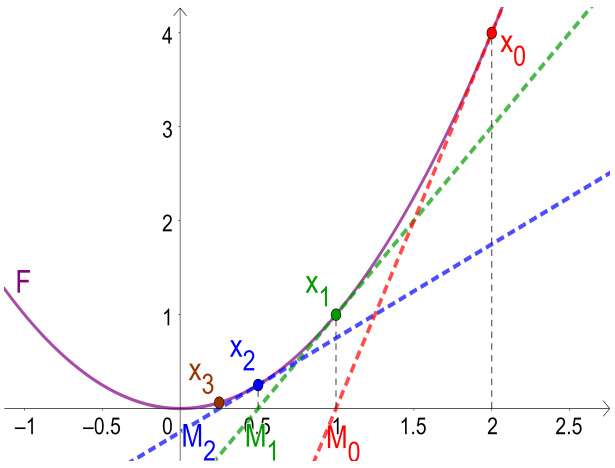

Figura 1 – Primeiras iterações do método de Newton para a função $F(x) = x^2$, partindo do ponto inicial $x_0 = 2$.

O algoritmo a seguir sintetiza todos procedimentos supracitados.

Algoritmo 1: Método de Newton para Equações Não Lineares

Entrada: $x_0 \in \mathbb{R}^n$;
Saída: $x^* \in \mathbb{R}^n$;
início
  $k \longleftarrow 0$;
  Calcule F_k;
  Calcule J_k;
  repita
    1) Obtenha p_k^N resolvendo o sistema

$$J_k p_k^N = -F_k;$$

    2) $x_{k+1} \longleftarrow x_k + p_k^N$;
    3) Calcule F_{k+1};
    4) Calcule J_{k+1};
    5) $k \longleftarrow k+1$;
  até $F_k = 0$;
fim
$x^* \longleftarrow x_k$.

A estrutura e a fundamentação do método de Newton já apontam algumas de suas desvantagens. Primeiramente, a exigência de derivadas de primeira ordem no cálculo da matriz Jacobiana de F torna o método mais suscetível a falhas, principalmente quando estas derivadas (analíticas) não são fáceis de serem obtidas. Recorrer a métodos numéricos específicos de diferenciação, como por exemplo o método de Diferenças Finitas [20], tende a aumentar a carga de trabalho e tornar o método ineficiente em alguns problemas de médio e grande porte.

Quando a matriz Jacobiana de F for não singular, a resolução do sistema linear (7) pode ser obtida via fatoração LU [14], com um custo computacional da ordem de $O\left(\frac{n^3}{3}\right)$ operações [25]. Acrescido das avaliações de F e J em x_k, o trabalho realizado em uma iteração acaba, assim, crescendo de forma súbita com o aumento da dimensão do problema, podendo tornar inviável sua resolução.

Por outro lado, se a matriz Jacobiana de F for singular ao longo do processo iterativo, teremos um provável comportamento inadequado do método de Newton em virtude do mau condicionamento do sistema (7). Nesta perspectiva, o sistema tende a ter um alto grau de sensibilidade, de modo que pequenas variações nos seus dados de entrada (os coeficientes da matriz Jacobiana) acarretam grandes alterações para a sua solução. Isto significa que a solução que será encontrada está mais suscetível a erros, como pode ser visto com maiores detalhes na obra de Golub e Loan [14].

Quando a singularidade ocorre para $J(x^*)$ o método de Newton tende a ter uma convergência lenta à raiz x^*. Como exemplo deste fato, podemos citar a função $F : \mathbb{R} \to \mathbb{R}$ dada por $F(x) = x^2$, representada na Figura 1. Neste caso dado um ponto inicial não nulo, o Algoritmo 1 gera a sequência de iterados

$$x_k = \frac{1}{2^k} x_0,$$

que converge para $x^* = 0$ apenas com uma taxa linear.

Em contrapartida, quando $J(x^*)$ for não singular e o ponto inicial estiver suficientemente próximo de x^*, o método de Newton mostra-se atrativo por apresentar uma taxa de convergência quadrática. Analisaremos tal fato com maiores detalhes posteriormente na subseção 2.2.1.

Outra vantagem do método de Newton é trabalhar bem em problemas que se aproximem de problemas lineares, já que, em casos de não haver singularidades, ele identifica a raiz de uma função afim em uma única iteração. Além disto, se alguma função componente de F for linear, cada iteração do método de Newton será uma solução para esta equação, já que o modelo utilizado pelo método de Newton é linear e, portanto, sempre fornecerá soluções exatas para este tipo de funções.

Veremos a seguir uma variante do método de Newton que busca superar os problemas citados anteriormente sem grandes perdas nas boas propriedades de convergência.

2.1.2 Métodos Quase-Newton

Na última subseção examinamos o método de Newton e verificamos que sua estruturação torna o método caro do ponto de vista computacional, principalmente devido à necessidade de calcular $J(x)$ a cada iteração. É natural pensarmos, então, em métodos que requeiram um custo computacional menor, mantendo, sempre que possível, as boas propriedades de convergência do método de Newton. Este é o cerne dos métodos Quase-Newton, cujo processo iterativo é caracterizado por formar uma sequência $\{x_k\}_{k\in\mathbb{N}}$ no qual cada novo iterado é gerado pela fórmula $x_{k+1} = x_k + p_k^{QN}$, em que p_k^{QN} é a solução do sistema

$$B_k p_k^{QN} = -F_k, \tag{8}$$

$B_k \in \mathbb{R}^{n\times n}$. A ideia central aqui é substituir a matriz J_k, utilizada no método de Newton, por uma aproximação B_k, preferencialmente mais fácil de ser obtida. Evidentemente, se B_k for não singular, teremos

$$p_k^{QN} = -B_k^{-1} F_k.$$

É nítido que se $B_k = J_k$ para toda iteração, teremos o método de Newton. Isto nos leva a concluir que o método de Newton pode ser considerado, também, um método Quase-Newton.

Um dos métodos Quase-Newton mais simples é o método de Newton Estacionário, na qual fixamos $B_k \equiv J_0$. Uma variação deste método é a incorporação de recomeços a cada rodada

de uma quantidade específica de iterações. Fixado um número inteiro m, se k é múltiplo de m, tomamos $B_k = J_k$; caso contrário, consideramos $B_k = B_{k-1}$. O método de Newton Estacionário pode ser visto com maiores detalhes nos estudos de Martínez e Santos [25] e de Shamanski [32], onde, inclusive, são feitas análises sobre o parâmetro m ótimo para alguns problemas específicos.

Outra ramificação importante dos métodos Quase-Newton são os métodos secantes. De forma similar ao método de Newton, consideramos, a cada iteração, o modelo linear

$$M_k(x) = F_k + B_k(x - x_k).$$

Desta forma, na iteração $k+1$, teremos

$$M_{k+1}(x) = F_{k+1} + B_{k+1}(x - x_{k+1}).$$

A ideia secante consiste, assim, em impor que M_{k+1} interpole a função F nos pontos x_k e x_{k+1}, ou seja, que

$$M_{k+1}(x_{k+1}) = F_{k+1} \qquad \text{e} \qquad M_{k+1}(x_k) = F_k.$$

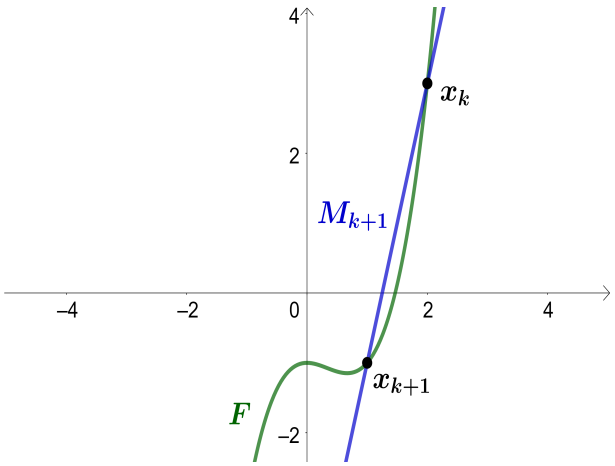

Figura 2 – Aproximação (local) da função não linear $F(x) = x^3 - x^2 - 1$ pelo modelo linear $M_{k+1}(x) = 4x - 5$.

Perceba que a primeira condição acima é satisfeita automaticamente pela definição de M_{k+1}, já que

$$\begin{aligned} M_{k+1}(x_{k+1}) &= F_{k+1} + B_{k+1}(x_{k+1} - x_{k+1}) \\ &= F_{k+1}. \end{aligned}$$

Resta-nos, portanto, exigir que seja satisfeita a condição

$$F_{k+1} + B_{k+1}(x_k - x_{k+1}) = F_k,$$

isto é, que seja satisfeita a equação secante

$$B_{k+1}p_k = y_k, \tag{9}$$

onde $p_k = x_k - x_{k+1}$ e $y_k = F_k - F_{k+1}$.

A análise da quantidade de matrizes que satisfazem (9) pode ser feita considerando a referida equação como um sistema linear cuja incógnita é a matriz B_{k+1}. Assim, o sistema possui n^2 variáveis (as entradas de B_{k+1}) e apenas n equações, o que implica que a possibilidade de solução única é restrita ao caso em que $n = 1$. Em contrapartida, nos casos em que $n > 1$ e $p_k \neq 0$, existirão infinitas soluções $B \in \mathbb{R}^{n\times n}$ que satisfazem $Bp_k = y_k$.

Para cada escolha da matriz B teremos um método secante específico. Nesta perspectiva, destacamos três métodos secantes muito utilizados em sistemas quadrados de equações não lineares: o método Broyden, Fletcher, Goldfarb e Shanno (BFGS), o método de Atualização Simétrica de Posto Um (*Simetric Rank One* ou SR1) e o primeiro método de Broyden. Nestes dois últimos, a escolha de B_{k+1} parte do princípio de que a diferença $\Delta B_k = B_{k+1} - B_k$ seja uma matriz de posto unitário. Assim, segue de (9) que

$$\Delta B_k p_k = y_k - B_k p_k.$$

Tomando $w_k \in \mathbb{R}^n$ não ortogonal a p_k, verificamos que

$$\Delta B_k = \frac{(y_k - B_k p_k)w_k^T}{w_k^T p_k},$$

ou seja,

$$B_{k+1} = B_k + \frac{(y_k - B_k p_k)w_k^T}{w_k^T p_k}. \tag{10}$$

A escolha $w_k = p_k$ define o primeiro método de Broyden, enquanto a escolha $w_k = y_k - B_k p_k$ define o método SR1. Para que a atualização (10) esteja bem definida e possa ser utilizada é comum exigirmos que seja satisfeita a condição $\left|w_k^T p_k\right| > \gamma \|w_k\| \|p_k\|$, para algum $\gamma > 0$. Nestes métodos a matriz inicial B_0 deve ser não singular — exigência mais fraca que o método BFGS, que impõe que o processo iterativo seja iniciado com uma matriz definida positiva.

Já no método BFGS a escolha de B_{k+1} parte do princípio de que esta seja a matriz simétrica definida positiva que satisfaz a equação secante (9), e que seja a mais próxima da matriz corrente B_k, no sentido de satisfazer o problema

$$\min_B \|B - B_k\| \text{ sujeito a } B = B^T, \ Bs_k = y_k.$$

A solução deste problema é dada pela fórmula a seguir

$$B_{k+1} = B_k - \frac{B_k p_k p_k^T B_k}{p_k^T B_k p_k} + \frac{y_k y_k^T}{y_k^T p_k},$$

e pode ser vista com maiores detalhes na obra de Nocedal e Wright [27].

Perceba que pré-multiplicando a equação (9) por p_k^T, temos que

$$p_k^T B_{k+1} p_k = p_k^T y_k.$$

Logo, almejando preservar a característica de B_{k+1} ser definida positiva e estar bem definida, a cada iteração o método BFGS impõe que seja satisfeita a condição de curvatura

$$p_k^T y_k > 0.$$

A estruturação dos métodos Quase-Newton é similar a do método de Newton. Inicialmente resolvemos o sistema linear (8) para identificar p_k^{QN}. Em seguida, tomamos $x_{k+1} = x_k + p_k^{QN}$, calculamos F_{k+1} e B_{k+1}, e refazemos o mesmo procedimento para a iteração $k+1$. Este processo iterativo se repete até que seja identificada uma raiz de F, como evidenciado no algoritmo abaixo.

Algoritmo 2: Método Quase-Newton para Equações Não Lineares

Entrada: $x_0 \in \mathbb{R}^n$ e $B_0 \in \mathbb{R}^{n\times n}$;
Saída: $x^* \in \mathbb{R}^n$;
início
  $k \longleftarrow 0$;
  Calcule F_k;
  repita
    1) Obtenha p_k^{QN} resolvendo o sistema

$$B_k p_k^{QN} = -F_k;$$

    2) $x_{k+1} \longleftarrow x_k + p_k^{QN}$;
    3) Calcule F_{k+1};
    4) Calcule B_{k+1};
    5) $k \longleftarrow k+1$;
  até $F_k = 0$;
fim
$x^* \longleftarrow x_k$.

Embora a substituição da matriz J por uma aproximação B tenda a melhorar a eficiência do algoritmo, os métodos Quase-Newton também são apenas localmente convergentes. Outra desvantagem herdada do método de Newton é a necessidade de resolução de um sistema linear a cada iteração. Esta barreira, porém, pode ser superada nos métodos secantes através de atualizações das matrizes B_k^{-1}, cujas fórmulas estão explicitadas para cada método secante aqui abordado em Nocedal e Wright [27]. Desta forma, ao invés de resolver (8), atualizamos B_k^{-1} a partir de informações coletadas na iteração anterior e tomamos

$$p_k^{QN} = -B_k^{-1} F_k. \tag{11}$$

Outra alternativa para aliviar a carga de trabalho de métodos secantes é obter B_k a partir de uma estratégia conhecida como memória limitada [27]. Os métodos secantes que usufruem de memória limitada buscam obter aproximações para as matrizes B_k a partir do armazenamento de alguns vetores de dimensão n, em que n é o número de variáveis do problema. Desta forma é evitado o armazenamento de matrizes densas $n \times n$, o que pode demandar um esforço computacional muito alto quando n for grande. Em contrapartida, tais métodos costumam oferecer uma taxa de convergência mais lenta (geralmente linear).

Especificando-se aos métodos secantes de posto um, Ortega e Rheinboldt [29] exploram uma estratégia de memória limitada a partir da fórmula de Sherman-Morison [14] para obter uma equação de atualização de B_k^{-1} a partir do armazenamento de alguns vetores relativos às informações adquiridas nas últimas iterações. Desta forma, evitamos resolver o sistema linear (8) com técnicas trabalhosas, e tomamos apenas como solução o produto (11).

Além dos métodos clássicos de Newton e Quase-Newton, existem outros métodos para resolver sistemas não lineares de equações. Embora possam ser considerados menos utilizados na Otimização, seu uso pode ser vantajoso quando consideramos problemas com características específicas.

Em alguns problemas de equações não lineares, por exemplo, a matriz Jacobiana de F é facilmente obtida. Ainda assim, usufruir de um método direto para obter p através da resolução do sistema linear (7) pode ser um meio altamente custoso do ponto de vista computacional, principalmente quando nos reportamos a problemas de grandes dimensões. Nestes casos, métodos de Newton Inexatos são uma alternativa atrativa a ser considerada, já que evitam resolver (7) exatamente. A cada iteração de um método de Newton Inexato, são utilizadas direções de busca p_k que satisfazem a condição

$$\|F_k + J_k p_k\| \leq \gamma_k \|F_k\|, \tag{12}$$

para algum $\gamma_k \in [0, \gamma]$, onde $\gamma \in [0, 1)$. A teoria de convergência para esses métodos não depende da técnica utilizada para calcular p_k, mas sim da condição (12). Maiores detalhes sobre a implementação e a convergência superlinear do método podem ser consultados nos estudos de Nocedal e Wright [27].

Por outro lado, existem problemas cujo comportamento não linear da função F não permite uma representação adequada pelos modelos lineares utilizados para os métodos de Newton e Quase-Newton. Pensando nisto, os métodos Tensores são caracterizados por modificar o modelo linear $M_k(x)$, acrescentando um termo extra que visa capturar alguns dos comportamentos não lineares de ordens superiores da função F. A expectativa é de que esta alteração provoque uma convergência mais rápida e confiável para identificar as raízes x^* procuradas, inclusive para as quais $J(x^*)$ não tenha posto completo. Maiores informações relativas a estes métodos podem ser vistas nos estudos de Schnabel e Frank [31].

Tendo em vista nosso interesse nos métodos de Newton e Quase-Newton, restringiremos

a análise de convergência para estes dois métodos.

2.2 ANÁLISE DE CONVERGÊNCIA

Um dos principais indicadores de eficiência de um método iterativo é sua taxa de convergência a uma solução ou a um ponto estacionário do problema em questão. Por meio dela quantificamos a rapidez com que uma determinada sequência convergente, gerada pelo método, se aproxima de seu limite. Para tanto, verificamos a velocidade na redução da distância $\|x_{k+1}-x^*\|$ em relação a $\|x_k-x^*\|$, e dizemos que a sequência tem convergência:

1. *Linear:* Se existem $k_0 \in \mathbb{N}$ e $s \in (0,1)$ tais que, para todo $k \geq k_0$,
$$\|x_{k+1}-x^*\| \leq s\|x_k-x^*\|.$$
2. *Superlinear:* Se existe uma sequência s_k tendendo a zero, quando $k \to \infty$, tal que
$$\|x_{k+1}-x^*\| \leq s_k\|x_k-x^*\|,$$
para todo $k \in \{0,1,2,\dots\}$.
3. *Quadrática:* Se existem $k_0 \in \mathbb{N}$ e $\tilde{s} > 0$ tais que, para todo $k \geq k_0$,
$$\|x_{k+1}-x^*\| \leq \tilde{s}\|x_k-x^*\|^2.$$

De modo mais geral, se $\{x_k\}$ converge para x^*, dizemos que essa convergência tem ordem $q+1$ se existirem $k_0 \in \mathbb{N}$, $\tilde{s} > 0$ e $q > 0$ tais que, para todo $k \geq k_0$,
$$\|x_{k+1}-x^*\| \leq \tilde{s}\|x_k-x^*\|^{q+1}.$$
Nesta perspectiva, quanto maior o valor de q mais rápida será a convergência local.

A fim de analisar, de modo geral, a convergência dos métodos Quase-Newton e, especificamente, do próprio método de Newton, veremos, inicialmente, que se o ponto inicial x_0 e todas as matrizes B_k estiverem suficientemente próximos, respectivamente, de uma raiz x^* e de $J(x^*)$, a sequência gerada por $x_{k+1} = x_k - B_k^{-1}F_k$ converge linearmente para x^*. Para tanto, assumiremos que $J(x^*)$ e todas as matrizes que se encontram numa certa vizinhança da referida matriz são não singulares. O Lema 2.2.2, a ser visto mais adiante, mostrará com precisão o tamanho desta vizinhança, porém antes necessitamos de um resultado auxiliar clássico: o Lema de Banach. Ambos resultados teóricos são, também, encontrados nas obras de Martínez e Santos [25] e de Ortega e Rheinboldt [29].

Lema 2.2.1 (Lema de Banach)**.** *Seja $\|\cdot\|$ uma norma qualquer em $\mathbb{R}^n$, que denota também a norma matricial subordinada. Seja $A \in \mathbb{R}^{n\times n}$ uma matriz com $\|A\| < 1$, então $I+A$ é não singular e*
$$\frac{1}{1+\|A\|} \leq \left\|(I+A)^{-1}\right\| \leq \frac{1}{1-\|A\|}.$$

Demonstração. Assuma, para obter contradição, que $I+A$ é singular. Assim, existe $x \in \mathbb{R}^n$ não nulo tal que $(I+A)x = 0$, isto é,$-x = Ax$. Logo,

$$\|x\| \leq \|A\|\|x\|,$$

ou seja, $\|A\| \geq 1$. Isto contradiz a hipótese de que $\|A\| < 1$. Portanto, $I+A$ é inversível.

Tome $B = (I+A)^{-1}$. Desta forma, $B(I+A) = I$ e, portanto,

$$1 = \|B(I+A)\| \leq \|B\|\|I+A\| \leq \|B\|(1+\|A\|),$$

isto é,

$$\frac{1}{1+\|A\|} \leq \left\|(I+A)^{-1}\right\|. \tag{13}$$

Além disto, temos que $B = I - BA$. Assim, $\|B\| \leq 1 + \|B\|\|A\|$, ou seja, $\|B\|(1-\|A\|) \leq 1$. Logo,

$$\left\|(I+A)^{-1}\right\| \leq \frac{1}{1-\|A\|}. \tag{14}$$

De (13) e de (14) concluímos a prova. □

Lema 2.2.2. *Se $B \in \mathbb{R}^{n\times n}$ é tal que $\|B - J(x^*)\| \leq \frac{1}{2\|J(x^*)^{-1}\|}$ então B^{-1} existe e satisfaz $\left\|B^{-1}\right\| \leq 2\left\|J(x^*)^{-1}\right\|$.*

Demonstração. Seja $A = BJ(x^*)^{-1} - I = [B - J(x^*)]J(x^*)^{-1}$. Das propriedades de normas e das hipóteses do lema, temos, então, que

$$\|A\| = \left\|[B-J(x^*)]J(x^*)^{-1}\right\| \leq \|B-J(x^*)\|\left\|J(x^*)^{-1}\right\| \leq \frac{1}{2} < 1.$$

Desta forma, pelo Lema 2.2.1, temos que $BJ(x^*)^{-1}$ é não singular. Por consequência, existe B^{-1}, vale $[BJ(x^*)^{-1}]^{-1} = J(x^*)B^{-1}$ e

$$\left\|J(x^*)B^{-1}\right\| \leq \frac{1}{1-\left\|BJ(x^*)^{-1}-I\right\|} \leq \frac{1}{1-\frac{1}{2}} = 2.$$

Como $\left\|B^{-1}\right\| = \left\|J(x^*)^{-1}J(x^*)B^{-1}\right\| \leq \left\|J(x^*)^{-1}\right\|\left\|J(x^*)B^{-1}\right\|$, concluímos, por fim, que $\left\|B^{-1}\right\| \leq 2\left\|J(x^*)^{-1}\right\|$. □

Com tais resultados, conseguimos mostrar a convergência linear da sequência supracitada a ser demonstrada nos próximos dois resultados. Destacamos que ambos são encontrados na obra de Martínez e Santos [25].

Lema 2.2.3 (Lema das Duas Vizinhanças)**.** *Para cada $x \in \Omega$ e $B \in \mathbb{R}^{n\times n}$, definimos a função $\Phi(x,B) = x - B^{-1}F(x)$. Seja $r \in (0,1)$. Então existem $\varepsilon_1 = \varepsilon_1(r)$, $\delta_1 = \delta_1(r) > 0$ tais que se $\|x-x^*\| \leq \varepsilon_1$, $\|B-J(x^*)\| \leq \delta_1$ a função $\Phi(x,B)$ está bem definida e satisfaz $\|\Phi(x,B)-x^*\| \leq r\|x-x^*\|$.*

Demonstração. Seja $\delta_1' = \frac{1}{2\|J(x^*)^{-1}\|}$.

Pelo Lema 2.2.2, se $\|B - J(x^*)\| \leq \delta_1'$, então B^{-1} existe e satisfaz

$$\left\|B^{-1}\right\| \leq 2\left\|J(x^*)^{-1}\right\|. \tag{15}$$

Logo, basta tomar $x \in \Omega$ e $\delta_1 \leq \delta_1'$ para que $\Phi(x,B)$ esteja bem definida.

Defina

$$A_1 = \left\|x - x^* - B^{-1}J(x^*)(x-x^*)\right\| \qquad \text{e} \qquad A_2 = \left\|B^{-1}\left[F(x) - J(x^*)(x-x^*)\right]\right\|.$$

Perceba que

$$\begin{aligned}
\|\Phi(x,B) - x^*\| &= \left\|x - B^{-1}F(x) - x^*\right\| \\
&= \left\|x - x^* - B^{-1}\left[J(x^*)(x-x^*) - J(x^*)(x-x^*) + F(x)\right]\right\| \\
&\leq \left\|x - x^* - B^{-1}J(x^*)(x-x^*)\right\| + \left\|B^{-1}\left[F(x) - J(x^*)(x-x^*)\right]\right\| \\
&= A_1 + A_2.
\end{aligned} \tag{16}$$

De (15), sabemos que

$$\begin{aligned}
A_1 &= \left\|x - x^* - B^{-1}J(x^*)(x-x^*) - B^{-1}B(x-x^*) + B^{-1}B(x-x^*)\right\| \\
&= \left\|x - x^* - B^{-1}B(x-x^*) + B^{-1}\left[B - J(x^*)\right](x-x^*)\right\| \\
&= \left\|B^{-1}\left[B - J(x^*)\right](x-x^*)\right\| \\
&\leq \left\|B^{-1}\right\| \|B - J(x^*)\| \|x-x^*\| \\
&\leq 2\left\|J(x^*)^{-1}\right\| \delta_1 \|x-x^*\|.
\end{aligned} \tag{17}$$

Além disso, pela diferenciabilidade de F e por (15),

$$A_2 \leq \left\|B^{-1}\right\| \|F(x) - J(x^*)(x-x^*)\| \leq 2\left\|J(x^*)^{-1}\right\| \beta(x), \tag{18}$$

onde $\lim_{x \to x^*} \frac{\beta(x)}{\|x-x^*\|} = 0$.

Seja ε_1 tal que

$$2\left(\delta_1 + \sup_{\|x-x^*\| \leq \varepsilon_1} \left\{\frac{\beta(x)}{\|x-x^*\|}\right\}\right) \leq \frac{r}{\left\|J(x^*)^{-1}\right\|}. \tag{19}$$

Então, para $\|B - J(x^*)\| \leq \delta_1$ e $\|x-x^*\| \leq \varepsilon_1$, segue de (16) – (19) que

$$\begin{aligned}
\|\Phi(x,B) - x^*\| &\leq 2\left\|J(x^*)^{-1}\right\| \delta_1 \|x-x^*\| + 2\left\|J(x^*)^{-1}\right\| \beta(x) \\
&= 2\left\|J(x^*)^{-1}\right\| \left(\delta_1 + \frac{\beta(x)}{\|x-x^*\|}\right) \|x-x^*\| \\
&\leq r\|x-x^*\|,
\end{aligned}$$

como queríamos demonstrar. □

Teorema 2.2.4 (Teorema das Duas Vizinhanças)**.** *Seja $r \in (0,1)$. Então existem $\varepsilon = \varepsilon(r)$ e $\delta = \delta(r) > 0$ tais que se $\|x_0 - x^*\| \leq \varepsilon$ e $\|B_k - J(x^*)\| \leq \delta$ para todo k, então a sequência gerada por $x_{k+1} = x_k - B_k^{-1}F_k$ está bem definida, converge para x^* e $\|x_{k+1} - x^*\| \leq r\|x_k - x^*\|$, para todo k.*

Demonstração. Como $\Phi(x,B) = x - B^{-1}F(x)$, temos que $x_{k+1} = \Phi(x_k, B_k)$, $k = \{0,1,2,\dots\}$. Por indução e pelo Lema 2.2.3, concluímos a tese do Teorema. □

A utilização do Teorema das Duas Vizinhanças nos métodos de Newton e Quase-Newton será fundamental para provar as respectivas velocidades de convergência. Veremos sua aplicação nos resultados que serão apresentados nas seções subsequentes, comprovando a importância de que realizemos escolhas adequadas para x_0 e B_k na convergência local.

2.2.1 Convergência do Método de Newton

As boas propriedades de convergência do método de Newton tornam-no um método muito utilizado em alguns tipos de sistemas de equações não lineares. Sob determinadas condições, o referido método, descrito pelo Algoritmo 1, atinge taxa de convergência quadrática, convergindo com agilidade para uma solução do problema em questão.

Como parte da comprovação de convergência quadrática do método de Newton, assumimos a existência de constantes $L > 0$ e $q > 0$ tal que, em uma vizinhança de x^*,

$$\|J(x) - J(x^*)\| \leq L\|x - x^*\|^q. \tag{20}$$

Quando $q = 1$ dizemos que a Jacobiana de F é localmente Lipschitz contínua. Ressaltamos que uma função Lipschitz contínua configura um critério de suavidade mais forte que a condição de continuidade uniforme (logo, de continuidade), podendo ser interpretada como uma condição intermediária entre a continuidade e a diferenciabilidade de uma função [22].

A partir de (20), conseguimos provar dois resultados importantes para as análises de convergência que desejamos fazer neste estudo. O primeiro deles será utilizado no teorema que garante convergência superlinear aos métodos secantes; já o segundo, auxiliará na garantia de convergência quadrática do método de Newton.

Lema 2.2.5. *Suponha que vale (20). Então, para todo $x, z \in \Omega$,*

$$\|F(z) - F(x) - J(x^*)(z - x)\| \leq L\|x - z\| \max\left\{\|x - x^*\|^q, \|z - x^*\|^q\right\}.$$

Demonstração. Disponível em [5]. □

Lema 2.2.6. *Suponha que vale (20). Então, para todo $x \in \Omega$,*

$$\|F(x) - J(x^*)(x - x^*)\| \leq \frac{L}{1+q}\|x - x^*\|^{q+1}.$$

Demonstração. Disponível em [8]. □

Ressaltamos que é comum supormos que a desigualdade (20) vale, na melhor das hipóteses, para $q = 1$, tendo em vista que a referida relação acontece com menor frequência para o caso $q > 1$. Por consequência, os Lemas 2.2.5 e 2.2.6 geralmente são explorados para o caso $q = 1$, fazendo com que o teorema a seguir, também encontrado em [25], comprove a taxa de convergência quadrática almejada.

Teorema 2.2.7. *Suponha que F, L e q satisfazem (20). Então existem* ε, $\gamma > 0$ *tais que para todo* x_0 *que satisfaz* $\|x_0 - x^*\| \leq \varepsilon$, *a sequência gerada por*

$$x_{k+1} = x_k - J_k^{-1} F_k, \qquad k = 0, 1, \ldots$$

está bem definida, converge a x^* *e satisfaz*

$$\|x_{k+1} - x^*\| \leq \gamma \|x_k - x^*\|^{q+1}$$

Demonstração. Escolha $r \in (0,1)$ arbitrário. Seja $\varepsilon_1 = \varepsilon_1(r)$, definido pelo Lema 2.2.3.

Como $J(x)$ é contínua, existe $\varepsilon_2 > 0$ tal que, sempre que $\|x - x^*\| \leq \varepsilon_2$, $\|J(x) - J(x^*)\| \leq \delta_1(r)$.

Defina $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$. Assim, $\|J_0 - J(x^*)\| \leq \delta_1(r)$ e, pelo Lema 2.2.3,

$$\|x_1 - x^*\| \leq r\|x_0 - x^*\| < \varepsilon_1.$$

Portanto, $\|J_1 - J(x^*)\| \leq \delta_1(r)$. Repetindo tal raciocínio indutivamente, concluímos que $\{x_k\}$ converge para x^* linearmente com taxa r.

Agora, de (15), temos que, para todo k,

$$\begin{aligned}
\|x_{k+1} - x^*\| &= \left\| x_k - x^* - J_k^{-1} F_k \right\| \\
&= \left\| J_k^{-1}(-F_k - J_k(x^* - x_k)) \right\| \\
&\leq \left\| J_k^{-1} \right\| \|F_k - J_k(x_k - x^*)\| \\
&\leq 2 \left\| J(x^*)^{-1} \right\| \|F_k - J_k(x_k - x^*)\|. \qquad (21)
\end{aligned}$$

Mas, por (20) e pelo Lema 2.2.6,

$$\begin{aligned}
\|F_k - J_k(x_k - x^*)\| &= \|F_k - J(x^*)(x_k - x^*) + J(x^*)(x_k - x^*) - J_k(x_k - x^*)\| \\
&\leq \|F_k - J(x^*)(x_k - x^*)\| + \|[J(x^*) - J_k](x_k - x^*)\| \\
&\leq \|F_k - J(x^*)(x_k - x^*)\| + \|J_k - J(x^*)\| \|x_k - x^*\| \\
&\leq \|F_k - J(x^*)(x_k - x^*)\| + L\|x_k - x^*\|^{q+1} \\
&\leq 2L\|x_k - x^*\|^{q+1}.
\end{aligned}$$

Disto e de (21), concluímos que

$$\|x_{k+1}-x^*\| \leq 4L\left\|J(x^*)^{-1}\right\|\|x_k-x^*\|^{q+1},$$

isto é,

$$\|x_{k+1}-x^*\| \leq \gamma\|x_k-x^*\|^{q+1},$$

onde $\gamma = 4L\left\|J(x^*)^{-1}\right\|$.

Isto completa a prova. □

Com este resultado, podemos concluir que o método de Newton funciona melhor quando o ponto inicial está suficientemente próximo a uma solução x^*, onde $J(x^*)$ é não singular, a ponto de garantir, no mínimo, convergência quadrática.

2.2.2 Convergência dos Métodos Quase-Newton

Assim como o método de Newton, a convergência dos métodos Quase-Newton é sustentada, principalmente, pelo Teorema das Duas Vizinhanças. Este resultado nos garante a convergência linear deste tipo de método sob a condição do ponto inicial estar suficientemente próximo da solução, e de todas as matrizes B_k's estarem suficientemente próximas de $J(x^*)$. Porém, mesmo que B_0 esteja em uma vizinhança razoável de $J(x^*)$, alguns métodos Quase-Newton, como é o caso dos métodos secantes, não garantem que, ao longo do processo iterativo, todas as B_k's posteriores permaneçam dentro da vizinhança que garante a convergência linear oferecida pelo Teorema das Duas Vizinhanças. Assim, existe o risco destas matrizes estarem fora da referida vizinhança, o que, em termos técnicos, definimos como uma deterioração excessiva sofrida por B_{k+1} em relação a B_k.

Neste contexto, tornam-se necessários resultados teóricos que garantam que, apesar de possíveis deteriorações, todas as B_k's estejam na boa vizinhança tratada pelo Teorema 2.2.4. Desta forma, teoremas de deterioração limitada são desenvolvidos e utilizados nas garantias de convergência dos métodos Quase-Newton. Como salientado nos estudos de Dennis e Walker [7], uma propriedade de deterioração limitada típica é

$$\|B_{k+1}-J(x^*)\| \leq (1+c_1)\|B_k-J(x^*)\| + c_2\sigma_k, \tag{22}$$

onde c_1 e c_2 são constantes positivas e $\sigma_k = \max\left\{\|x_k-x^*\|, \|x_{k+1}-x^*\|\right\}$.

Entre os métodos Quase-Newton que satisfazem a propriedade de deterioração limitada (22), sob a condição de continuidade Lipschitz de $J(x)$, destacamos os métodos secantes de posto um. O lema a seguir, também encontrado em [4], evidencia esse fato.

Lema 2.2.8. *Suponha que vale (20), com $q = 1$, e que a atualização secante é de posto um, definida por*

$$B_{k+1} = B_k + \frac{(y_k - B_k p_k)w_k^T}{w_k^T p_k},$$

onde $\left|w_k^T p_k\right| \geq \gamma \|w_k\| \|p_k\|$, *para algum* $\gamma > 0$. *Aqui,* $y_k = F_{k+1} - F_k$, $p_k = x_{k+1} - x_k$, *e* w_k *é qualquer vetor não ortogonal a* p_k.

Sob tais hipóteses, vale (22).

Demonstração. Perceba que

$$\begin{aligned} B_{k+1} - J(x^*) &= B_k - J(x^*) + \frac{(y_k - B_k p_k) w_k^T}{w_k^T p_k} \\ &= B_k - J(x^*) + \frac{(y_k - J(x^*) p_k + J(x^*) p_k - B_k p_k) w_k^T}{w_k^T p_k} \\ &= B_k - J(x^*) + \frac{(J(x^*) p_k - B_k p_k) w_k^T}{w_k^T p_k} + \frac{(y_k - J(x^*) p_k) w_k^T}{w_k^T p_k} \\ &= (B_k - J(x^*)) \left(I - \frac{p_k w_k^T}{w_k^T p_k} \right) + \frac{(y_k - J(x^*) p_k) w_k^T}{w_k^T p_k}. \end{aligned} \tag{23}$$

Sabemos do Lema 2.2.5 que

$$\|y_k - J_* p_k\| \leq L \|p_k\| \max \left\{ \|x_k - x^*\|, \|x_{k+1} - x^*\| \right\}.$$

Disto e de (23),

$$\begin{aligned} \|B_{k+1} - J(x^*)\| &\leq \|B_k - J(x^*)\| \left(1 + \frac{\|p_k\| \|w_k\|}{\left|w_k^T p_k\right|} \right) + \frac{\|y_k - J(x^*) p_k\| \|w_k\|}{\left|w_k^T p_k\right|} \\ &\leq \left(1 + \frac{1}{\gamma} \right) \|B_k - J(x^*)\| + \left(\frac{L}{\gamma} \right) \max \left\{ \|x_k - x^*\|, \|x_{k+1} - x^*\| \right\} \\ &= (1 + c_1) \|B_k - J(x^*)\| + c_2 \sigma_k, \end{aligned}$$

onde $c_1 = \frac{1}{\gamma}$, $c_2 = \frac{L}{\gamma}$ e $\sigma_k = \max \left\{ \|x_k - x^*\|, \|x_{k+1} - x^*\| \right\}$.

Portanto, vale (22). □

A propriedade de deterioração limitada recém abordada é utilizada para mostrar a convergência linear dos métodos que a satisfazem, como será evidenciado no próximo teorema. Este resultado também é encontrado no artigo de Broyden, Dennis e Moré [5]. Para tanto, necessitaremos, ainda, utilizar uma norma matricial $\|\cdot\|_M$ e uma norma vetorial $\|\cdot\|$ não relacionadas. Uma vez que todas normas em um espaço vetorial de dimensão finita são equivalentes, existe uma constante $\tilde{\eta}$ tal que

$$\|A\| \leq \tilde{\eta} \|A\|_M, \tag{24}$$

onde $\|\cdot\|$ denota a norma induzida pela correspondente norma vetorial. Denotaremos, ainda, $\mathscr{L}(\mathbb{R}^n)$ o espaço linear de todas matrizes reais de ordem n, e $\mathscr{P}\{\mathscr{L}(\mathbb{R}^n)\}$ a coleção de todos subconjuntos não vazios de $\mathscr{L}(\mathbb{R}^n)$.

Teorema 2.2.9. *Seja $F : \mathbb{R}^n \to \mathbb{R}^n$ uma função diferenciável em um conjunto aberto e convexo Λ, e assuma que para algum $x^* \in \Lambda$ e $q > 0$ a desigualdade (20) vale, onde $F(x^*) = 0$ e $J(x^*)$ é não singular. Seja $U : \mathbb{R}^n \times \mathscr{L}(\mathbb{R}^n) \to \mathscr{P}\{\mathscr{L}(\mathbb{R}^n)\}$ definida numa vizinhança $N = N_1 \times N_2$ de $(x^*, J(x^*))$, onde N_1 está contido em Λ e N_2 contém apenas matrizes não singulares. Suponha que existam constantes não negativas c_1 e c_2 tais que para cada $(x,B) \in N$, e para $\bar{x} = x - B^{-1}F(x)$, a função U satisfaz*

$$\|\bar{B} - J(x^*)\| \leq (1 + c_1)\|B - J(x^*)\| + c_2 \max\{\|\bar{x} - x^*\|^q, \|x - x^*\|^q\}, \tag{25}$$

para cada $\bar{B} \in U(x,B)$. Então, para cada $r \in (0,1)$, existem constantes positivas $\varepsilon(r)$ e $\delta(r)$ tais que se $\|x_0 - x^*\| \leq \varepsilon(r)$ e $\|B_0 - J(x^*)\| \leq \delta(r)$ e qualquer $B_{k+1} \in U(x_k, B_k)$, a sequência

$$x_{k+1} = x_k - B_k^{-1} F_k$$

está bem definida e converge para x^*. Além disso,

$$\|x_{k+1} - x^*\| \leq r\|x_k - x^*\|,$$

para todo $k \geq 0$, e $\{\|B_k\|\}$ e $\{\|B_k\|^{-1}\}$ são uniformemente limitadas.

Demonstração. Seja $r \in (0,1)$ dado e tome $\gamma \geq \|J(x^*)\|^{-1}$. Agora escolha $\varepsilon(r) = \varepsilon$ e $\delta(r) = \delta$ tais que

$$[2c_1\delta + c_2]\frac{\varepsilon^q}{1 - r^q} \leq \delta, \tag{26}$$

e para $\tilde{\eta}$ dado por (24),

$$\gamma(1 + r)[L\varepsilon^q + 2\tilde{\eta}\delta] \leq r. \tag{27}$$

Se necessário restrinja ainda mais ε e δ tal que $(x,B) \in N$ toda vez que $\|B - J(x^*)\|_M < 2\delta$ e $\|x - x^*\| < \varepsilon$. Agora suponha que $\|B_0 - J(x^*)\|_M < \delta$ e $\|x_0 - x^*\| < \varepsilon$. Então, $\|B_0 - J(x^*)\| < \tilde{\eta}\delta < 2\tilde{\eta}\delta$.

Tomemos $A = B_0 J(x^*)^{-1} - I$. Como vale (27), temos

$$2\gamma(1 + r)\tilde{\eta}\delta \leq r, \tag{28}$$

isto é,

$$2\tilde{\eta}\delta\gamma \leq \frac{r}{1 + r} < 1. \tag{29}$$

Desta forma,

$$\begin{aligned} \|A\| &= \left\| [B_0 - J(x^*)] J(x^*)^{-1} \right\| \\ &\leq \|B_0 - J(x^*)\| \left\| J(x^*)^{-1} \right\| \\ &< 2\tilde{\eta}\delta\gamma < 1. \end{aligned}$$

Segue, então, do Lema de Banach (Lema 2.2.1) que A e $I+A$ são não singulares. Além disto, B_0^{-1} existe, vale que $(I+A)^{-1} = \left[B_0 J(x^*)^{-1}\right]^{-1} = J(x^*)B_0^{-1}$, e

$$\left\|(I+A)^{-1}\right\| \leq \frac{1}{1-\|A\|}.$$

Disto e de (29), temos

$$\left\|J(x^*)B_0^{-1}\right\| \leq \frac{1}{1-\frac{r}{1+r}} = 1+r.$$

Assim,

$$\begin{aligned}\left\|B_0^{-1}\right\| &= \left\|J(x^*)^{-1}J(x^*)B_0^{-1}\right\| \\ &\leq \left\|J(x^*)^{-1}\right\|\left\|J(x^*)B_0^{-1}\right\| \\ &\leq \gamma(1+r).\end{aligned}$$

Usando esta desigualdade, o Lema 2.2.5 e (27), obtemos

$$\begin{aligned}\|x_1-x^*\| &= \left\|x_0-B_0^{-1}F_0-x^*\right\| \\ &= \left\|B_0^{-1}\left[B_0(x_0-x^*)-F_0+J(x^*)(x_0-x^*)-J(x^*)(x_0-x^*)\right]\right\| \\ &\leq \left\|B_0^{-1}\right\|\left[\|F_0-F(x^*)-J(x^*)(x_0-x^*)\|+\|B_0-J(x^*)\|\|x_0-x^*\|\right] \\ &\leq \gamma(1+r)\left[L\varepsilon^q+2\tilde{\eta}\,\delta\right]\|x_0-x^*\| \\ &\leq r\|x_0-x^*\| < \varepsilon.\end{aligned}$$

Portanto, $x_1 \in \Lambda$. Completaremos a prova com um argumento de indução. Assuma que $\|B_k-J(x^*)\|_M \leq 2\delta$ e $\|x_{k+1}-x^*\| \leq r\|x_k-x^*\|$ para $k=0,1,\ldots,m-1$. Segue, então, de (25) que

$$\|B_{k+1}-J(x^*)\|_M - \|B_k-J(x^*)\|_M \leq 2c_1\delta\varepsilon^q r^{kq} + c_2\varepsilon^q r^{kq}$$

e somando ambos os lados de $k=0$ até $m-1$, obtemos

$$\|B_m-J(x^*)\|_M \leq \|B_0-J(x^*)\|_M + \left[2c_1\delta+c_2\right]\frac{\varepsilon^q}{1-r^q}.$$

Disto e de (26), $\|B_m-J(x^*)\|_M \leq 2\delta$. Para completar a indução resta provarmos que $\|x_{m+1}-x^*\| \leq r\|x_m-x^*\|$. Isto segue de um procedimento similar ao caso $m=1$. De fato, como $\|B_m-J(x^*)\| \leq 2\tilde{\eta}\,\delta$, do Lema de Banach e de (28) obtemos

$$\left\|B_m^{-1}\right\| \leq \gamma(1+r),$$

e, assim, usando do Lema 2.2.5, teremos

$$\|x_{m+1}-x^*\| = \left\|x_m-B_m^{-1}F_m-x^*\right\|$$

$$
\begin{aligned}
&= \left\| B_m^{-1} \left[B_m(x_m - x^*) - F_m + J(x^*)(x_m - x^*) - J(x^*)(x_m - x^*) \right] \right\| \\
&\leq \left\| B_m^{-1} \right\| \left[\|F_m - F(x^*) - J(x^*)(x_m - x^*)\| + \|B_m - J(x^*)\| \|x_m - x^*\| \right] \\
&\leq \gamma(1+r) \left[L\varepsilon^q + 2\tilde{\eta}\,\delta \right] \|x_m - x^*\| \\
&\leq r\|x_m - x^*\|.
\end{aligned}
$$

Isto completa a prova.

□

A convergência linear dos métodos Quase-Newton que satisfazem a propriedade de deterioração limitada (22), por si só, não fornece uma estimativa muito boa da velocidade de convergência destes métodos. Assim como realizado em [25], veremos, posteriormente, que os métodos secantes que satisfazem uma condição específica, conhecida como condição de Dennis-Moré, convergem superlinearmente. Antes, porém, necessitamos de um resultado auxiliar que nos mostrará que, quando $J(x^*)$ é não singular, $\|F(x)\|$ pode ser utilizado como uma medida da distância entre x e x^*.

Lema 2.2.10. *Seja $F : \mathbb{R}^n \to \mathbb{R}^n$ continuamente diferenciável em $\mathbb{R}^n$ e suponha que $J(x^*)$ é não singular. Então, existem $\varepsilon, c_1, c_2 > 0$ tais que, sempre que $\|x - x^*\| \leq \varepsilon$, $x \in \mathbb{R}^n$, vale*

$$c_1\|x - x^*\| \leq \|F(x)\| \leq c_2\|x - x^*\|.$$

Demonstração. Disponível em [25]. □

Teorema 2.2.11 (Teorema de Dennis-Moré)**.** *Seja $F : \mathbb{R}^n \to \mathbb{R}^n$ continuamente diferenciável em $\mathbb{R}^n$, e suponha que vale a relação (20), que a sequência gerada por*

$$x_{k+1} = x_k - B_k^{-1} F_k$$

está bem definida, converge a x^*, e satisfaz

$$\lim_{k \to \infty} \frac{\|[B_k - J(x^*)]p_k\|}{\|p_k\|} = 0. \tag{30}$$

Então a referida sequência converge superlinearmente.

Demonstração. Sabemos que

$$
\begin{aligned}
[B_k - J(x^*)]p_k &= [B_k - J(x^*)](x_{k+1} - x_k) \\
&= B_k(x_{k+1} - x_k) - J(x^*)(x_{k+1} - x_k) \\
&= B_k(-B_k^{-1} F_k) - J(x^*)(x_{k+1} - x_k) \\
&= -F_k - J(x^*)(x_{k+1} - x_k)
\end{aligned}
$$

$$= F_{k+1} - F_k - J(x^*)(x_{k+1} - x_k) - F_{k+1}. \tag{31}$$

Sabemos, também, pelo Lema 2.2.5, que

$$\|F_{k+1} - F_k - J(x^*)(x_{k+1} - x_k)\| \leq L\|x_{k+1} - x_k\| \max\left\{\|x_k - x^*\|^q, \|x_{k+1} - x^*\|^q\right\}.$$

Disto, de (31), da convergência de $\{x_k\}$ e de (30), concluímos, então, que

$$\lim_{k\to\infty} \frac{\|F_{k+1}\|}{\|x_{k+1} - x_k\|} = 0. \tag{32}$$

Agora, pelo Lema 2.2.10, sabemos que, para k suficientemente grande, existe $c_1 > 0$ tal que $\|F_{k+1}\| \geq c_1\|x_{k+1} - x^*\|$. Disto e da desigualdade triangular de normas, segue de (32) que

$$0 = \lim_{k\to\infty} \frac{\|F_{k+1}\|}{\|x_{k+1} - x_k\|} \geq \lim_{k\to\infty} \frac{c_1\|x_{k+1} - x^*\|}{\|x_{k+1} - x^*\| + \|x_k - x^*\|}.$$

Multiplicando o numerador e o denominador da fração à direita da última inequação por $\frac{1}{\|x_k - x^*\|}$, temos que

$$\lim_{k\to\infty} \frac{c_1\|x_{k+1} - x^*\|}{\|x_{k+1} - x^*\| + \|x_k - x^*\|} = c_1 \lim_{k\to\infty} \frac{\tilde{r}_k}{\tilde{r}_k + 1} = 0,$$

onde $\tilde{r}_k = \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|}$.

Portanto,

$$\lim_{k\to\infty} \tilde{r_k} = 0,$$

o que completa a prova de convergência superlinear da sequência em questão. □

É importante observarmos que a condição de Dennis-Moré (30) é automaticamente satisfeita pelo método de Newton e por todos os métodos nos quais B_k converge para $J(x^*)$. Apesar disto, é possível que a condição valha mesmo que B_k não convirja para $J(x^*)$, já que o que deve tender para zero não é a diferença $B_k - J(x^*)$, mas sim a aplicação desta diferença na direção incremental $\frac{p_k}{\|p_k\|}$.

Assim como o método de Newton, os métodos Quase-Newton não oferecem garantia de que convergirão para uma solução do problema $F(x) = 0$, a menos de que eles sejam iniciados suficientemente próximos dessa solução. Desta forma, a fim de aumentar a robustez destes métodos, podemos usufruir de estratégias de globalização, isto é, de alternativas para se obter um método que convirja a partir de qualquer ponto inicial.

As duas estratégias de globalização comumente exploradas nos métodos para sistemas de equações não lineares são a busca linear e as regiões de confiança [27]. Naquela, a partir de um iterado x_k, o algoritmo escolhe uma direção p_k, na qual irá procurar um novo iterado com valor de função menor do que o anterior; enquanto nesta, as informações coletadas sobre a função objetivo f são usadas para construir um modelo m_k, cujo comportamento perto do ponto

atual x_k é semelhante ao de f. Como o modelo m_k pode não ser uma boa aproximação de f quando x está longe de x_k, restringimos a busca por um minimizador de m_k para alguma região em torno de x_k — a região de confiança. Este minimizador (restrito) será tomado como passo da k-ésima iteração se fornecer um decréscimo suficientemente grande para a função f.

Conforme Martínez e Santos [25], quando tratamos de métodos para sistemas de equações não lineares, as regiões de confiança costumam ser mais eficazes do que as buscas lineares no processo de globalização de convergência. Por este motivo destinamos uma seção para explicar os fundamentos que sustentam tal técnica.

2.3 MÉTODO DE REGIÃO DE CONFIANÇA

Os métodos de região de confiança constituem importantes procedimentos iterativos para a solução de problemas de otimização, sendo utilizados, principalmente, com o intuito de controlar, a cada iteração, o tamanho do passo dado por um determinado algoritmo. Sendo assim, configuram estratégias passíveis de combinações com diferentes escolhas de passos — a exemplo de passos Newton e Quase-Newton.

Consideremos o problema de otimização irrestrito

$$\min_{x\in\mathbb{R}^n} f(x), \tag{33}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$ é uma função duas vezes continuamente diferenciável.

A fundamentação dos métodos de região de confiança consiste, basicamente, em substituir, a cada iteração, a função f por um modelo que a aproxime em uma vizinhança do iterado x_k. Para tanto, escolhemos uma região em torno de x_k onde este modelo, a ser denotado por m_k, é confiável, isto é, na qual m_k representa suficientemente bem o comportamento de f. Neste contexto, o passo a ser dado a partir de x_k para identificar x_{k+1} é baseado na minimização do modelo na referida região de confiança, implicando na substituição do problema irrestrito (33) pelo problema restrito

$$\min_{p\in\mathbb{R}^n} m_k(p) \text{ sujeito a } \|p\| \leq \Delta_k, \tag{34}$$

onde Δ_k representa o raio da região de confiança na k-ésima iteração. Em geral, definimos $\|\cdot\|$ como a norma Euclidiana e, por consequência, a solução p_k^* de (34) é o minimizador de m_k na bola centrada em x_k com raio Δ_k.

Um dos modelos mais utilizados em métodos de região de confiança é o modelo quadrático, definido, a cada iteração, por

$$m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T H_k p, \tag{35}$$

onde H_k é uma matriz simétrica que aproxima $\nabla^2 f_k$, podendo ser a própria matriz Hessiana de f. Este modelo é baseado na série de Taylor de segunda ordem dada por

$$f(x_k + p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T \nabla^2 f(x_k + tp) p,$$

para algum $t \in (0,1)$. Desta forma, a diferença entre $m_k(p)$ e $f(x_k + p)$ é da ordem $O(\|p\|^2)$, levando-nos a concluir que o erro de aproximação é pequeno quando p é pequeno.

A escolha do raio da região de confiança na k-ésima iteração é baseada na concordância entre o modelo m_k e a função objetivo f. Para tanto, dado um passo $p \in \mathbb{R}^n$, definimos a razão

$$r_k(p) = \frac{f(x_k) - f(x_k + p)}{m_k(0) - m_k(p)}.$$

O numerador de r_k é denominado redução real da função, e o denominador é a redução prevista pelo modelo.

Note que $m_k(0) = f_k$. Disto e do fato de p_k minimizar m_k em uma região em torno de x_k, temos que a redução prevista sempre deverá ser não negativa. Consequentemente, se $r_k(p_k)$ for negativo, o novo valor da função $f(x_k + p_k)$ é maior do que o valor corrente $f(x_k)$, implicando na rejeição do passo p_k.

Neste contexto, $r_k(p_k)$ próximo de 1 indica uma boa concordância entre $m_k(p_k)$ e $f(x_k + p_k)$, sendo seguro expandir a região de confiança na próxima iteração. Em contrapartida, $r_k(p_k)$ negativo ou próximo de zero induz a necessidade de diminuir a região de confiança. Nos casos em que $r_k(p_k)$ for positivo, mas não muito próximo de um e nem de zero, não alteramos o tamanho da região de confiança para a próxima iteração.

O tamanho da região de confiança é fundamental para a eficácia de cada iteração. Por um lado, se a região for muito pequena e representar bem a função, o algoritmo perde a oportunidade de dar um passo substancial que o moverá mais perto do minimizador da função objetivo. Em contrapartida, se a região de confiança for muito grande, o minimizador do modelo pode estar longe de ser o minimizador da função objetivo na região, obrigando-nos a reduzir seu tamanho.

O algoritmo a seguir, baseado na obra de Nocedal e Wright [27], descreve o processo supracitado.

Algoritmo 3: Método de Região de Confiança

Entrada: $x_0 \in \mathbb{R}^n$; $\bar{\Delta} > 0$; $\Delta_0 \in (0, \bar{\Delta})$; e $\delta \in \left[0, \frac{1}{4}\right)$;
Saída: $x^* \in \mathbb{R}^n$;
início
  $k \longleftarrow 0$;
  Calcule ∇f_k e H_k;
  repita
    1) Obtenha p_k resolvendo o subproblema

$$\min_{p \in \mathbb{R}^n} m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T H_k p, \text{ sujeito a } \|p\| \leq \Delta_k;$$

    2) Calcule

$$r_k(p_k) = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)};$$

    se $r_k(p_k) < \frac{1}{4}$ **então**
      $\Delta_{k+1} \longleftarrow \frac{1}{4}\Delta_k$;
    senão
      se $r_k(p_k) > \frac{3}{4}$ *e* $\|p_k\| = \Delta_k$ **então**
        $\Delta_{k+1} \longleftarrow \min(2\Delta_k, \bar{\Delta}_k)$;
      senão
        $\Delta_{k+1} \longleftarrow \Delta_k$;
      fim
    fim
    se $r_k(p_k) > \delta$ **então**
      $x_{k+1} \longleftarrow x_k + p_k$;
      Calcule ∇f_{k+1} e H_{k+1};
    senão
      $x_{k+1} \longleftarrow x_k$;
    fim
    3) $k \longleftarrow k + 1$;
  até $\nabla f_k = 0$;
fim
$x^* \longleftarrow x_k$.

Diferentemente da otimização irrestrita, nos sistemas de equações não lineares geralmente é necessário definir uma função auxiliar para medir o progresso do método em direção a uma solução do problema — denominada função de mérito. A função de mérito é uma função

com valor escalar em x que indica se uma nova iteração é melhor ou pior do que a iteração atual. Em sistemas de equações não lineares esta função é comumente obtida pela combinação das n componentes da função vetorial F, sendo uma das mais utilizadas a metade do quadrado da norma euclidiana de F, ou seja, a função $f : \mathbb{R}^n \to \mathbb{R}$ definida por

$$f(x) = \tfrac{1}{2}\|F(x)\|^2. \tag{36}$$

Desta forma, as derivadas de $f(x)$ podem ser expressas em termos da Jacobiana $J(x)$, já que

$$\nabla f(x) = \sum_{j=1}^{m} F_j(x)\nabla F_j(x) = J(x)^T F(x), \tag{37}$$

e, assim,

$$\begin{aligned}\nabla^2 f(x) &= \sum_{j=1}^{m} \nabla F_j(x)\nabla F_j(x)^T + \sum_{j=1}^{m} F_j(x)\nabla^2 F_j(x) \\ &= J(x)^T J(x) + \sum_{j=1}^{m} F_j(x)\nabla^2 F_j(x).\end{aligned} \tag{38}$$

Frequentemente $J(x)^T J(x)$ é mais importante do que o segundo termo da soma em (38), já que os resíduos F_j costumam se comportar de maneira próxima da linearidade em regiões adjacentes a uma solução e, assim, $\nabla^2 F_j$ são relativamente pequenos. Por não exigir derivadas de segunda ordem (que podem não estar disponíveis ou ser de difícil obtenção), a possibilidade de assumir a aproximação $\nabla^2 f(x) \approx J(x)^T J(x)$ mostra-se, desta forma, viável e atrativa. Pensando nisto, os métodos de região de confiança mais amplamente usados para equações não lineares utilizam a função de mérito (36) e o modelo Gauss-Newton $m_k^{GN}(p)$ definido por

$$\begin{aligned}m_k^{GN}(p) &= \tfrac{1}{2}\|F_k + J_k p\|^2 \\ &= \tfrac{1}{2}\|F_k\|^2 + F_k^T J_k p + \tfrac{1}{2} p^T J_k^T J_k p \\ &= f_k + \nabla f_k^T p + \tfrac{1}{2} p^T J_k^T J_k p.\end{aligned} \tag{39}$$

Desta forma, o passo p_k é gerado mediante a procura de uma solução do subproblema

$$\text{Minimizar } m_k^{GN}(p) \text{ sujeito a } \|p\| \leq \Delta_k,$$

onde Δ_k é o raio de região de confiança definido na k-ésima iteração. Assim, a razão $r_k(p)$ entre a redução atual da função e a redução prevista pelo modelo é dada por

$$r_k(p) = \frac{\|F_k\|^2 - \|F(x_k + p)\|^2}{\|F_k\|^2 - \|F_k + J_k p\|^2}.$$

A efetividade do método de região de confiança depende, crucialmente, da resolução do subproblema estabelecido no passo 1 do Algoritmo 3. Duas estratégias comumente utilizadas

para encontrar soluções aproximadas deste subproblema são o método Dogleg e o método Steihaug [27]. O primeiro é apropriado para os casos em que a matriz H_k é positiva definida para todo k; já o segundo é mais recomendado quando $H_k = \nabla^2 f_k$, e quando esta matriz é grande e esparsa.

No método que iremos propor no Capítulo 3, iremos recorrer a uma versão Quase-Newton do modelo quadrático (39), na qual as matrizes H_k são semidefinidas positivas. Portanto, utilizaremos apenas o método Dogleg para a resolução de (34) — razão pela qual destinamos uma seção a seguir para abordar o método. Antes, porém, abordaremos um importante conceito inerente à referida técnica: o ponto de Cauchy.

2.3.1 O ponto de Cauchy

A convergência global dos métodos de região de confiança é garantida a partir do fundamento de que, a cada iteração, encontramos um passo, com direção e tamanho adequado, que forneça redução suficiente no modelo considerado. Esta redução pode ser quantificada em termos do ponto, na direção de máxima descida $-\nabla f_k$, que causa o maior decréscimo do modelo m_k dentro da região de confiança: o ponto de Cauchy. Defina, então,

$$d_k^C = -\nabla f_k.$$

A razão pela qual buscamos minimizar o modelo quadrático m_k a partir da direção d_k^C é justificada pelo fato de que, localmente, m_k se comporta de maneira semelhante à sua versão linear l_k, dada por

$$l_k(p) = f_k + \nabla f_k^T p. \tag{40}$$

Ora, a solução p_k^* do problema

$$\underset{p\in\mathbb{R}^n}{\arg\min}\ f_k + \nabla f_k^T p \quad \text{sujeito à} \quad \|p\| \leq \Delta_k$$

é simplesmente o ponto de interseção entre a direção de máxima descida e o bordo da região de confiança, dado por

$$p_k^* = -\frac{\Delta_k}{\|\nabla f_k\|}\nabla f_k \tag{41}$$

isto é, um múltiplo positivo do vetor $-\nabla f_k = d_k^C$.

Denotamos o ponto de Cauchy na k-ésima iteração por p_k^C, e definimos

$$p_k^C = \tau_k d_k^C,$$

onde

$$\tau_k = \underset{\tau>0}{\arg\min}\left\{ m_k(\tau d_k^C) \mid \left\| \tau\, d_k^C \right\| \leq \Delta_k \right\},$$

e Δ_k é o raio de região de confiança na k-ésima iteração.

O tamanho do passo de Cauchy τ_k pode ser calculado explicitamente, conforme a proposição a seguir.

Proposição 2.3.1. *Seja p_k^C o ponto de Cauchy. Então,*

$$\tau_k = \min\left\{ \frac{\|\nabla f_k\|^2}{\nabla f_k^T H_k \nabla f_k}, \frac{\Delta_k}{\|\nabla f_k\|} \right\}$$

Demonstração. Perceba que

$$m_k(\tau d_k^C) = f_k - \tau\|\nabla f_k\|^2 + \tfrac{1}{2}\tau^2 \nabla f_k^T H_k \nabla f_k.$$

Assim, para o caso $\nabla f_k^T H_k \nabla f_k \leq 0$, temos que, sempre que $\nabla f_k \neq 0$, a função $m_k(\tau d_k^C)$ decresce monotonamente conforme τ cresce. Daí, para que a redução do modelo seja a maior possível, respeitando o limite da região de confiança, τ_k deves ser dado por

$$\tau_k = \frac{\Delta_k}{\|\nabla f_k\|}, \tag{42}$$

toda vez que $\nabla f_k \neq 0$.

Já para o caso $\nabla f_k^T H_k \nabla f_k > 0$, $m_k(\tau d_k^C)$ será uma quadrática convexa em τ. Logo, τ_k será o minimizador irrestrito desta quadrática, definido por

$$\tau_k = \frac{\|\nabla f_k\|^2}{\nabla f_k^T H_k \nabla f_k},$$

quando este minimizador estiver contido na região de confiança; caso contrário, τ_k será o limite da região de confiança definido por (42). □

O passo de Cauchy mostra-se barato do ponto de vista computacional, principalmente por não exigir nenhuma fatoração de matriz. Além disto, ele é suficiente para garantir convergência global a um método de região de confiança, porém tal convergência pode ser lenta em determinados problemas. Pensando na aceleração deste processo deu-se o desenvolvimento do método Dogleg, a ser explorado na subseção a seguir.

2.3.2 O método Dogleg

O método Dogleg é uma estratégia de resolução do subproblema de região de confiança que possibilita a combinação do passo de Cauchy com passos Newton ou Quase-Newton, buscando usufruir das vantagens ofertadas pelos dois tipos de passos envolvidos. Devido a sua estruturação e fundamentação, a utilização do referido método é recomendada apenas nos casos em que a matriz H_k do modelo quadrático (35) é definida positiva.

Assuma, portanto, que H_k é definida positiva, e denote a solução de (34) por $p^*(\Delta_k)$, onde Δ_k é o raio de região de confiança na k-ésima iteração. Sob tal condição, o minimizador irrestrito do modelo m_k definido por (35) será dado por $p^B = -H_k^{-1}\nabla f_k$. Então, quando a região de confiança for grande o suficiente a ponto de conter tal minimizador, ou seja, quando $\|p^B\| \leq \Delta_k$, teremos

$$p^*(\Delta_k) = p^B.$$

Por outro lado, quando a região de confiança não contiver o minimizador irrestrito de m_k ($\|p^B\| > \Delta_k$), a restrição $\|p\| \leq \Delta_k$ impõe que a parte quadrática de m_k não terá muita influência para o modelo. Neste caso, a solução $p^*(\Delta_k)$ pode ser aproximada pela solução de

$$\min_{p\in\mathbb{R}^n} l_k(p) \text{ sujeito a } \|p\| \leq \Delta_k,$$

onde l_k é definido por (40). Desta forma, tomaremos

$$p^*(\Delta_k) = \frac{-\Delta_k}{\|\nabla f_k\|}\nabla f_k,$$

isto é, tomaremos como solução a interseção entre a direção de máxima descida $-\nabla f_k$ e o bordo da região de confiança.

Para valores intermediários de Δ, a solução $p^*(\Delta)$ segue a trajetória curva evidenciada na figura abaixo:

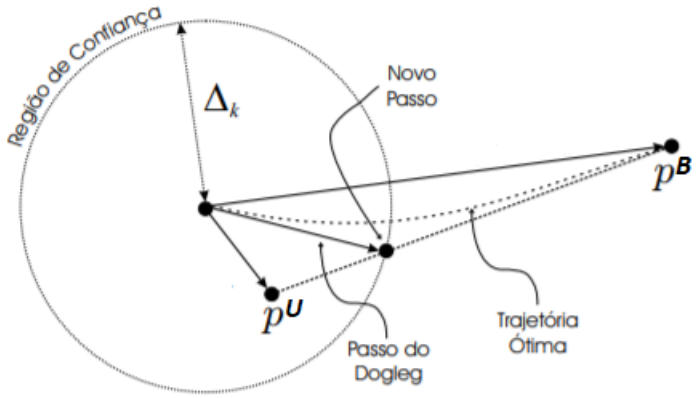

Figura 3 – Passo Dogleg.

O método Dogleg calcula uma solução aproximada de (34), substituindo a trajetória ótima por uma aproximação composta por dois segmentos de reta. O primeiro segmento vai da origem até o minimizador de m_k na direção de máxima descida, definido por

$$p^U = \frac{-\|\nabla f_k\|^2}{\nabla f_k^T H_k \nabla f_k}\nabla f_k.$$

Já o segundo, segmento percorre p^U até p^B.

Assim, representamos o passo Dogleg por $p^D(\tilde{\tau})$, para $\tilde{\tau} \in [0,2]$ da seguinte maneira:

$$p^D(\tilde{\tau}) = \begin{cases} \tilde{\tau} p^U, & 0 \leq \tilde{\tau} \leq 1, \\ p^U + (\tilde{\tau} - 1)(p^B - p^U), & 1 \leq \tilde{\tau} \leq 2. \end{cases} \tag{43}$$

Assim como feito na obra de Nocedal e Wright [27], veremos no lema a seguir que à medida que $\tilde{\tau}$ cresce o tamanho do passo Dogleg aumenta, simultaneamente com a redução do modelo. Disto e do fato de que o caminho Dogleg intersepta a região de confiança no máximo uma vez, o minimizador do modelo na região de confiança ao longo do referido caminho pode ser obtido analiticamente.

Lema 2.3.2. *Seja H positiva definida. Então,*

(i) $\left\|p^D(\tilde{\tau})\right\|$ *é uma função crescente de* $\tilde{\tau}$*, e*

(ii) $m\left(p^D(\tilde{\tau})\right)$ *é uma função decrescente de* $\tilde{\tau}$*.*

Demonstração. Consideremos, inicialmente, $\tilde{\tau} \in [0,1]$. Perceba que

$$\left\|p^D(\tilde{\tau})\right\| = \left\|\tilde{\tau} p^U\right\|$$

cresce quando $\tilde{\tau}$ cresce, o que confirma a validade de (i). Além disto,

$$\begin{aligned} m\left(p^D(\tilde{\tau})\right) &= m\left(\tilde{\tau} p^U\right) \\ &= f + \tilde{\tau} \nabla f^T p^U + \frac{1}{2} \tilde{\tau}^2 (p^U)^T H p^U. \end{aligned}$$

Derivando esta expressão com relação a $\tilde{\tau}$, temos que

$$\begin{aligned} m'\left(p^D(\tilde{\tau})\right) &= \nabla f^T p^U + \tilde{\tau} (p^U)^T H p \\ &= -\frac{\|\nabla f\|^4}{\nabla f^T H \nabla f} + \tilde{\tau} \frac{\|\nabla f\|^4}{\nabla f^T H \nabla f} \\ &\leq -\frac{\|\nabla f\|^4}{\nabla f^T H \nabla f} + \frac{\|\nabla f\|^4}{\nabla f^T H \nabla f} \\ &= 0. \end{aligned}$$

Logo, $m\left(p^D(\tilde{\tau})\right)$ é uma função decrescente com relação a $\tilde{\tau}$, e, neste caso, vale (ii).

Considere, agora, $\tilde{\tau} \in [1,2]$ e $\alpha \in (0,1)$. Defina $h(\alpha)$ por

$$\begin{aligned} h(\alpha) &= \tfrac{1}{2} \left\|p^D(1+\alpha)\right\|^2 \\ &= \tfrac{1}{2} \left\|p^U + \alpha(p^B - p^U)\right\|^2 \end{aligned}$$

$$= \frac{1}{2}\left\|p^U\right\|^2 + \alpha(p^U)^T(p^B - p^U) + \frac{1}{2}\alpha^2\left\|p^B - p^U\right\|^2.$$

Para que valha (i), precisamos mostrar que $h'(\alpha) \geq 0$ para $\alpha \in (0,1)$. Ora, mas

$$\begin{aligned} h'(\alpha) &= -(p^U)^T(p^U - p^B) + \alpha\left\|p^U - p^B\right\| \\ &\geq -(p^U)^T(p^U - p^B) \\ &= \frac{\|\nabla f\|^2}{\nabla f^T H \nabla f}\nabla f^T\left(-\frac{\|\nabla f\|^2}{\nabla f^T H \nabla f}\nabla f + H^{-1}\nabla f\right) \\ &= \|\nabla f\|^2\frac{\nabla f^T H^{-1}\nabla f}{\nabla f^T H \nabla f}\left(1 - \frac{\|\nabla f\|^4}{(\nabla f^T H \nabla f)(\nabla f^T H^{-1}\nabla f)}\right). \end{aligned} \tag{44}$$

Do fato de H ser positiva definida, $\nabla f^T H \nabla f > 0$ e $\nabla f^T H^{-1}\nabla f > 0$. Por consequência, o primeiro fator de (44) é não negativo, restando-nos a tarefa de mostrar tal característica, também, para o segundo fator.

Pela desigualdade de Cauchy-Schwarz, sabemos que, dados $u, v \in \mathbb{R}^n$,

$$(u^T v)^2 \leq \|u\|^2 \|v\|^2 = (u^T u)(v^T v).$$

Tomemos $u = H^{1/2}\nabla f$ e $v = H^{-1/2}\nabla f$. Desta forma, $u^T v = \nabla f^T \nabla f$, $u^T u = \nabla f^T H \nabla f$ e $v^T v = \nabla f^T H^{-1}\nabla f$. Logo,

$$\begin{aligned} \|\nabla f\|^4 &= (\nabla f^T \nabla f)^2 \\ &\leq (\nabla f^T H \nabla f)(\nabla f^T H^{-1}\nabla f). \end{aligned}$$

Portanto,

$$\frac{\|\nabla f\|^4}{(\nabla f^T H \nabla f)(\nabla f^T H^{-1}\nabla f)} \leq 1,$$

isto é,

$$1 - \frac{\|\nabla f\|^4}{(\nabla f^T H \nabla f)(\nabla f^T H^{-1}\nabla f)} \geq 0.$$

Disto e de (44), concluímos que $h'(\alpha) \geq 0$.

Para (ii), definimos $\tilde{h}(\alpha) = m(p^D(1+\alpha))$ e mostraremos que $\tilde{h}'(\alpha) \leq 0$ para $\alpha \in (0,1)$.

Substituindo (43) em (35), temos que

$$\begin{aligned} \tilde{h}(\alpha) &= f + \nabla f^T(p^U + \alpha(p^B - p^U)) + \tfrac{1}{2}(p^U + \alpha(p^B - p^U))^T H(p^U + (\alpha(p^B - p^U)) \\ &= f + \nabla f^T p^U + \tfrac{1}{2}(p^U)^T H p^U + \alpha\left[(p^B - p^U)^T(\nabla f + H p^U)\right] + \tfrac{1}{2}\alpha^2(p^B - p^U)^T H(p^B - p^U). \end{aligned}$$

Derivando esta expressão com relação a α, obtemos que

$$\tilde{h}'(\alpha) = (p^B - p^U)^T(\nabla f + H p^U) + \alpha(p^B - p^U)^T H(p^B - p^U)$$

$$\begin{aligned}
&\leq (p^B - p^U)^T(\nabla f + H p^U + H(p^B - p^U)) \\
&= (p^B - p^U)^T(\nabla f + H p^B) \\
&= (p^B - p^U)^T(\nabla f - H H^{-1}\nabla f) \\
&= 0.
\end{aligned}$$

□

Como já mencionado anteriormente, p^B será tomado como solução do subproblema de minimização na região de confiança quando $\left\|p^B\right\| \leq \Delta$. Por outro lado, quando $\left\|p^B\right\| > \Delta$ o caminho $p^D(\tilde{\tau})$ interceptará o limite da região de confiança em exatamente um ponto, e, do fato de m ser decrescente ao longo do referido caminho, tomaremos como solução do subproblema este ponto de interseção. Assim, computamos $\tilde{\tau}$ resolvendo a seguinte equação quadrática:

$$\left\|p^U + (\tilde{\tau} - 1)(p^B - p^U)\right\|^2 = \Delta^2.$$

2.3.3 Escalamento Afim

O desempenho de um algoritmo de otimização depende, entre outros fatores, da estrutura e do escalamento dos problemas a serem resolvidos. Funções mal escaladas são aquelas em que pequenas mudanças em uma ou mais coordenadas de um ponto de seu domínio produzem grandes variações na imagem da função, quando comparadas com variações de outras coordenadas. Em contrapartida, problemas bem escalados são menos sensíveis a pequenas variações em um ponto do domínio, facilitando, muitas vezes, o processo de minimização almejado.

Topologicamente, um sintoma do mal escalamento de uma função f é que seu minimizador x^* costuma ficar em um vale estreito, de modo que os contornos de f próximos de x^* tendem ser elipses altamente excêntricas. Tal complicação configura um dos principais motivos pelos quais foram desenvolvidas regiões de confiança elípticas, nas quais os eixos são mais curtos nas direções mais sensíveis e mais largos nas direções menos sensíveis. O cerne das regiões de confiança elípticas consiste em considerar uma matriz diagonal com elementos diagonais positivos, a ser denominada por D, e encontrar um passo p adequado que minimize um determinado modelo da função e satisfaça

$$\|Dp\| \leq \Delta,$$

onde Δ é o raio da região de confiança. Quando $f(x)$ é altamente sensível ao valor da i-ésima componente x_i, tomamos o elemento diagonal correspondente da matriz D, d_{ii}, grande. O valor de d_{ii} deve ser próximo de zero para componentes x_i menos sensíveis.

O escalamento das regiões de confiança necessita que ajustemos algumas variáveis anteriormente abordadas. Definimos, portanto,

$$\tilde{p} = Dp,$$

e passamos a considerar o modelo quadrático

$$\tilde{m}_k(\tilde{p}) = f_k + \nabla f_k^T D^{-1}\tilde{p} + \tfrac{1}{2}\tilde{p}^T D^{-1} H_k D^{-1}\tilde{p}.$$

Desta forma, o subproblema (34) passa a ser reinterpretado como

$$\min_{\tilde{p}\in\mathbb{R}^n} \tilde{m}_k(\tilde{p}) \text{ sujeito a } \|\tilde{p}\| \leq \Delta_k.$$

Isto evidencia que o desenvolvimento da teoria e dos algoritmos para regiões de confiança elípticas pode ser derivado das regiões de confiança tradicionais (baseadas em bolas euclidianas). Para tanto basta substituir p por $\tilde{p}$, ∇f_k por $D^{-1}\nabla f_k$, H_k por $D^{-1}H_kD^{-1}$. Disto e de (41), temos que a solução p_k^* do problema

$$\arg\min_{p\in\mathbb{R}^n} f_k + \nabla f_k^T p \quad \text{sujeito à} \quad \|Dp\| \leq \Delta_k.$$

satisfaz

$$Dp_k^* = -\frac{\Delta_k}{\|D^{-1}\nabla f_k\|} D^{-1}\nabla f_k.$$

Portanto,

$$p_k^* = -\frac{\Delta_k}{\|D^{-1}\nabla f_k\|} D^{-2}\nabla f_k,$$

isto é, um múltiplo positivo do vetor $-D^{-2}\nabla f_k$. Por este motivo, a direção de máxima descida escalada a ser considerada é $\tilde{d}_k^C = -D_k^{-2}\nabla f_k$, onde D_k é a matriz de escalamento na k-ésima iteração.

Assim, o ponto de Cauchy escalado é definido por

$$p_k^C = \tilde{\tau}_k \tilde{d}_k, \tag{45}$$

onde

$$\tilde{\tau}_k = \arg\min_{\tau>0} \left\{ m_k(\tau\tilde{d}_k) \mid \left\| \tau D_k d_k^C \right\| \leq \Delta_k \right\},$$

e Δ_k é o raio de região de confiança na k-ésima iteração.

De forma similar ao realizado na Proposição 2.3.1, garantimos que

$$\tilde{\tau}_k = \min\left\{ \frac{\left\| D_k^{-1}\nabla f_k \right\|^2}{\nabla f_k^T D_k^{-2} H_k D_k^{-2}\nabla f_k}, \frac{\Delta_k}{\left\| D_k^{-1}\nabla f_k \right\|} \right\}. \tag{46}$$

A escolha da matriz D é, evidentemente, fundamental para a eficácia das regiões de confiança elípticas. Quando nos reportamos a métodos de regiões de confiança aplicados a sistemas não lineares com restrições de caixa, as matrizes afim-escala propostas por Coleman e Li [6] ganham destaque pela maneira pela qual conseguem lidar com as respectivas restrições.

Consideremos uma caixa n-dimensional $\Omega \subseteq \mathbb{R}^n$, definida por

$$\Omega = \left\{ x \in \mathbb{R}^n \;\middle|\; l \leq x \leq u \right\},$$

e o problema de otimização restrito à caixa

$$\min_{x \in \Omega} f(x), \tag{47}$$

onde $f : \Omega \to \mathbb{R}$ é definida por $f(x) = \frac{1}{2}\|F(x)\|^2$.

As condições de otimalidade de primeira ordem do problema (47), são apresentadas no teorema a seguir.

Teorema 2.3.3. *Seja $x^* \in \Omega$ um minimizador local do problema de otimização (47), e f continuamente diferenciável em uma vizinhança aberta de x^*. Então,*

$$\begin{cases} \nabla f_i(x^*) = 0 & \text{se } l_i < x_i^* < u_i, \\ \nabla f_i(x^*) \leq 0 & \text{se } x_i^* = u_i, \\ \nabla f_i(x^*) \geq 0 & \text{se } x_i^* = l_i. \end{cases} \tag{48}$$

Demonstração. Verifiquemos as três possibilidades de ordem das coordenadas de x^* em relação aos limites da caixa:

a) $l_i < x_i^* < u_i$.

Suponha, para obter contradição, que $\nabla f_i(x^*) \neq 0$.

Defina $p_i = -\nabla f_i(x^*)$, e note que $p_i \nabla f_i(x^*) = -\nabla f_i^2(x^*) < 0$.

Como ∇f é contínuo próximo de x^*, suas componentes também são contínuas nesta vizinhança. Logo, existe um escalar $T > 0$ tal que

$$p_i \nabla f_i(x^* + tp) < 0, \tag{49}$$

para todo $t \in [0, T]$.

Assim, do Teorema de Taylor, segue que, para todo $\bar{t} \in (0, T]$,

$$f_i(x^* + \bar{t}p) = f_i(x^*) + \bar{t} p_i \nabla f_i(x^* + tp),$$

para algum $t \in (0, \bar{t})$.

Disto e de (49), $f_i(x^* + \bar{t}p) < f_i(x^*)$ para todo $\bar{t} \in (0, T]$. Escolhamos $\bar{t}$ suficientemente pequeno, de tal forma que

$$l_i < x_i^* + \bar{t} p_i < u_i.$$

Perceba que encontramos uma direção a partir de x_i^* que reduz o valor da função dentro da caixa, o que leva a concluir que x^* não é um minimizador local — uma contradição. Portanto, $\nabla f_i(x^*) = 0$.

b) $x_i^* = u_i$

Suponha, para obter contradição, que $\nabla f_i(x^*) > 0$.

Ora, como x_i^* encontra-se em um limite superior da caixa, podemos tomar a direção de máxima descida $p_i = -\nabla f_i(x^*) < 0$ e, assim, para um escalar positivo $\bar{t}$ suficientemente pequeno, garantimos que

$$l_i < x_i^* + \bar{t}p_i < u_i.$$

Perceba que encontramos uma direção a partir de x_i^* que reduz o valor da função dentro da caixa, o que leva a concluir que x^* não é um minimizador local — uma contradição. Portanto, $\nabla f_i(x^*) \leq 0$.

c) $x_i^* = l_i$

Suponha, para obter contradição, que $\nabla f_i(x^*) < 0$.

Ora, como x_i^* encontra-se em um limite inferior da caixa, podemos tomar a direção de máxima descida $p_i = -\nabla f_i(x^*) > 0$ e, assim, para um escalar positivo $\bar{t}$ suficientemente pequeno, garantimos que

$$l_i < x_i^* + \bar{t}p_i < u_i.$$

Perceba que encontramos uma direção a partir de x_i^* que reduz o valor da função dentro da caixa, o que leva a concluir que x^* não é um minimizador local — uma contradição. Portanto, $\nabla f_i(x^*) \geq 0$.

□

Pontos que satisfazem as condições estabelecidas em (48) são chamados de pontos estacionários do problema de otimização (47). Tendo em vista que essas condições são escritas em termos das coordenadas do vetor gradiente de f, dado um ponto $x \in Int(\Omega)$, a matriz afim-escala $D(x)$ é definida, estrategicamente, por

$$\begin{aligned} D(x) &= \begin{bmatrix} |v_1(x)|^{-1/2} & & \\ & \ddots & \\ & & |v_n(x)|^{-1/2} \end{bmatrix}, \\ &= diag\left(|v_1(x)|^{-1/2}, \cdots, |v_n(x)|^{-1/2}\right). \end{aligned} \tag{50}$$

onde

$$v_i(x) = \begin{cases} u_i - x_i, & \text{se } \nabla f_i(x) < 0 \text{ e } u_i < \infty, \\ x_i - l_i, & \text{se } \nabla f_i(x) \geq 0 \text{ e } l_i < -\infty, \\ -1, & \text{se } \nabla f_i(x) < 0 \text{ e } u_i = \infty, \\ 1, & \text{se } \nabla f_i(x) \geq 0 \text{ e } l_i = -\infty. \end{cases} \tag{51}$$

Desta forma, fica evidente que $D(x)$ não está definido na fronteira de Ω. Em contrapartida, $D(x)^{-1}$ pode ser estendido continuamente até lá, já que os elementos de sua diagonal principal não possuem restrições que os impeçam de serem nulos. Assim,

$$D(x)^{-1} = \begin{bmatrix} |v_1(x)|^{1/2} & & \\ & \ddots & \\ & & |v_n(x)|^{1/2} \end{bmatrix},$$

para todo $x \in \Omega$, em que $v_i(x)$ é dado por (51) e, por consequência, $D(x)^{-1}$ é contínuo para todo $x \in \Omega$.

Devido à maneira como é definida a matriz afim-escala, quando o ponto atual está próximo de um limite da caixa e distante de um ponto estacionário do problema (47) a estratégia de região de confiança Coleman-Li costuma forçar um grande passo, um recurso geralmente necessário para métodos de otimização eficientes e práticos. A Figura 4 evidencia o caso em que o iterado está próximo da fronteira da caixa, e o eixo maior da elipse propicia que passos maiores sejam explorados caso haja necessidade.

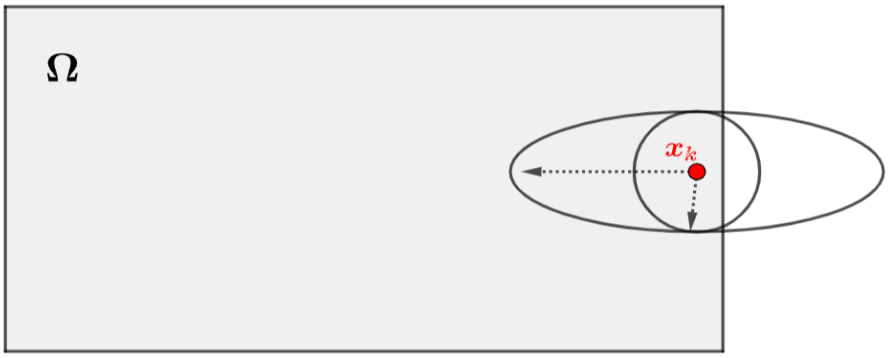

Figura 4 – Possibilidade de passo maior para uma região de confiança elíptica em um problema com restrição de caixa.

As condições de otimalidade de primeira ordem expressas por (48) são equivalentes a um sistema não linear de equações envolvendo a matriz afim-escala, cujo tratamento é, frequentemente, mais simples de ser feito.

Lema 2.3.4. *Seja $x \in \Omega$. Então $D(x)^{-1}\nabla f(x) = 0$ se, e somente se, x é um ponto estacionário do problema (47).*

Demonstração. Suponha inicialmente que $D(x)^{-1}\nabla f(x) = 0$. Note que, por $D(x)^{-1}$ ser uma matriz diagonal, tal hipótese equivale a

$$|v_i(x)|^{1/2} \cdot \nabla f_i(x) = 0,$$

para $i = 1, \ldots, n$. Neste contexto, é necessário que analisemos as possibilidades de posicionamento das componentes de x em relação aos limites da caixa:

a) $l_i < x_i < u_i$.
Neste caso, por (51), necessariamente, $|v_i(x)|^{1/2} \neq 0$. Consequentemente, $\nabla f_i(x) = 0$.

b) $x_i = l_i$.
Neste caso, temos duas situações para analisar. Se $|v_i(x)|^{1/2} \neq 0$, então, necessariamente, $\nabla f_i(x) = 0$. Por outro lado, se $|v_i(x)|^{1/2} = 0$, então, $v_i(x) = 0 = x_i - l_i$. Logo, de (51), temos que $\nabla f_i(x) \geq 0$.
Portanto, $\nabla f_i(x) \geq 0$.

c) $x_i = u_i$.
Neste caso, temos duas situações para analisar. Se $|v_i(x)|^{1/2} \neq 0$, então, necessariamente, $\nabla f_i(x) = 0$. Por outro lado, se $|v_i(x)|^{1/2} = 0$, então, $v_i(x) = 0 = u_i - x_i$. Logo, de (51), temos que $\nabla f_i(x) < 0$.
Portanto, $\nabla f_i(x) \leq 0$.

Isto completa a primeira parte da prova.

Considere, agora, x um ponto estacionário de f. Assim, é necessário que analisemos as possibilidades de posicionamento das componentes de x em relação aos limites da caixa:

a) $l_i < x_i < u_i$
Então, por hipótese, $\nabla f_i(x) = 0$. Consequentemente, $|v_i(x)|^{1/2} \cdot \nabla f_i(x) = 0$.

b) $x_i = u_i$
Então, por hipótese, $\nabla f_i(x) \leq 0$. Se $\nabla f_i(x) = 0$, o resultado é confirmado automaticamente. Assuma, portanto, $\nabla f_i(x) < 0$. Assim, de (51), $v_i(x) = u_i - x_i = 0$, e, por consequência, $|v_i(x)|^{1/2} \cdot \nabla f_i(x) = 0$.

c) $x_i = l_i$
Então, por hipótese, $\nabla f(x)_i \geq 0$. Se $\nabla f_i(x) = 0$, o resultado é confirmado automaticamente. Assuma, portanto, $\nabla f_i(x) > 0$. Assim, de (51), $v_i(x) = x_i - l_i = 0$, e, por consequência, $|v_i(x)|^{1/2} \cdot \nabla f_i(x) = 0$.

Isto completa a prova. □

Como D^{-1} é não singular em $Int(\Omega)$ e vale (37), a partir do Lema 2.3.4 podemos concluir que se $x^* \in Int(\Omega)$ é um ponto estacionário do problema de otimização (47) e $J(x^*)$ é não singular, então $F(x^*) = 0$, isto é, x^* é solução do sistema não linear. Além disto, x^* é um mínimo global de f. Para maiores detalhes, este e outros resultados teóricos relativos à abordagem afim-escala são encontrados nas bibliografias [1, 6, 12].

3 MÉTODO QUASE-NEWTON PARA RESOLUÇÃO DE SISTEMAS QUADRADOS DE EQUAÇÕES NÃO LINEARES COM RESTRIÇÕES DE CAIXA

Neste capítulo adaptaremos as ideias discutidas anteriormente a fim de propor um método de região de confiança escalada Quase-Newton para sistemas não lineares quadrados sujeitos a restrições de caixa. A partir desta proposição discutiremos resultados teóricos que fundamentam o referido algoritmo, e que garantem, sob determinadas condições, taxa de convergência superlinear.

3.1 DESCRIÇÃO DO MÉTODO

Neste estudo desejamos resolver o seguinte sistema de equações:

$$F(x) = 0,\ x \in \Omega, \tag{52}$$

onde

$$\Omega = \left\{ x \in \mathbb{R}^n \;\middle|\; l \leq x \leq u \right\}.$$

$l_i \in \mathbb{R} \cup \{-\infty\}$, $u_i \in \mathbb{R} \cup \{+\infty\}$, para $i = 1, \ldots, n$. Assumimos que $F : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^n$ é continuamente diferenciável num conjunto aberto que contém Ω.

Este sistema pode ser convertido em um problema de mínimos quadrados. Para isto, basta definir $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ por

$$f(x) = \frac{1}{2}\|F(x)\|^2. \tag{53}$$

Expandindo a expressão que define $f(x)$, verificamos que

$$\begin{aligned} f(x) &= \frac{1}{2}\sum_{i=1}^{n} F_i^2(x) \\ &= \frac{1}{2}\left(F_1^2(x) + F_2^2(x) + \ldots + F_n^2(x)\right). \end{aligned}$$

Assim, a i-ésima coordenada do gradiente de f pode ser escrita como

$$\nabla f_i(x) = \sum_{j=1}^{n} \frac{\partial F_j(x)}{\partial x_i} F_j(x),$$

em que $i = 1, ..., n$. Portanto,

$$\nabla f(x) = J(x)^T F(x).$$

Com estas definições básicas, passaremos a considerar o problema de otimização associado a (52)

$$\min_{x \in \Omega} f(x), \tag{54}$$

onde $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ é dada por (53).

Note que toda solução de (52) é, também, solução de (54), porém é possível que existam minimizadores locais de (54) que não sejam solução de (52) (veja a Figura 5). Desta forma, quando consideramos o problema de otimização, a resolução de (52) depende diretamente da identificação de soluções globais de (54). Conforme destacado no final do Capítulo 2, quando $x^* \in Int(\Omega)$ é um ponto estacionário de (54) e $J(x^*)$ é não singular, então x^* é um minimizador global do referido problema.

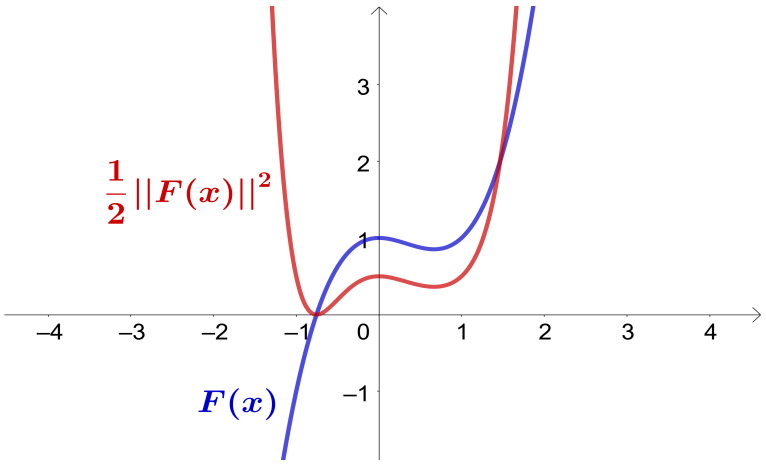

Figura 5 – Comparação entre F e f para a função $F(x) = x^3 - x^2 + 1$

Para métodos de regiões de confiança afim-escala que buscam resolver (52), a cada iteração o passo p_k costuma ser obtido por meio da resolução do subproblema

$$\text{Minimizar } m_k^{GN}(p) \text{ sujeito a } \|D_k p\| \leq \Delta,$$

onde $m_k^{GN}(p) = \frac{1}{2}\|F_k + J_k p\|^2$, Δ é o raio da região de confiança e $D_k \equiv D(x_k)$ é a matriz afim-escala, simétrica definida positiva, definida no capítulo anterior pela equação (50). Aqui, m_k^{GN} denota o modelo Gauss-Newton na k-ésima iteração, e foi utilizado em grande parte dos estudos de métodos de pontos interiores que utilizam regiões de confiança [1, 12].

Pensando em diminuir a carga de trabalho do método que estamos propondo, substituímos a matriz Jacobiana de F por uma matriz de atualização secante B. Desta forma, assim como realizado por Yuan e Wei [35] e por Zeng e Fu [36], passamos a considerar um modelo Quase-Newton e, por consequência, a resolução do subproblema

$$\text{Minimizar } m_k^{QN}(p) \text{ sujeito a } \|D_k p\| \leq \Delta, \tag{55}$$

onde

$$m_k^{QN}(p) = \tfrac{1}{2}\|F_k + B_k p\|^2$$

$$= \tfrac{1}{2}\|F_k\|^2 + F_k^T B_k p + \tfrac{1}{2} p^T B_k^T B_k p.$$

A direção Quase-Newton que minimiza $m_k^{QN}(p)$ será aquela que satisfaz

$$F_k + B_k p = 0.$$

Evidentemente, sob a condição de B_k ter posto completo,

$$p = -B_k^{-1} F_k.$$

De forma sucinta, com exceção da substituição de J por B (por consequência, a substituição de um modelo quadrático Gauss-Newton por um modelo quadrático Quase-Newton, e de uma matriz afim-escala por uma matriz baseada na abordagem afim-escala), o funcionamento do método que propomos é similar ao método de região de confiança afim-escala desenvolvido em [1]. Dada a iteração interna atual, definimos a região de confiança escalada com raio $\Delta > 0$ e calculamos o passo corrente a partir da resolução do subproblema (55). A redução prevista do modelo quadrático e a redução real da função objetivo proporcionadas por este passo são calculadas. Se a redução real for grande o suficiente quando comparada à redução prevista, um múltiplo viável deste passo é então calculado e somado ao iterado atual x_k, gerando um ponto de teste. Se a redução real gerada pelo passo viável for suficientemente grande em relação à redução real do passo (inteiro), o ponto de teste será aceito e uma nova iteração será iniciada. Caso contrário, o raio da região de confiança é reduzido e repetem-se os procedimentos para identificar um passo que satisfaça as condições de redução real da função. Veremos a seguir, com maiores detalhes, os procedimentos relativos à escolha do raio de região de confiança, da matriz de escalamento, da viabilidade do passo escolhido, das condições de decréscimo do modelo em relação à função objetivo utilizada e da escolha da matriz de atualização secante.

Tendo em vista as vantagens obtidas na convergência dos métodos de região de confiança, comprovadas por Zhang e Wang [37], consideraremos o raio de região de confiança a cada iteração como $\Delta_k = c^t \eta_k$, onde $0 < c < 1$, t é um número inteiro não negativo e $\eta_k = \min\{\max\{1, \|F_k\|\}, 10^2\}$. Além disto, dado um iterado $x_k \in Int(\Omega)$, a matriz de escalamento $D(x_k)$ da região de confiança que utilizaremos é definida por

$$\begin{aligned} D(x_k) &= \begin{bmatrix} |v_1(x_k)|^{-1/2} & & \\ & \ddots & \\ & & |v_n(x_k)|^{-1/2} \end{bmatrix}, \\ &= diag\left(|v_1(x_k)|^{-1/2}, \cdots, |v_n(x_k)|^{-1/2}\right). \end{aligned} \tag{56}$$

onde

$$v_i(x_k) = \begin{cases} u_i - (x_k)_i, & \text{se } (B_k^T F_k)_i < 0 \text{ e } u_i < \infty, \\ (x_k)_i - l_i, & \text{se } (B_k^T F_k)_i \geq 0 \text{ e } l_i < -\infty, \\ -1, & \text{se } (B_k^T F_k)_i < 0 \text{ e } u_i = \infty, \\ 1, & \text{se } (B_k^T F_k)_i \geq 0 \text{ e } l_i = -\infty. \end{cases} \tag{57}$$

Como consequência desta definição, temos que, diferentemente de D_k^{-1}, D_k não está definida na fronteira de Ω.

Assim como os métodos de região de confiança afim-escala, buscamos desenvolver um método de pontos interiores para caixas, ou seja, um método no qual todos os iterados gerados pelo algoritmo estão no interior de Ω. Seja, então, $x_k \in Int(\Omega)$ um iterado do algoritmo e seja $p \in \mathbb{R}^n$ uma direção. Defina

$$\lambda(p) = \arg\max\{t \geq 0 \mid x_k + tp \in \Omega\}. \tag{58}$$

Por consequência, se $\lambda(p) > 1$, então $x_k + p \in Int(\Omega)$; por outro lado, se $\lambda(p) \leq 1$, então $x_k + p \notin Int(\Omega)$ (veja a Figura 6). Em Otimização, tal fato implica na necessidade de reduzir o tamanho do passo, a fim de garantir que o processo de minimização esteja limitado a trabalhar com pontos viáveis.

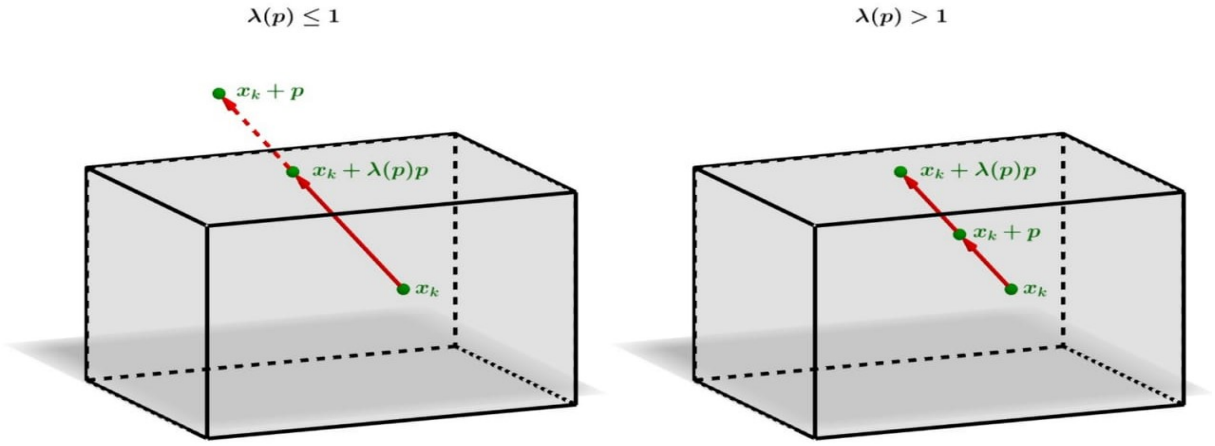

Figura 6 – Comportamento de λ em uma caixa em $\mathbb{R}^3$.

Também, dado $\theta \in (0,1)$, definimos

$$\xi(p) = \begin{cases} 1 & \text{se } \lambda(p) > 1, \\ \max\{\theta, 1 - \|p\|\}\lambda(p) & \text{caso contrário.} \end{cases} \tag{59}$$

Defina, agora,

$$\alpha(p) = \xi(p)p. \tag{60}$$

Desta forma, $x_k + \alpha(p) \in Int(\Omega)$. Logo, independentemente de $x_k + p$ estar fora ou na fronteira da caixa, $\alpha(p)$ é um múltiplo interior de p.

Determinado o passo $p \in \mathbb{R}^n$, definimos a redução real e prevista na k-ésima iteração, denotadas, respectivamente, por $ared_k(p)$ e $pred_k(p)$, pelas expressões

$$ared_k(p) = f(x_k) - f(x_k + p),$$
$$pred_k(p) = m_k^{QN}(0) - m_k^{QN}(p).$$

Um critério inicial para que o passo seja aceito é de que a redução real fornecida por p seja grande o suficiente em comparação com a redução prevista. Neste caso, a direção p deverá satisfizer a condição

$$r_k(p) = \frac{ared_k(p)}{pred_k(p)} = \frac{f(x_k) - f(x_k + p)}{m_k^{QN}(0) - m_k^{QN}(p)} \geq \rho,$$

em que $\rho \in (0, 1)$ é dado. Além disto, a redução real fornecida pelo múltiplo viável $\alpha(p)$ deve ser suficientemente grande em relação à redução fornecida por p, isto é, deve valer a condição

$$ared_k(\alpha(p_k)) \geq \gamma ared_k(p_k),$$

onde $\gamma \in (0, 1)$ é dado.

Com relação a escolha das matrizes de atualização secante, propomos que seja tomado $B_0 = J_0$, onde J_0 é a matriz Jacobiana de F no ponto inicial x_0 calculada pelo método das Diferenças Finitas. Esta escolha é justificada pelo fato de estarmos baseando o desenvolvimento do método na abordagem afim-escala de Coleman e Li [6], cuja construção da matriz de escalamento depende fundamentalmente do vetor $\nabla f = J^T F$. Como propomos aproximar J por uma matriz de atualização B, para evitar um passo inicial que se distancie da solução procurada e/ou se aproxime inapropriadamente das fronteiras da caixa, é importante que iniciemos o processo iterativo do método com uma matriz B muito próxima da Jacobiana exata. Neste contexto, tomar $B_0 = I$, como é feito usualmente em sistemas não lineares irrestritos [35, 36], implica em assumir a aproximação $\nabla f_0 \approx F_0$ e correr o risco de promover um escalamento inadequado que encaminhe o processo iterativo à falha.

Para fins de aproximação do modelo com a função, dado um passo viável $s_k = \alpha(p_k)$ exigiremos que as matrizes B_{k+1}, aproximações de J_{k+1}, satisfaçam a equação secante

$$B_{k+1} s_k = y_k,$$

em que $y_k = F_{k+1} - F_k$ e $s_k = x_{k+1} - x_k$. Conforme visto no capítulo anterior, as atualizações BFGS, SR1 e Broyden são tradicionais fórmulas que satisfazem a referida equação, e serão, portanto consideradas em nosso método.

Diferentemente da resolução de sistemas não lineares irrestritos, a atualização BFGS aqui proposta não contemplará a condição de curvatura $y_k^T s_k > 0$, que almeja preservar a característica de B_{k+1} ser sempre definida positiva ao longo do processo iterativo. Tal fato ocorre,

primeiramente, devido à escolha $B_0 = J_0$, que não oferece a garantia de iniciar o processo iterativo com uma matriz definida positiva. Além disto, a cada iteração o vetor $B_k^T F_k$ é utilizado como base para indicar a direção a ser percorrida pelo método, a partir da construção da matriz de escalamento D_k. Neste sentido, manter $B_{k+1} = B_k$ por várias iterações seguidas mostra-se uma alternativa que, além de impedir que B_k possa capturar informações importantes com relação à curvatura da função f, encaminha o distanciamento dos vetores $\nabla f_{k+1} = J_{k+1}^T F_{k+1}$ e $B_{k+1}^T F_{k+1}$. Estes fatores tendem conduzir a sequência gerada pelo método a locais inadequados da caixa, comprometendo a resolução do problema.

Sem a condição de curvatura, a atualização BFGS perde a característica de B_{k+1} ficar definida positiva e, por consequência, de estar bem definida a cada iteração. Além disto, quando a viabilidade do problema limitar a escolha do passo a vetores muito pequenos, $\|s_k\|$ e $\|y_k\|$ poderão ser próximos de zero e dificultar a utilização das fórmulas secantes, principalmente as atualizações BFGS e Broyden. Assim, a formulação das atualizações secantes utilizadas será dada por:

a) Atualização Simétrica de Posto Um (SR1), definida por

$$B_{k+1}^{SR1} = \begin{cases} B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}, & \text{se } \left|s_k^T (y_k - B_k s_k)\right| \geq \tilde{\varepsilon} \|s_k\| \|y_k - B_k s_k\|; \\ B_k^{SR1}, & \text{caso contrário.} \end{cases} \tag{61}$$

Aqui, $\tilde{\varepsilon} > 0$ representa uma tolerância que evita a existência de possíveis singularidades na atualização.

b) Atualização Broyden-Fletcher-Goldfarb-Shanno (BFGS), definida por

$$B_{k+1}^{BFGS} = \begin{cases} B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, & \text{se } \left|y_k^T s_k\right|, \left|s_k^T B_k s_k\right| \geq \tilde{\varepsilon}; \\ B_{k+1}^{SR1}, & \text{caso contrário.} \end{cases} \tag{62}$$

c) Atualização Broyden, definida por

$$B_{k+1}^{BRO} = \begin{cases} B_k + \frac{(y_k - B_k s_k) s_k^T}{s_k^T s_k}, & \text{se } s_k^T s_k \geq \tilde{\varepsilon}; \\ B_{k+1}^{SR1}, & \text{caso contrário.} \end{cases} \tag{63}$$

Com tais informações, torna-se possível definir o algoritmo principal do método proposto.

Algoritmo 4: Método STRQN (*Scaled Trust-Region Quasi-Newton*)

Entrada: $x_0 \in Int(\Omega)$; $B_0 \in \mathbb{R}^{n\times n}$; $c, \theta, \rho, \gamma \in (0,1)$;

Saída: $x^* \in \Omega$;

início

$k \longleftarrow 0$ e $t \longleftarrow 0$;

repita

1) Calcule F_k e D_k;

2) Resolva o sistema linear

$$B_k p_k^{QN} = -F_k;$$

3) Calcule $\Delta_k = c^t \eta_k$, onde $\eta_k = \min\left\{\max\left\{1, \|F_k\|\right\}, 10^2\right\}$.

se $\left\|D_k p_k^{QN}\right\| \leq \Delta_k$ **então**

Defina $p_k^t = p_k^{QN}$;

senão

Calcule a solução p_k^t do subproblema de região de confiança (55);

fim

4) Calcule

$$r_k(p_k^t) = \frac{ared_k(p_k^t)}{pred_k(p_k^t)},$$

e $ared_k(\alpha(p_k^t))$, onde $\alpha(p_k^t)$ é calculado conforme (60);

se $r_k(p_k^t) < \rho$ *ou* $ared_k(\alpha(p_k^t)) < \gamma ared_k(p_k^t)$ **então**

$t \longleftarrow t+1$;

Volte para o passo 3;

senão

$x_{k+1} \longleftarrow x_k + \alpha(p_k^t)$;

fim

5) Calcule B_{k+1} de acordo com a atualização secante escolhida;

6) $k \longleftarrow k+1$ e $t \longleftarrow 0$;

até $F_k = 0$;

fim

$x^* \longleftarrow x_k$.

Convém ressaltar que o procedimento "passo 3 - passo 4 - passo 3" será aqui denominado como *círculo interno*.

Na próxima seção apresentaremos os resultado teóricos referentes ao Algoritmo 4.

3.2 RESULTADOS DE CONVERGÊNCIA

A fim de provar os resultados de convergência do Algoritmo 4, consideraremos as seguintes hipóteses:

(H1) O conjunto de nível $L(x_0) := \{x \in \Omega \mid f(x) \leq f(x_0)\}$ é limitado para todo $x_0 \in \Omega$;

(H2) $F(x)$ é duas vezes continuamente diferenciável em um conjunto aberto, limitado e convexo Ω_1 contendo Ω;

(H3) A seguinte relação vale

$$\left\| [J_k - B_k]^T F_k \right\| = O(\left\| p_k^t \right\|);$$

(H4) Para todo k os valores singulares da matriz B_k, a serem denotados por σ_i^k, são uniformemente limitados, de modo que existem constantes $0 < M_0 < M$ que satisfazem

$$M_0 \leq \sigma_i^k \leq M, \ i = 1, \ldots, n;$$

Importante observar que de *(H4)* podemos concluir que, para todo k, B_k é não singular e

$$M_0 \leq \|B_k\| \leq M \qquad \text{e} \qquad \tfrac{1}{M} \leq \|B_k^{-1}\| \leq \tfrac{1}{M_0}.$$

Assim, para todo k, $\left\| B_k^T B_k \right\| \leq M^2$. Além disto, de *(H2)*, podemos concluir que existe $M_1 > 0$ tal que, para todo k, $\left\| J_k^T J_k \right\| \leq M_1$.

Partindo das hipóteses iniciais, desenvolveremos alguns resultados teóricos auxiliares para garantir a convergência da sequência gerada pelo Algoritmo 4. O lema a seguir definirá a ordem das diferenças entre a redução real da função e a redução prevista pelo modelo utilizado na k-ésima iteração, sendo fundamental para a demonstração de resultados posteriores.

Lema 3.2.1. *Suponha que valem* (H1) – (H4). *Seja p_k^t a solução de (55), e seja $\{x_k\}$ a sequência gerada pelo Algoritmo 4. Então*

$$\left| ared_k(p_k^t) - pred_k(p_k^t) \right| = O(\left\| p_k^t \right\|^2).$$

Demonstração. Pelo Teorema de Taylor [27], temos que

$$F(x_k + p_k^t) = F(x_k) + J(x_k) p_k^t + O(\left\| p_k^t \right\|^2).$$

Disto e de *(H1) – (H4)*, deduzimos que

$$\begin{aligned} \left| ared_k(p_k^t) - pred_k(p_k^t) \right| &= \left| f(x_k) - f(x_k + p_k^t) - m_k(0) + m_k(p_k^t) \right| \\ &= \frac{1}{2} \left| \left\| F_k + B_k p_k^t \right\|^2 - \left\| F(x_k + p_k^t) \right\|^2 \right| \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\left| \left\|F_k + B_k p_k^t\right\|^2 - \left\|F_k + J_k p_k^t + O(\left\|p_k^t\right\|^2)\right\|^2 \right| \\
&= \left| F_k^T (B_k - J_k) p_k^t + \frac{1}{2}(p_k^t)^T (B_k^T B_k - J_k^T J_k)(p_k^t) + O(\left\|p_k^t\right\|^4) \right| \\
&\leq \left\|(B_k - J_k)^T F_k\right\| \left\|p_k^t\right\| + O(\left\|p_k^t\right\|^2) + O(\left\|p_k^t\right\|^4) \\
&= O(\left\|p_k^t\right\|^2).
\end{aligned}
$$

□

A fim de obter garantia de que a solução do subproblema de região de confiança (55) diminuirá o valor do modelo na k-ésima iteração, um resultado clássico necessário para métodos de região de confiança é a limitação inferior da redução prevista ($pred_k(p_k^t) > 0$). Desta forma, quando a razão entre $ared_k(p_k^t)$ e $pred_k(p_k^t)$ for maior que zero, garantimos que $ared_k(p_k^t) > 0$, isto é, que p_k^t fornece um decréscimo para a função f. É o que nos mostrará o lema a seguir.

Lema 3.2.2. *Se p_k^t é solução de (55), então*

$$
pred_k(p_k^t) \geq \frac{1}{2}\left\|D_k^{-1} B_k^T F_k\right\| \min\left\{ c^t \eta_k, \frac{\left\|D_k^{-1} B_k^T F_k\right\|}{\left\|D_k^{-1} B_k^T B_k D_k^{-1}\right\|} \right\}.
$$

Demonstração. Como p_k^t é solução de (55), para todo $\beta \in [0,1]$, vale que

$$
\begin{aligned}
pred_k(p_k^t) &\geq pred_k\left(-\beta \frac{\Delta_k}{\left\|D_k^{-1} B_k^T F_k\right\|} D_k^{-2} B_k^T F_k \right) \\
&= \frac{1}{2}\left(\|F_k\|^2 - \left\| F_k - \frac{\beta \Delta_k}{\left\|D_k^{-1} B_k^T F_k\right\|} B_k D_k^{-2} B_k^T F_k \right\|^2 \right) \\
&\geq \beta \Delta_k \left\| D_k^{-1} B_k^T F_k \right\| - \frac{1}{2}\beta^2 \Delta_k^2 \left\| D_k^{-1} B_k^T B_k D_k^{-1} \right\|. \qquad (64)
\end{aligned}
$$

Como em (64) temos uma expressão quadrática com concavidade voltada para baixo, vale que

$$
pred_k(p_k^t) \geq \max_{0 \leq \beta \leq 1} \left[\beta \Delta_k \left\| D_k^{-1} B_k^T F_k \right\| - \frac{1}{2}\beta^2 \Delta_k^2 \left\| D_k^{-1} B_k^T B_k D_k^{-1} \right\| \right]. \qquad (65)
$$

Derivando a quadrática acima em relação a β e igualando a zero, identificamos que o maximizador da quadrática em questão ocorre para

$$
\beta = \frac{\left\| D_k^{-1} B_k^T F_k \right\|}{\Delta_k \left\| D_k^{-1} B_k^T B_k D_k^{-1} \right\|}. \qquad (66)
$$

Este maximizador é alcançado por passos Quase-Newton, já que nesse caso garantimos que

$$
0 \leq \beta \leq \frac{\left\| D_k^{-1} B_k^T F_k \right\|}{\left\| D_k^{-1} B_k^T B_k D_k^{-1} \right\| \left\| D_k p_k^t \right\|}
$$

$$\leq \frac{\left\|D_k^{-1}B_k^T F_k\right\|}{\left\|D_k^{-1}B_k^T B_k p_k^t\right\|}$$
$$= \frac{\left\|D_k^{-1}B_k^T F_k\right\|}{\left\|D_k^{-1}B_k^T F_k\right\|} = 1.$$

Disto e da substituição de (66) em (65), concluímos que

$$pred_k(p_k^t) \geq \frac{1}{2}\left\|D_k^{-1}B_k^T F_k\right\| \min\left\{\Delta_k, \frac{\left\|D_k^{-1}B_k^T F_k\right\|}{\left\|D_k^{-1}B_k^T B_k D_k^{-1}\right\|}\right\}$$
$$= \frac{1}{2}\left\|D_k^{-1}B_k^T F_k\right\| \min\left\{c^t\eta_k, \frac{\left\|D_k^{-1}B_k^T F_k\right\|}{\left\|D_k^{-1}B_k^T B_k D_k^{-1}\right\|}\right\}.$$

□

Além da garantia de redução do modelo ao longo do processo iterativo, a convergência do Algoritmo 4 depende da finitude da quantidade de círculos internos realizados a cada iteração (para que o algoritmo esteja bem definido), e da convergência da sequência $\{f_k\}$. Abordaremos, nos próximos dois lemas, resultados relativos aos aspectos supracitados.

Lema 3.2.3. *Assuma que valem* (H1) – (H4) *e considere x_k um iterado da sequência gerada pelo Algoritmo 4 tal que $F_k \neq 0$. Então existe um número inteiro não negativo $t < \infty$, tal que as condições $r_k(p_k^t) \geq \rho$ e $ared_k(\alpha(p_k^t)) \geq \gamma ared_k(p_k^t)$ relativas ao passo 4 do referido algoritmo são satisfeitas, isto é, na k-ésima iteração o círculo interno termina em um número finito de passos.*

Demonstração. Suponha, para obter contradição, que o Algoritmo 4 realiza o círculo interno infinitamente na k-ésima iteração, de forma que $t \to \infty$, $r_k(p_k^t) < \rho$ e $c^t \to 0$.

Evidentemente, existe $\varepsilon > 0$ tal que $\|F_k\| \geq \varepsilon$. Disto e de *(H4)*, $D_k^{-1}B_k^T F_k \neq 0$.

Tomemos, agora, p_k^t de forma que $\left\|D_k p_k^t\right\| \leq \Delta_k$, onde $\Delta_k = c^t\eta_k \to 0$, quando $t \to \infty$. Dos Lemas 3.2.1 e 3.2.2 temos então que

$$\left|r_k(p_k^t) - 1\right| = \frac{\left|ared_k(p_k^t) - pred_k(p_k^t)\right|}{\left|pred_k(p_k^t)\right|} \leq \frac{2O(\left\|p_k^t\right\|^2)}{\Delta_k\left\|D_k^{-1}B_k^T F_k\right\|}.$$

Como $\left\|D_k^{-1}B_k^T F_k\right\| > 0$ e $\left\|p_k^t\right\| \to 0$, quando $t \to \infty$, concluímos que $\left|r_k(p_k^t) - 1\right| \to 0$. Assim, existe $t_0 > 0$ tal que para todo $t > t_0$,

$$r_k(p_k^t) \geq \rho.$$

Além disto, do fato de x_k pertencer a $Int(\Omega)$, temos que existe $t_1 > 0$ (dependente de x_k) tal que para todo $t > t_1$ vale $p_k^t = \alpha(p_k^t)$ e, consequentemente, $ared_k(\alpha(p_k^t)) \geq \gamma ared_k(p_k^t)$.

Defina $t_2 = \max\{t_0, t_1\}$. Então para todo $t > t_2$ as condições para encerrar o círculo interno na k-ésima iteração são satisfeitas. Isto contradiz a hipótese inicial da demonstração, e completa a prova. □

Lema 3.2.4. *Assuma que valem* (H1) – (H4)*, e que* $\{x_k\}$ *é a sequência gerada pelo Algoritmo 4. Então* $\{x_k\} \subseteq L(x_0)$*. Além disto,* $\{f_k\}$ *converge.*

Demonstração. Da definição do referido algoritmo, temos que

$$r_k(p_k^t) \geq \rho > 0.$$

Disto e do Lema 3.2.2 segue que $ared_k(p_k^t) > 0$. Portanto,

$$ared_k(\alpha(p_k^t)) \geq \gamma ared_k(p_k^t) > 0.$$

Assim,

$$f(x_{k+1}) \leq f(x_k) \leq \cdots \leq f(x_0).$$

Isto comprova que $\{x_k\} \subseteq L(x_0)$. Considerando que, para todo k, $f_k \geq 0$, concluímos que $\{f_k\}$ converge. A prova está completa. □

Lema 3.2.5. *Seja* $\{x_k\}$ *a sequência gerada pelo Algoritmo 4, e assuma que vale* (H1)*. Então, para todo k existe* $\chi_D > 0$ *tal que*

$$\left\|D_k^{-1}\right\| \leq \chi_D$$

Demonstração. De *(H1)* sabemos que existem $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ e $b = (b_1, \ldots, b_n) \in \mathbb{R}^n$, tais que, para todo k,

$$a_i \leq (x_k)_i \leq b_i, \qquad i = 1, \ldots, n. \tag{67}$$

Considere Z_1 o conjunto de índices i para os quais l_i é finito e Z_2 o conjunto de índices para os quais u_i é finito. Escolheremos, por conveniência, a e b que satisfazem simultaneamente (67) e as condições

$$\begin{aligned} b_i - a_i &\geq 1, && i = 1, \ldots, n \\ a_i &\leq l_i, && \forall i \in Z_1 \\ u_i &\leq b_i, && \forall i \in Z_2. \end{aligned}$$

Como D_k^{-1} é diagonal com elementos positivos, temos que $\|D_k^{-1}\|$ equivale ao maior autovalor da referida matriz, isto é, ao seu maior elemento. Disto e de (57), o resultado segue automaticamente com

$$\chi_D = \max_i \{b_i - a_i\}.$$

□

Teorema 3.2.6. *Assuma que valem* (H1) – (H4)*, e que* $\{x_k\}$ *é a sequência gerada pelo Algoritmo 4. Então, ou o algoritmo termina em um número finito de iterações (identifica uma solução viável do sistema não linear), ou gera uma sequência infinita* $\{x_k\}$ *tal que*

$$\lim_{k\to\infty} \left\| D_k^{-1} B_k^T F_k \right\| = 0.$$

Demonstração. Assuma que o algoritmo não termina após um número finito de iterações. Precisamos, então, mostrar que a relação

$$\lim_{k\to\infty} \left\| D_k^{-1} B_k^T F_k \right\| = 0$$

é verdadeira.

A fim de obter contradição, assumiremos, então, que existe $\varepsilon_1 > 0$ e uma subsequência infinita $\{x_{k_j}\}$ tal que $\left\| D_{k_j}^{-1} B_{k_j}^T F_{k_j} \right\| \geq \varepsilon_1$, Naturalmente, do fato de $\{x_{k_j}\} \subset Int(\Omega)$, $D_{k_j}^{-1}$ é não singular para todo k_j. Assim, $B_{k_j}^T F_{k_j} \neq 0$, e existe $\varepsilon_2 > 0$ tal que $\left\| F_{k_j} \right\| \geq \varepsilon_2$. Seja $\varepsilon = \min\{\varepsilon_1, \varepsilon_2, 1\}$. Seja $\tilde{K} = \left\{ k \mid \left\| D_k^{-1} B_k^T F_k \right\| \geq \varepsilon, \|F_k\| \geq \varepsilon \right\}$.

Usando da definição do Algoritmo 4, de *(H4)*, do Lema 3.2.2 e do Lema 3.2.5, obtemos

$$\begin{aligned}
\sum_{k\in\tilde{K}} [f(x_k) - f(x_{k+1})] &\geq \gamma \sum_{k\in\tilde{K}} ared_k(p_k^{t_k}) \\
&\geq \gamma \sum_{k\in\tilde{K}} \rho \cdot pred_k(p_k^{t_k}) \\
&\geq \sum_{k\in\tilde{K}} \gamma\rho \cdot \frac{\varepsilon}{2} \min\left\{ c^{t_k}\varepsilon, \frac{\varepsilon}{(\chi_D M)^2} \right\},
\end{aligned}$$

onde t_k é o maior valor de t obtido no círculo interno no ponto iterativo x_k.

Pelo Lema 3.2.4, sabemos que $\{f_k\}$ é convergente. Logo,

$$\sum_{k\in\tilde{K}} \gamma\rho \cdot \frac{\varepsilon}{2} \min\left\{ c^{t_k}\varepsilon, \frac{\varepsilon}{(\chi_D M)^2} \right\} < \infty.$$

Então, $t_k \to \infty$ quando $k \to \infty$ e $k \in \tilde{K}$. Assim, podemos assumir que $t_k \geq 1$ para todo $k \in \tilde{K}$.

De acordo com a determinação de p_k, $k \in \tilde{K}$, no círculo interno, a solução $\tilde{p}_k$ correspondente ao seguinte subproblema

$$\min_p m_k^{QN}(p) = \frac{1}{2}\|F_k + B_k p\|^2, \text{ sujeito a } \|D_k p\| \leq c^{t_k - 1}\eta_k,$$

não é aceitável, já que teremos

$$\frac{ared_k(\tilde{p}_k)}{pred_k(\tilde{p}_k)} < \rho \tag{68}$$

e/ou

$$ared_k(\alpha(\tilde{p}_k)) < \gamma ared_k(\tilde{p}_k). \tag{69}$$

Suponha que vale (69). Sabemos que $t_k \to \infty$ quando $k \to \infty$, $k \in K$. Disto e do fato de $\{x_k\} \subset Int(\Omega)$, teremos $\tilde{p}_k = \alpha(\tilde{p}_k)$ para t_k suficientemente grande. Isto contradiz (69).

Suponha, agora, que vale (68). Ora, do Lema 3.2.2 sabemos que

$$pred_k(\tilde{p}_k) \geq \frac{\varepsilon}{2} \min\left\{c^{t_k-1}\varepsilon, \frac{\varepsilon}{(\chi_D M)^2}\right\}.$$

Além disto, do Lema 3.2.1 temos que

$$|ared_k(\tilde{p}_k) - pred_k(\tilde{p}_k)| = O(\|\tilde{p}_k\|^2) = O(c^{2(t_k-1)}).$$

Portanto,

$$\left|\frac{ared_k(\tilde{p}_k) - pred_k(\tilde{p}_k)}{pred_k(\tilde{p}_k)}\right| = \left|\frac{ared_k(\tilde{p}_k)}{pred_k(\tilde{p}_k)} - 1\right| \leq \frac{O(c^{2(t_k-1)})}{\frac{\varepsilon}{2}\min\left\{c^{t_k-1}\varepsilon, \frac{\varepsilon}{(\chi_D M)^2}\right\}}.$$

Como $t_k \to \infty$ quando $k \to \infty$, obtemos

$$\frac{ared_k(\tilde{p}_k)}{pred_k(\tilde{p}_k)} \to 1, k \in \tilde{K}.$$

Isto contradiz (68), e completa a prova. □

Corolário 3.2.7. *Seja $\{x_k\}$ a sequência gerada pelo Algoritmo 4 e suponha que valem* (H1) – (H4). *Sob tais condições, se um ponto limite x^* da sequência pertence a $Int(\Omega)$, então*

$$F(x^*) = 0.$$

Demonstração. Para o caso em que a sequência gerada pelo algoritmo é finita, o resultado segue de modo imediato. Suponha, então, que ela é infinita. Do Teorema 3.2.6 temos que é gerada uma sequência infinita $\{x_k\}$ que satisfaz

$$\lim_{k\to\infty} \left\| D_k B_k^T F_k \right\| = 0.$$

Como, para todo k, x_k e x^* pertencem a $Int(\Omega)$, D_k é não singular. Portanto,

$$\lim_{k\to\infty} \left\| B_k^T F_k \right\| = 0.$$

Finalmente, de *(H4)* e da continuidade de F, concluímos então que

$$\lim_{k\to\infty} \|F_k\| = \|F(x^*)\| = 0.$$

Portanto, $F(x^*) = 0$. □

A tese deste corolário configura um importante resultado teórico relativo ao algoritmo que propomos neste estudo, já que direciona à resolução do problema central que motiva o desenvolvimento deste trabalho. Porém, até agora, ele está limitado a convergência de uma subsequência do Algoritmo 4. Desta forma necessitamos ainda provar, sob condições adequadas, que a sequência (inteira) converge a uma solução do problema, e sob qual velocidade esta convergência ocorre.

Teorema 3.2.8. *Seja* $\{x_k\}$ *a sequência gerada pelo Algoritmo 4 e* $x^* \in Int(\Omega)$ *um ponto limite da sequência. Suponha que valem* (H1) – (H4)*, que* $\lim_{k\to\infty} \|x_k - x_{k+1}\| = 0$ *e que o conjunto de raízes interiores de* F *em* $L(x_0)$ *é finito. Então,* $\lim_{k\to\infty} x_k = x^*$ *e* $F(x^*) = 0$*.*

Demonstração. Seja $\Lambda \subseteq L(x_0)$ o conjunto dos pontos limite de $\{x_k\}$ que pertencem a $Int(\Omega)$, e X^* o conjunto de raízes interiores de F em $L(x_0)$. Tome $z \in \Lambda$. Então, existe um subconjunto $K \subseteq \mathbb{N}$ tal que $\lim_{k\in K} x_k = z$. Pelo Corolário 3.2.7 e pela continuidade de F, temos que

$$0 = \lim_{k\in K} F(x_k) = F(z),$$

ou seja, $\Lambda \subseteq X^*$, e, pela hipótese de finitude de X^*, Λ é finito.

Seja $\Lambda = \{z_1, \ldots, z_m\}$. Mostraremos que $m = 1$. Defina

$$\delta = \min\left\{ \left\|z_i - z_j\right\| \;\middle|\; i \neq j;\ i,j = 1,\ldots,m \right\} > 0.$$

Como z_i são pontos de acumulação de $\{x_k\}$ e $\lim_{k\to\infty} \|x_k - x_{k+1}\| = 0$, existe $k_0 \in \mathbb{N}$ tal que $x_k \in \cup_{i=1}^{m} \mathscr{B}(z_i, \delta/4)$ e $\|x_k - x_{k+1}\| \leq \delta/4$ para todo $k \geq k_0$. Assim, se $x_{k_1} \in \mathscr{B}(z_1, \delta/4)$, com $k_1 \geq k_0$, vale para todo $i \geq 2$ que

$$\begin{aligned}
\left\|z_i - x_{k_1+1}\right\| &= \left\|z_i - z_1 + z_1 - x_{k_1} + x_{k_1} - x_{k_1+1}\right\| \\
&\geq \|z_i - z_1\| - \left\|z_1 - x_{k_1}\right\| - \left\|x_{k_1} - x_{k_1+1}\right\| \\
&\geq \delta - \frac{\delta}{4} - \frac{\delta}{4} \\
&= \frac{\delta}{2},
\end{aligned}$$

e, portanto, necessariamente, $x_{k_1+1} \in \mathscr{B}(z_1, \delta/4)$. Usando o princípio da indução, $x_k \in \mathscr{B}(z_1, \delta/4)$ para todo $k \geq k_1$ e

$$\lim_{k\to\infty} x_k = z_1.$$

Portanto, $\Lambda = \{z_1\}$ e, da continuidade de F, $F(z_1) = 0$. □

A hipótese de finitude das raízes interiores de F no conjunto de nível $L(x_0)$ não configura uma hipótese muito exigente, uma vez que para valores de x_0 suficientemente próximos de uma raiz interior isolada ela acaba sendo satisfeita automaticamente.

Ao provar sob as devidas condições que a sequência $\{x_k\}$ gerada pelo Algoritmo 4 converge a uma solução do problema de estudo, podemos concluir que, se $x^* \in Int(\Omega)$, a partir de uma determinada iteração, os passos Quase-Newton calculados pelo Algoritmo 4 não exigirão a redução do raio de região de confiança através do círculo interno e estarão automaticamente dentro da caixa Ω. É o que mostraremos no Lema 3.2.10, porém, antes, necessitamos do resultado auxiliar apresentado a seguir.

Lema 3.2.9. *Seja $z \in Int(\Omega)$. Então existe $r' > 0$ e $\mathscr{D}_1 > 0$ tal que $\|D(x_k)\| < \mathscr{D}_1$ para todo $x_k \in \mathscr{B}(z, r') \subset Int(\Omega)$.*

Demonstração. Como z é ponto interior de Ω, existe $r' \in (0, 1]$ tal que $\mathscr{B}(z, 2r') \subset Int(\Omega)$. Defina $\mathscr{D}_1 = \sqrt{\frac{1}{r'}}$. Logo, para todo $x_k \in \mathscr{B}(z, r')$,

$$|(x_k)_i - l_i|, |u_i - (x_k)_i| > r',$$

para $i = 1, ..., n$.

Assim, de (56), segue que $\|D(x_k)\| < \sqrt{\frac{1}{r'}} = \mathscr{D}_1$. □

Lema 3.2.10. *Suponha que valem* (H1) – (H4). *Considere p_k^t a solução do subproblema de região de confiança (55), gerada pelo Algoritmo 4, e $x^* \in Int(\Omega)$ um ponto limite de $\{x_k\}_{k \in K}$, onde K é uma sequência infinita de índices, tal que*

$$\lim_{k \in K} \|F_k\| = 0.$$

Então existe $k_0 \in \{0, 1, 2, \dots\}$, tal que, para todo $k \geq k_0$, $k \in K$,

$$x_{k+1} = x_k + p_k^{QN}, \tag{70}$$

onde $p_k^{QN} = -B_k^{-1} F_k$.

Demonstração. Da continuidade de F,

$$\lim_{k \in K} \|F_k\| = \|F(x^*)\| = 0. \tag{71}$$

Como $\{x_k\}_{k \in K}$ é limitada e $x^* \in Int(\Omega)$, pelo Lema 3.2.9 existe $\mathscr{D}_1 > 0$ tal que $\|D_k\| \leq \mathscr{D}_1$ para todo $k \in K$. Disto e de *(H4)* vale que

$$\left\| D_k p_k^{QN} \right\| \leq \mathscr{D}_1 \left\| -B_k^{-1} F_k \right\| \leq \frac{\mathscr{D}_1}{M_0} \|F_k\|.$$

Desta inequação e de (71),

$$\lim_{k \in K} \left\| D_k p_k^{QN} \right\| = 0. \tag{72}$$

Ora, $\{x_k\} \subset Int(\Omega)$ e, dessa forma, D_k é não singular para todo $k \in K$. Assim, por (72),

$$\lim_{k\in K} \left\| p_k^{QN} \right\| = 0.$$

Disto, de (58) e de (59), segue que existe $\tilde{k} \in \{0,1,2,\dots\}$, tal que $\left\| D_k p_k^{QN} \right\| \leq \Delta_k$ e $\lambda(p_k^{QN}) > 1$ para $k \geq \tilde{k}$, $k \in K$. Consequentemente, $p_k^0 = p_k^{QN}$, $\xi(p_k^{QN}) = 1$ e $\alpha(p_k^{QN}) = p_k^{QN}$ para $k \geq \tilde{k}$, $k \in K$.

Perceba que para $k \geq \tilde{k}$, $k \in K$, vale, portanto, que

$$\begin{aligned} pred_k(p_k^0) &= m_k^{QN}(0) - m_k^{QN}(p_k^0), \\ &= -F_k^T B_k p_k^0 - \frac{1}{2}(p_k^0)^T B_k^T B_k (p_k^0), \\ &= O(\left\| p_k^0 \right\|). \end{aligned}$$

Disto e do Lema 3.2.1, segue, portanto, que

$$\lim_{k\in K} \left| r_k(p_k^0) - 1 \right| = \lim_{k\in K} \frac{\left| ared_k(p_k^0) - pred_k(p_k^0) \right|}{\left| pred_k(p_k^0) \right|} = 0,$$

o que implica que

$$\lim_{k\in K} r_k(p_k^0) = 1.$$

Portanto, existe $k_0 \geq \tilde{k}$, $k_0 \in \{0,1,2,\dots\}$, tal que $r_k(p_k^0) \geq \rho$, para todo $k \geq k_0$, $k \in K$. Assim, para k suficientemente grande, $p_k^0 = p_k^{QN}$ é aceito no passo 4 do Algoritmo 4, e a atualização de x_{k+1} é dada por (70). Isto completa a prova. □

Antes de provar a convergência superlinear da sequência gerada pelo Algoritmo 4, apresentaremos duas hipóteses adicionais. A primeira delas diz respeito à continuidade Lipschitz de $J(x)$, e é necessária para que resultados auxiliares já apresentados no Capítulo 2 possam ser utilizados. Já a segunda, trata-se da condição de Dennis-Moré, que implica que o passo Quase-Newton converge para o passo Newton, em direção e magnitude.

(H5) Para todo $x, y \in \Omega_1$ tem-se que existe $L > 0$ tal que

$$\|J(x) - J(y)\| \leq L\|x - y\|; \tag{73}$$

(H6) B_k aproxima $J(x^*)$ na direção $\alpha(p_k^t)$, isto é,

$$\lim_{k\to\infty} \frac{\|[B_k - J(x^*)](x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0.$$

Com estas hipóteses estamos aptos para provar, sob condições específicas, a convergência superlinear da sequência gerada pelo Algoritmo 4. Tal taxa de convergência é típica dos métodos Quase-Newton, e mostra-se mais rápida do que a convergência linear, porém mais lenta que a convergência quadrática. Assim como acontece na convergência quadrática do método Newton, faz-se necessário que assumamos agora a não singularidade de J em um ponto limite x^*.

Teorema 3.2.11. *Seja $\{x_k\}$ a sequência gerada pelo Algoritmo 4 e $x^* \in Int(\Omega)$ um ponto limite da sequência. Assuma que valem* (H1) – (H6)*, que $\|x_k - x_{k+1}\| \to 0$, que o conjunto de raízes interiores de F em $L(x_0)$ é finito e que $J(x^*)$ é não singular. Então x_k converge superlinearmente para x^*. Além disto, $F(x^*) = 0$.*

Demonstração. Do Teorema 3.2.8 já sabemos que $x_k \to x^* \in Int(\Omega)$ e que $F(x^*) = 0$. Resta apenas provarmos que esta convergência é superlinear.

Do Lema 3.2.10 sabemos que existe $k_0 \in \{0, 1, 2, \dots\}$, tal que, para todo $k \geq k_0$, $x_{k+1} = x_k + p_k^0$, onde $p_k^0 = p_k^{QN}$. Assim,

$$\begin{aligned} 0 &= B_k p_k^0 + F_k, \\ &= [B_k - J(x^*)]p_k^0 + F_k + J(x^*)p_k^0, \end{aligned}$$

para todo $k \geq k_0$. Portanto,

$$-F_{k+1} = [B_k - J(x^*)]p_k^0 + (-F_{k+1} + F_k + J(x^*)p_k^0),$$

onde $p_k^0 = x_{k+1} - x_k$.

Agora, do Lema 2.2.5, segue que, para $k \geq k_0$,

$$\begin{aligned} \frac{\|F_{k+1}\|}{\left\|p_k^0\right\|} &\leq \frac{\left\|[B_k - J(x^*)]p_k^0\right\|}{\left\|p_k^0\right\|} + \frac{\left\|-F_{k+1} + F_k + J(x^*)p_k^0\right\|}{\left\|p_k^0\right\|} \\ &\leq \frac{\|[B_k - J(x^*)](x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} + \frac{L\|x_{k+1} - x_k\| \max\{\|x_{k+1} - x^*\|, \|x_k - x^*\|\}}{\|x_{k+1} - x_k\|}, \\ &\leq \frac{\|[B_k - J(x^*)](x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} + L(\|x_{k+1} - x^*\| + \|x_k - x^*\|). \end{aligned}$$

Assim, de $x_k \to x^*$ e de *(H6)*,

$$\lim_{k \to \infty} \frac{\|F_{k+1}\|}{\left\|p_k^0\right\|} = 0.$$

Agora, pelo Lema 2.2.10, existem $c_1 \geq 0$, $k' \geq 0$, tal que

$$\|F_{k+1}\| \geq c_1 \|x_{k+1} - x^*\|,$$

para todo $k \geq k'$. Portanto,

$$\begin{aligned} 0 &= \lim_{k\to\infty} \frac{\|F_{k+1}\|}{\left\|p_k^0\right\|} \geq \lim_{k\to\infty} \frac{c_1\|x_{k+1}-x^*\|}{\left\|p_k^0\right\|}, \\ &\geq \lim_{k\to\infty} \frac{c_1\|x_{k+1}-x^*\|}{\|x_{k+1}-x^*\|+\|x_k-x^*\|}, \\ &= \lim_{k\to\infty} \frac{c_1 e_k}{e_k+1}, \end{aligned}$$

onde $e_k = \frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}$. Isto implica que

$$\lim_{k\to\infty} e_k = 0,$$

o que completa a prova de convergência superlinear do algoritmo. □

Com este resultado de convergência superlinear do Algoritmo 4, concluímos os resultados teóricos que fundamentam o método proposto. No próximo capítulo analisaremos o desempenho do método STRQN em sistemas não lineares quadrados sujeitos a restrições de caixa e verificaremos, desta forma, se os resultados recentemente provados podem ser observados na prática.

4 RESULTADOS NUMÉRICOS

Os resultados numéricos constituem um dos principais meios para análise de desempenho de um método computacional, uma vez que permitem quantificar eficiência e robustez a partir de sua execução em um conjunto de problemas. A comparação destes resultados com os de outros métodos numéricos de mesma classe possibilita identificar em quais aspectos o método em questão é mais vantajoso em relação aos demais, indicando os tipos de problemas mais recomendados para sua utilização.

Neste capítulo serão detalhados os resultados numéricos do Algoritmo 4, proposto anteriormente, baseado no seu desempenho em trinta problemas práticos disponíveis na literatura [1, 3, 11, 15, 19, 28]. Tais problemas incluem equações relativas aos processos de combustão, equilíbrio químico, cinemática de robôs, transferência radiativa — evidenciando a importância da análise destes resultados numéricos para as áreas de engenharias e de ciências aplicadas. Neste contexto, foram explorados diversos sistemas restritos representativos de equações não lineares, com diferentes níveis de esparsidade, que foram úteis para testar os métodos considerados na identificação de soluções no interior e na fronteira da região viável Ω.

O conjunto de problemas utilizado, com diferentes valores de dimensão n, está explicitado na Tabela 1. Dos trinta problemas utilizados, dois têm limite inferior infinito e limite superior finito (problemas 28 e 30); seis têm limite inferior finito e limite superior infinito (problemas 16, 17, 19, 20, 21 e 23); e os outros vinte e dois têm limites superior e inferior finitos. Todos os problemas foram escolhidos por serem utilizados com frequência na literatura, principalmente no que tange à obtenção de resultados numéricos de métodos focados em resolver sistemas não lineares quadrados sujeitos a restrições de caixa.

Tabela 1 – Problemas-teste.

Prb	Função	n
1	Himmelblau Function [11]	2
2	Equilibrium Combustion Function [11]	5
3	Powell's Badly Scaled Function [11]	2
4	Ferraris and Tronconi System [11]	2
5	Brown's Almost Linear System [11]	5
6	Robot Kinematics Problem [11]	8
7	Circuit Design Problem [11]	9
8	Series of CSTRs, R=0.995 [11]	2
9	Smith Nonisothermal CSTR [11]	1
10	Discrete Integral Function [1]	50
11	Discrete Boundary Value Function [1]	500
12	Trigonometric Exponential System [3]	5000
13	Troesh Problem [3]	500
14	Trigonometric System [3]	1000
15	Countercurrent Reactor Problem [3]	1000
16	Five Diagonal System [3]	500
17	Seven Diagonal System [3]	500
18	Zero Jacobian Function [28]	2000
19	Chandrasekhar H-Equation, $c = 0.99$ [15]	1000
20	Chandrasekhar H-Equation, $c = 0.9999$ [15]	1000
21	Chandrasekhar H-Equation, $c = 1$ [15]	1000
22	Boundary Value Problem 2 [15]	2500
23	Boundary Value Problem 3 [15]	7500
24	Brent Problem [19]	3000
25	Yamamutra Problem [19]	10000
26	Tridiagonal System [19]	750
27	Extended Wood Problem [19]	200
28	Singular Broyden Problem [19]	500
29	Extended Powell Singular Problem [19]	5000
30	Structured Jacobian Problem [19]	8000

Todos os experimentos realizados foram conduzidos a partir de três iterados iniciais x_0 diferentes para cada um dos trinta problemas da Tabela 1. Em particular, utilizamos $x_0 = l + 0.25\,\kappa\,(u - l)$, com $\kappa = 1, 2, 3$, para os problemas com limites inferior (l) e superior (u) finitos. Para os demais casos, a fórmula para geração dos pontos iniciais foi a mesma, porém, utilizamos, exclusivamente para este fim, $l_i = -4$ (nos casos em que a i-ésima coordenada era ilimitada inferiormente) e $u_i = 8$ (nos casos em que a i-ésima coordenada era ilimitada superiormente). Estes parâmetros foram os mesmos utilizados no artigo de Francisco, Krejic e Martínez [12]. Quando o ponto inicial gerado era a solução do problema substituímos o respectivo valor de κ por $\kappa + 0.5$.

Convém salientar que as implementações computacionais desenvolvidas neste estudo foram realizadas no *software* Matlab - Versão R2020b, e executadas em um computador com as

seguintes especificações técnicas: Processador *Intel(R) Core*^{TM} i7-8550U; Memória RAM 8GB; Arquitetura 64 *bits*; e Sistema Operacional *Windows 10*.

A fim de melhorar a organização deste capítulo, detalharemos, inicialmente, a maneira como ocorreu a implementação computacional do método STRQN. Em seguida, compararemos três métodos STRQN com diferentes escolhas de atualizações secantes (consequentemente, de passos Quase-Newton). Aquele que tiver obtido maior eficiência e robustez, será utilizado, posteriormente, para a comparação do método proposto neste estudo com os métodos clássicos STRN [1] e PLM [17]. Por fim, analisaremos o desempenho do método STRQN, ressaltando suas potencialidades e fragilidades na resolução de sistemas de equações não lineares sujeitos a restrições de caixa.

4.1 IMPLEMENTAÇÃO COMPUTACIONAL DO MÉTODO STRQN

O método STRQN, estruturado conforme o Algoritmo 4, necessita que definamos alguns parâmetros. Nesta perspectiva, a implementação computacional do método foi realizada com $c = 0.25, \theta = 0.99995$, $\rho = 0.0001$ e $\gamma = 0.001$. Já a matriz B_0 tomada foi a matriz Jacobiana de F em x_0, denotada por J_0, calculada numericamente pelo método de Diferenças Finitas (esta escolha, inclusive, está justificada no Capítulo 3).

No passo 3 do Algoritmo 4, quando o passo Quase-Newton p_k^{QN} não está contido inteiramente na região de confiança escalada, é exigido que encontremos uma direção p_k que satisfaça $\|D_k p_k\| \leq \Delta_k$, onde Δ_k é o raio de região de confiança da iteração corrente. Nestes casos, tomamos p_k como o passo Dogleg p_k^D. Este passo necessita do ponto de Cauchy escalado, que pode ser calculado a partir da substituição de ∇f_k por $B_k^T F_k$ e de H_k por $B_k^T B_k$ nas equações (45) e (46). Desta forma, dada a direção escalada $\tilde{d}_k^C = -D_k^{-2} B_k^T F_k$ e o raio $\Delta_k > 0$, o passo de Cauchy na região de confiança é expresso por

$$p_k^C = -\tilde{\tau}_k \tilde{d}_k^C,$$

onde

$$\tilde{\tau}_k = \min \left\{ \frac{\left\| D_k^{-1} B_k^T F_k \right\|^2}{\left\| B_k D_k^{-2} B_k^T F_k \right\|^2}, \frac{\Delta_k}{\left\| D_k^{-1} B_k^T F_k \right\|} \right\}.$$

Assim, o passo Dogleg p_k^D é computado da seguinte maneira:

$$p_k^D = \begin{cases} \frac{-\Delta_k D_k^{-2} B_k^T F_k}{\left\| D_k^{-1} B_k^T F_k \right\|} & \text{se } \left\| D_k p_k^C \right\| = \Delta_k, \\ p_k^C + (\mu - 1)\left(p_k^{QN} - p_k^C \right) & \text{caso contrário;} \end{cases}$$

em que μ é a solução positiva de

$$\left\| D_k \left(p_k^C + (\mu - 1)\left(p_k^{QN} - p_k^C \right) \right) \right\|^2 = \Delta_k^2.$$

Evidentemente, a escolha estratégica de μ fará com que o passo Dogleg esteja dentro da região de confiança. Para os casos excepcionais em que $\|D_k p_k^C\| < \Delta_k$ e p_k^{QN} não pode ser calculado devido à singularidade de B_k, tomamos $p_k^D = p_k^C$. Garantimos, desta forma, que o passo 4 do Algoritmo 4 estará bem definido. Ressaltamos que o comportamento do passo Dogleg aqui apresentado é similar àquele representado pela Figura 3, com a diferença do formato da região de confiança (passa a ser elíptica, ao invés de uma bola euclidiana).

Se para os métodos de região de confiança afim-escala, que calculam a matriz J a cada iteração, os pontos próximos da fronteira da caixa costumam evidenciar dificuldades no processo iterativo, é natural que para o método STRQN isto também aconteça. Mais ainda, as aproximações de ∇f pelo vetor $B^T F$ no referido método, decisivas para definir as direções a serem percorridas no interior da caixa, podem não ser suficientemente boas, a ponto de provocar um escalamento ruim para as regiões de confiança utilizadas. Neste sentido a sequência gerada pelo Algoritmo 4 pode se aproximar de fronteiras da caixa que não possuem pontos estacionários do problema $\min_{x\in\Omega} f(x)$ por perto, aumentando as chances de ocorrer falhas no processo iterativo. Assim, quando identificado que o iterado está na fronteira da caixa e distante de um ponto estacionário, o algoritmo foi programado a retornar duas iterações e calcular a respectiva matriz Jacobiana, tomando $B_{k-2} = J_{k-2}$. Esta estratégia, embora tenha um custo elevado do ponto de vista computacional, mostra-se uma alternativa para tentar salvar o processo iterativo, e passará a ser denominada de *step-back*.

Além disto, quando no círculo interno de uma iteração é evidenciado que o raio de região de confiança ficou muito próximo de zero (utilizamos a tolerância 10^{-6}), indicando que o modelo quadrático Quase-Newton utilizado não oferece uma boa aproximação para a função objetivo f, reiniciamos a iteração substituindo B_k por J_k. Esta estratégia mostrou-se necessária uma vez que a utilização de um raio excessivamente pequeno implica na escolha de um passo insignificante, podendo indicar ao método falha devido à impossibilidade de obtenção de melhorias no resíduo não linear $\|F_{k+1} - F_k\|$. Salientamos que tanto esta mudança quanto o *step-back* não interferem na convergência do Algoritmo 4, e foram implementadas como forma de superar, na medida do possível, alguns obstáculos enfrentados na prática pelo método STRQN. Ao final do processo iterativo o Algoritmo 4 indica o sucesso ou a falha na identificação de uma raiz de F em Ω, além das seguintes informações: número de iterações realizadas; número de avaliações de F realizadas (sem incluir as avaliações utilizadas para calcular a matriz J); tempo de execução do algoritmo; quantidade de *step-backs* realizados; quantidade de cálculos da matriz J no círculo interno; norma euclidiana do valor corrente de $F(x)$ e de $D(x)^{-1}\nabla f(x)$; número de passos Quase-Newton e número de passos Dogleg utilizados.

O sucesso do Algoritmo 4 é declarado se $\|F_k\| \leq 10^{-6}$, enquanto a falha é declarada nas seguintes ocasiões:

F1. o limite máximo de 5000 iterações foi atingido;

F2. o limite máximo de 10000 avaliações de F foi atingido;

F3. o limite máximo de 3600 segundos de execução do algoritmo foi atingido;

F4. $\|F_{k+1} - F_k\| \leq 100 \cdot eps \cdot \|F_k\|$, isto é, não foi possível obter melhoria para o resíduo não linear;

F5. A sequência está se aproximando de um mínimo local de f em Ω;

F6. Erro ao calcular a matriz D_k, já que a sequência se aproximou de uma fronteira de Ω.

Ressaltamos que o valor $eps = 2^{-52}$, usado em $F4$, é o valor de precisão relativa de ponto flutuante do *software* Matlab. Ainda, pelo fato do Algoritmo 4 substituir, a cada iteração, J_k por uma matriz de atualização B_k, a falha $F5$ não pode ser identificada diretamente. Por isso, a cada iteração, o algoritmo calcula $D_k^{-1} B_k^T F_k$ e avalia seu valor na norma euclidiana. Se tal norma é menor ou igual que a tolerância 10^{-6}, calculamos a matriz J_k e tomamos D_k como a matriz afim-escala em x_k. Neste caso, se $\|D_k^{-1} J_k^T F_k\| \leq 10^{-6}$ a falha é declarada e o algoritmo pára; caso contrário, o algoritmo prossegue com $B_k = J_k$. Já a falha $F6$ somente é declarada após a utilização sem sucesso da estratégia de *step – back*.

Analisaremos, na próxima seção, o desempenho do método STRQN baseado em três atualizações secantes diferentes. Desta forma, buscaremos identificar a melhor forma de atualizar a matriz B, considerando as especificidades de sistemas não lineares com restrições de caixa.

4.2 COMPARAÇÃO ENTRE MÉTODOS STRQN

O método STRQN foi formulado com o propósito principal de evitar o cálculo da matriz J_k a cada iteração, substituindo para tanto J_k por uma matriz de aproximação B_k, obtida por uma atualização secante. Consideraremos, neste estudo, as atualizações SR1, BFGS e Broyden, definidas no capítulo anterior, respectivamente, pelas equações (61), (62) e (63). O valor utilizado para o parâmetro $\tilde{\varepsilon}$ nas referidas expressões foi de 10^{-8}. Além disto, passaremos a denotar o método STRQN com atualização BFGS por STRQN-BFGS; com atualização SR1 por STRQN-SR1; com atualização Broyden por STRQN-Broyden.

A Tabela 3 (Apêndice B) detalha os resultados numéricos da execução dos métodos STRQN-BFGS, STRQN-SR1 e STRQN-Broyden nos problemas da Tabela 1. Como forma de complementar os dados apresentados, na Tabela 4 (Apêndice C) são apresentados maiores detalhes inerentes aos problemas que apresentaram alguma falha ao longo do processo iterativo: a norma de F no ponto inicial $\|F(x_0)\|$, a norma de F no ponto em que ocorreu a falha $\|F(x^*)\|$, e a distância entre este ponto ao ponto estacionário mais próximo $\|D(x^*)^{-1} \nabla f(x^*)\|$ (aqui $D(x^*)$ é a matriz afim-escala em x^*).

Sintetizamos os dados relativos ao tempo de execução dos referidos métodos em cada um dos noventa problemas utilizados no perfil de desempenho (Apêndice A) a seguir:

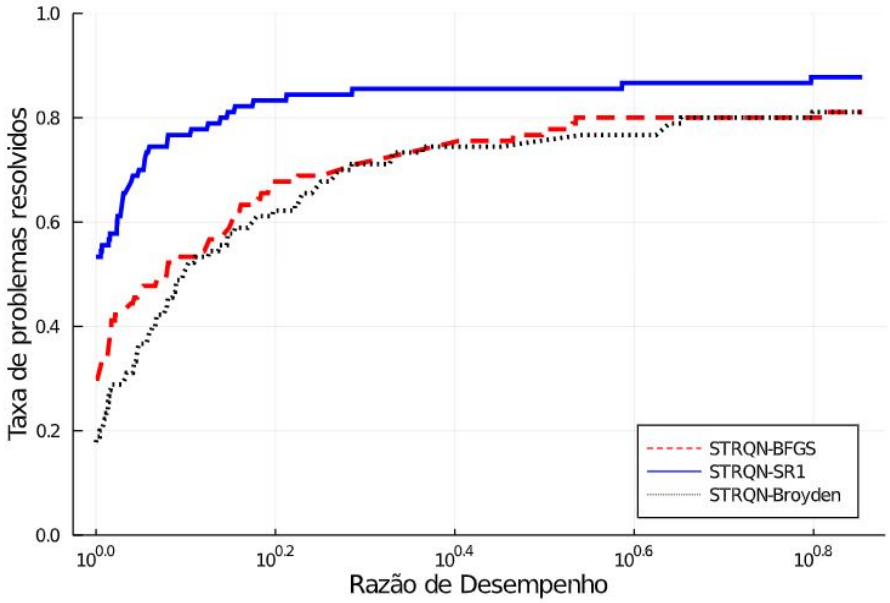

Figura 7 – Perfil de desempenho baseado no tempo de execução dos métodos STRQN-BFGS, STRQN-SR1 e STRQN-Broyden.

Podemos perceber que o método STRQN-SR1 foi o método mais eficiente, uma vez que, considerando a totalidade de problemas resolvidos, ele foi o solucionador mais ágil em aproximadamente 53.33% dos casos, ante 30% do método STRQN-BFGS e 17.78% do método STRQN-Broyden. Além disto, o método STRQN-SR1 foi o mais robusto, resolvendo cerca de 87.78% dos problemas ao qual foi submetido, evidenciando, assim, ser o melhor dentre os três métodos comparados. Os métodos STRQN-BFGS e STRQN-Broyden resolveram, cada um, cerca de 81.11% dos problemas.

Conforme constatado por Nocedal e Wright [27], as matrizes geradas pela fórmula SR1 tendem a ser boas aproximações para a verdadeira matriz Jacobiana de F — muitas vezes melhor do que as aproximações BFGS, já que ela não garante que a matriz de atualização mantenha-se definida positiva e pode, portanto, ser adequada em problemas indefinidos. Além disto, a condição $p_k^T(y_k - B_k p_k) \approx 0$ ocorre com pouca frequência, uma vez que requer certos vetores alinhados de uma maneira específica.

Com relação às falhas apresentadas pelos métodos comparados, fica evidente a predominância das falhas $F4$ e $F6$. Estas falhas são típicas de métodos de pontos interiores, uma vez que eles apresentam dificuldades em dar continuidade ao processo iterativo em regiões muito próximas das fronteiras da caixa. Nesta perspectiva, destacamos que embora $F4$ tenha sido declarada em grande parte dos casos devido à exigência da utilização de passos muito pequenos

para manter a viabilidade das iterações, esta falha também pode ocorrer próxima de pontos estacionários do problema de otimização ou quando o raio de região de confiança ficar muito próximo de zero.

Outro fator negativo enfrentado pelos métodos STRQN foi o trabalho computacional realizado na obtenção do passo Quase-Newton, por meio da resolução de um sistema linear de ordem n. Logo, nos casos em que n é grande o tempo para realização de uma iteração foi elevado. O problema 25 deixa isto nítido: a realização de poucas iterações e círculos internos (avaliações da função F), aliada a um único cálculo de J (no passo inicial), não impediram que os métodos ocupassem um tempo excessivo para a identificação de uma solução do problema.

Nesta perspectiva, a substituição da resolução do sistema linear por uma estratégia de atualizações sucessivas de B_k^{-1} mostra-se um desafio para os métodos STRQN. A dificuldade principal acontece quando, na tentativa de superar alguns obstáculos, o método necessita calcular a matriz Jacobiana e tomar $B_k = J_k$ (no passo inicial, no círculo interno ou, até mesmo, em um *step-back*). Nestes casos, a matriz inversa B_k^{-1} obtida da atualização não corresponderá necessariamente à J_k^{-1}. Desta forma, o passo $p_k = -B_k^{-1}F_k$ não irá, necessariamente, satisfazer a equação $J_k p = -F_k$, podendo configurar uma escolha inadequada na minimização do modelo quadrático Quase-Newton, a ponto de não satisfazer as condições de redução da função esperadas. Por outro lado, calcular a matriz inversa de J_k elevaria de forma súbita o custo computacional do método, principalmente para problemas em que n é grande. Apesar das dificuldades citadas, esta é uma estratégia plausível de estudos futuros que visem aprimorar o método STRQN, uma vez que proporcionará a redução do esforço computacional necessário para obtenção do passo Quase-Newton da $O\left(\frac{n^3}{3}\right)$ para $O(n^2)$.

4.3 COMPARAÇÃO ENTRE OS MÉTODOS STRQN-SR1, STRN E PLM

Analisaremos, nesta seção, o desempenho do método STRQN-SR1 baseado na comparação com dois métodos clássicos para sistemas quadrados de equações não lineares com restrições de caixa: o método de região de confiança afim-escala STRN [1] e o método projetivo PLM [17]. Esta investigação visa averiguar a eficácia do método proposto para a referida classe de problemas, além de identificar os aspectos que o tornam mais e menos atrativos na otimização em caixas.

O método STRN é um método de região de confiança afim-escala com passos Newton para sistemas não lineares quadrados sujeitos a restrições de caixa. Uma de suas características mais atrativas consiste em lidar com os limites da caixa implicitamente, através da utilização estratégica do escalamento afim. Conforme visto no Capítulo 2, este escalamento, por sua vez, é determinado pela proximidade da iteração atual aos limites da caixa, e costuma proporcionar passos largos em iterados próximos destes limites, mas distantes de um ponto estacionário do

problema de otimização.

Da mesma forma que o Algoritmo 4, o método STRN busca encontrar raízes de F minimizando, a cada iteração, um modelo quadrático sujeito à restrição de região de confiança escalada. Desta vez, porém, o modelo considerado é o modelo Gauss-Newton dado por

$$m_k^{GN}(p) = \|F_k + J_k p\|^2. \tag{74}$$

Dada a iteração interna atual x_k, a região de confiança Coleman-Li com raio $\Delta \geq \Delta_{min} > 0$ é definida. O ponto de Cauchy dentro desta região de confiança e o passo Newton são calculados. Se o passo Newton satisfazer a restrição de região de confiança e oferecer uma redução suficientemente grande ao modelo (74) em comparação com à redução proporcionada pelo passo de Cauchy (preservando a viabilidade), um múltiplo viável do passo Newton é definido como um ponto de teste. Caso contrário, um ponto de teste diferente que satisfaça as referidas condições é definido (um múltiplo viável do ponto de Cauchy é naturalmente um ponto de teste admissível). A redução prevista do modelo (74) e a redução real da função objetivo são calculados no ponto de teste como nos métodos de região de confiança padrão. Se a redução real da função for grande o suficiente quando comparada à redução prevista, o ponto de teste é aceito e se inicia uma nova iteração. Caso contrário, o raio da região de confiança é reduzido e repetem-se os procedimentos de identificação de passos que satisfaçam as condições estabelecidas pelo método.

O ponto de Cauchy escalado referente ao modelo Gauss-Newton na k-ésima iteração, p_k^C, pode ser facilmente obtido através da substituição de H_k por $J_k^T J_k$ em (45). Como descrito no parágrafo anterior, ele será usado para verificar se um determinado passo fornece reduções satisfatórias ao modelo (74) em relação ao passo de Cauchy. Para mensurar tal redução, definimos a razão

$$\rho_k^C(p) = \frac{m_k^{GN}(0) - m_k^{GN}(\alpha(p))}{m_k^{GN}(0) - m_k^{GN}(\alpha(p_k^C))}, \tag{75}$$

onde $\alpha(p)$ é dado por (60).

Já para mensurar a redução real da função em relação à redução prevista pelo modelo, utilizamos a razão

$$\rho_k^f(p) = \frac{f(x_k) - f(x_k + \alpha(p))}{m_k^{GN}(0) - m_k^{GN}(\alpha(p))}. \tag{76}$$

O algoritmo a seguir descreve os procedimentos supracitados. Sua implementação computacional pode ser encontrada no sítio eletrônico `http://strscne.de.unifi.it/`.

Algoritmo 5: Método STRN (*Scaled Trust-Region Newton*)

Entrada: $x_0 \in Int(\Omega)$; $\Delta_{min} > 0$; $\Delta > \Delta_{min}$; $\beta_1 \in (0,1]$; $\beta_2, \beta_3 \in (0,1)$, tais que $0 < \beta_2 < \beta_3 < 1$;

Saída: $x^* \in \Omega$;

início

$k \longleftarrow 0$;

repita

1) Calcule F_k, J_k e a matriz afim-escala D_k;

2) Resolva o sistema linear

$$J_k p_k^N = -F_k;$$

3) Calcule $\rho_k^C(p_k^N)$, conforme (75);

se $\left\|D_k p_k^N\right\| \leq \Delta$ *e* $\rho_k^C(p_k^N) \geq \beta_1$ **então**

Defina $p_k = p_k^N$;

senão

Encontre p_k tal que $\|D_k p_k\| \leq \Delta$ e $\rho_k^C(p_k) \geq \beta_1$;

fim

4) Calcule $\rho_k^f(p_k)$, conforme (76);

se $\rho_k^f(p_k) \geq \beta_2$ **então**

$x_{k+1} \longleftarrow x_k + \alpha(p_k)$, onde $\alpha(p_k)$ é calculado conforme (60);

senão

$\Delta = \min\left\{0.25\Delta, 0.5\|D_k p_k\|\right\}$;

Volte para o passo 3;

fim

5) Escolha $\Delta \geq \Delta_{min}$ de forma que

se $\rho_k^f(p_k) \geq \beta_3$ **então**

$\Delta = \max\left\{\Delta_{min}, \Delta, 2\|D_k p_k\|\right\}$;

senão

$\Delta = \Delta$;

fim

6) $k \longleftarrow k+1$;

Vá para o passo 1;

até $F_k = 0$;

fim

$x^* \longleftarrow x_k$.

Para fins de obtenção dos resultados numéricos, os parâmetros utilizados no Algoritmo 5 foram os mesmos utilizados em [1]: o raio inicial da região de confiança foi $\Delta = 1$; $\Delta_{min} = \sqrt{eps} = 2^{-26}$; $\beta_1 = 0.1$; $\beta_2 = 0.25$; e $\beta_3 = 0.75$.

Neste algoritmo o círculo interno é definido como o procedimento "passo 3 - passo 4 - passo 3". No passo 3, quando o passo Newton p_k^N não está contido na região de confiança e/ou a redução fornecida por ele em relação ao passo de Cauchy não é satisfatória, calculamos o passo Dogleg p_k^D definido por

$$p_k^D = \begin{cases} \frac{-\Delta D_k^{-2} J_k^T F_k}{\left\| D_k^{-1} J_k^T F_k \right\|} & \text{se } \left\| D_k p_k^C \right\| = \Delta_k, \\ p_k^C + (\mu - 1)\left(p_k^N - p_k^C\right) & \text{caso contrário,} \end{cases}$$

em que μ é a solução positiva de

$$\left\| D_k \left(p_k^C + (\mu - 1)\left(p_k^N - p_k^C\right) \right) \right\|^2 = \Delta^2.$$

Assim, efetivamos o passo 3 tomando $p_k = p_k^D$ se o passo Dogleg satisfazer $\rho_k^C(p_k^D) \geq \beta_1$; caso contrário, tomamos $p_k = p_k^C$ como forma de garantir tal desigualdade.

Nesta perspectiva, podemos destacar algumas das diferenças entre os métodos STRN e STRQN. Além de tomarem diferentes matrizes para representar a Jacobiana de F (J e B) e, por consequência, diferentes matrizes de escalamento e modelos quadráticos (Gauss-Newton e Quase-Newton), o método STRN exige que os passos escolhidos forneçam uma redução suficiente do modelo em relação à redução proporcionada pelo passo de Cauchy. Desta forma, quando os passos Newton e Dogleg não são considerados bons suficientes, o passo de Cauchy mostra-se uma alternativa vantajosa ao processo iterativo. Outra diferença substancial entre os métodos está na escolha do raio de região de confiança, incluindo a maneira como ocorre a redução do raio no círculo interno.

Diferentemente dos métodos STRN e STRQN, o método PLM é um método projetivo que pode ser aplicado em sistemas não lineares indeterminados com qualquer tipo de restrição convexa. Nele, o passo de Levenberg-Marquardt é calculado em cada iteração e, em seguida, projetado no conjunto viável. Se a norma do sistema no ponto projetado for uma determinada fração da norma do sistema no ponto atual, o ponto Levenberg-Marquardt projetado é tomado como novo iterado. Caso contrário, verifica-se se o ponto Levenberg-Marquardt projetado gera uma direção de descida. Em caso positivo, esta direção é usada para uma busca linear do novo iterado; em caso negativo, o método usa um procedimento de gradiente projetado para reduzir a função objetivo.

O algoritmo a seguir descreve, mais explicitamente, as etapas que constituem o método PLM.

Algoritmo 6: Método PLM (*Projected Levenberg-Marquardt*)

Entrada: $x_0 \in \Omega$; $\mu > 0$; $\beta, \sigma, \gamma \in (0,1)$;

Saída: $x^* \in \Omega$;

início

$k \longleftarrow 0$;

Calcule F_k;

repita

1) Calcule $\mu_k = \mu \|F_k\|^2$;

2) Calcule d_k^U por meio da resolução do sistema

$$\left(J_k^T J_k + \mu_k I\right) d^U = -J_k^T F_k; \tag{77}$$

se $\left\|F(P_\Omega(x_k + d_k^U))\right\| \leq \gamma \|F_k\|$ **então**

$x_{k+1} = P_\Omega(x_k + d_k^U))$;

$k \longleftarrow k+1$;

Volte para o passo "1";

senão

Vá para o passo "3";

fim

3) Calcule $s_k^{LM} = P_\Omega(x_k + d_k^U)) - x_k$;

se $\nabla f_k^T s_k^{LM} \leq -\rho \left\|s_k^{LM}\right\|^p$ **então**

Calcule $t_k = \max\{\beta^i \mid i = 0, 1, \dots\}$, tal que

$$f(x_k + t_k s_k^{LM}) \leq f_k + t_k \sigma \nabla f_k^T s_k^{LM};$$

$x_{k+1} = x_k + t_k s_k^{LM}$;

$k \longleftarrow k+1$;

Volte para o passo "1";

senão

Vá para o passo "4";

fim

4) Calcule $t_k = \max\{\beta^i \mid i = 0, 1, \dots\}$, tal que

$$f(x_k(t_k)) \leq f_k + \sigma \nabla f_k^T (x_k(t_k) - x_k),$$

onde $x_k(t) = P_\Omega(x_k - t\nabla f_k)$;

$x_{k+1} = x_k(t_k)$;

$k \longleftarrow k+1$;

Volte para o passo "1";

até $F_k = 0$;

fim

$x^* \longleftarrow x_k$.

Os parâmetros utilizados para o Algoritmo 6 foram os mesmos utilizados em [17]: $\beta = 0.9$, $\gamma = 0.99995$, $\rho = 10^{-8}$, $p = 2.1$ e $\sigma = 10^{-4}$. Além disto, iniciamos o algoritmo com $\mu_0 = \frac{1}{2}10^{-8}\|F_0\|^2$ e nas iterações subsequentes tomamos $\mu_{k+1} = \min\{\mu_k, \|F_{k+1}\|^2\}$.

Destacamos que tanto para o método STRN quanto para o método PLM, a matriz J_k, exigida em cada iteração, foi calculada pelo método de Diferenças Finitas. Esta escolha possibilita a expansão de sua utilização para problemas cujo processo de derivação não está disponível ou é de difícil obtenção. Ademais, o sucesso destes métodos será declarado se $\|F_k\| \leq 10^{-6}$ (mesmo critério utilizado para o método STRQN). Além das falhas $F1 - F6$ descritas na Seção 4.1, o processo iterativo é interrompido nos métodos STRN e PLM quando

F7. A matriz J_k torna-se singular na k-ésima iteração;

F8. O raio de região de confiança Δ tornou-se muito pequeno ($\Delta < \Delta_{min}$);

F9. O parâmetro t_k do Algoritmo 6 tornou-se muito pequeno ($t_k < 10^{-12}$).

Devido à estruturação dos métodos, a falha $F8$ é aplicável apenas ao método STRN e a falha $F9$ apenas ao método PLM.

A Tabela 2, a seguir, apresenta os dados relativos à execução dos métodos nos problemas da Tabela 1. Neste contexto, o número de iterações (IT), o número de avaliações da função F (AF) e o tempo de execução (TE) dos métodos STRN, PLM e STRQN-SR1 em cada um dos problemas-teste são comparados. Para o método STRQN-SR1 também são apresentados o número de *step-backs* realizados (S), a quantidade de iterações nas quais foi realizado o cálculo da matriz J no círculo interno (C), a quantidade de passos Quase-Newton (QN) e de passos Dogleg (D) tomados. Convém ressaltar que AF não contempla as avaliações de F necessárias para o cálculo numérico da Jacobiana pelo método de Diferenças Finitas, sendo restrita, portanto, às avaliações da função realizadas no círculo interno dos algoritmos.

Tabela 2 – Desempenho dos métodos STRQN-SR1, STRN e PLM.

Prb	κ	STRQN-SR1					STRN			PLM		
		IT	AF	S/C	QN/D	TE	IT	AF	TE	IT	AF	TE
1	1	9	19	0/0	7/2	0.07	6	7	0.06	6	7	0.07
	2	8	13	0/0	7/1	0.05	5	6	0.06	5	8	0.08
	3	9	10	0/0	9/0	0.05	5	6	0.06	5	6	0.06
2	1	37	91	0/1	24/13	0.16	12	13	0.11	205	206	0.67
	2	34	82	0/1	27/7	0.14	14	15	0.12	717	727	1.99
	3	42	90	0/1	29/13	0.18	16	17	0.13	2731	2740	7.22
3	1	6	28	0/1	5/1	0.08	20	25	0.13	15	29	0.09
	2	F4	—	—	—	—	17	21	0.11	18	110	0.13
	3	F4	—	—	—	—	F4	—	—	485	865	0.88
4	1	11	30	0/1	5/6	0.10	5	6	0.07	4	10	0.19
	2	13	24	0/0	12/1	0.07	5	6	0.07	5	6	0.05
	3	9	45	0/2	7/2	0.12	6	8	0.07	5	10	0.06
5	1	10	17	0/0	6/4	0.07	9	10	0.08	10	12	0.08
	2	7	8	0/0	6/1	0.04	7	8	0.08	5	6	0.05
	3.5	13	32	0/1	10/3	0.10	4	5	0.06	5	41	0.07
6	1	F4	—	—	—	—	9	10	0.14	4	6	0.07
	2	14	38	0/2	9/5	0.09	F7	—	—	6	7	0.10
	3	19	36	0/0	12/7	0.10	7	8	0.11	5	6	0.09
7	1	F2	—	—	—	—	F7	—	—	F2	—	—
	2	F2	—	—	—	—	F7	—	—	F2	—	—
	3	F2	—	—	—	—	F7	—	—	F2	—	—
8	1	35	40	0/0	30/5	0.05	19	20	0.08	22	23	0.11
	2	94	597	0/1	9/85	0.54	39	40	0.13	40	41	0.10
	3	108	121	1/1	53/55	0.18	58	59	0.31	59	60	0.20
9	1	4	5	0/0	3/1	0.01	3	4	0.04	3	4	0.03
	2	17	18	0/0	15/2	0.02	13	14	0.06	568	569	0.30
	3	24	25	0/0	22/2	0.03	17	18	0.06	2609	2610	1.32
10	1	6	18	0/1	6/0	0.11	7	8	0.42	5	6	0.31
	2	3	4	0/0	3/0	0.06	3	4	0.25	3	4	0.16
	3	13	22	0/0	12/1	0.12	7	8	0.47	6	7	0.22
11	1	91	118	0/0	17/74	1.60	14	15	5.20	4340	4341	1473.09
	2	2	3	0/0	2/0	0.25	2	3	0.60	2	3	0.55
	3	83	119	0/1	16/67	2.05	14	15	5.98	2013	2014	947.82

Tabela 2 – Desempenho dos métodos STRQN-SR1, STRN e PLM.

Prb	κ	STRQN-SR1					STRN			PLM		
		IT	AF	S/C	QN/D	TE	IT	AF	TE	IT	AF	TE
12	1	36	161	0/2	23/13	187.25	23	26	498.68	F3	—	—
	2	14	49	0/1	9/5	99.39	10	15	294.47	8	11	385.46
	3	23	35	0/1	23/0	93.10	26	29	567.63	F3	—	—
13	1	47	134	0/0	18/29	0.82	9	11	3.15	7	8	1.43
	2	11	12	0/0	8/3	0.33	6	7	2.68	7	8	1.41
	3	42	131	0/0	14/28	0.90	7	8	3.10	7	8	1.64
14	1	35	36	20/0	31/4	18.33	F6	—	—	10	11	8.97
	2	28	29	9/0	22/6	8.99	15	16	34.70	11	12	10.33
	3	35	36	4/0	24/11	5.35	14	15	23.57	10	11	10.75
15	1	21	63	0/2	14/7	2.71	15	17	13.51	20	21	13.55
	2	39	130	0/2	24/15	3.72	17	19	14.32	23	24	16.25
	3	24	64	0/2	19/5	3.01	18	20	16.42	25	26	18.03
16	1	24	70	4/3	20/4	2.20	11	12	4.24	11	12	2.51
	2	F4	—	—	—	—	17	20	5.49	735	736	156.21
	3	F4	—	—	—	—	F6	—	—	632	633	128.43
17	1	18	34	0/1	17/1	11.62	10	11	27.50	7	8	48.18
	2	F4	—	—	—	—	13	14	37.83	10	11	70.54
	3	24	37	0/1	23/1	12.84	15	16	43.15	16	17	109.36
18	1	24	25	0/0	24/0	12.29	22	23	35.58	18	19	72.21
	2	26	27	0/0	26/0	12.33	23	24	36.55	20	21	79.00
	3	27	28	0/0	27/0	12.28	24	25	38.87	23	24	91.15
19	1	7	8	0/0	3/4	3.19	5	6	19.02	4	14	12.83
	2	17	18	0/0	16/1	3.57	12	13	52.81	14	15	47.32
	3	19	62	0/1	5/14	6.71	F4	—	—	8	78	25.20
20	1	5	10	0/0	3/2	3.19	4	6	15.69	4	35	14.05
	2	33	50	0/1	28/5	6.94	22	23	116.13	237	238	787.90
	3	21	44	0/1	7/14	11.32	F8	—	—	11	102	36.47
21	1.5	20	21	0/0	16/4	6.22	15	16	77.55	14	15	45.60
	2	77	93	0/1	12/65	9.19	F7	—	—	F3	—	—
	3	26	27	0/0	12/14	7.10	F4	—	—	F4	—	—
22	1	30	31	0/0	1/29	13.85	5	6	33.46	71	72	574.26
	2	20	21	0/0	1/19	9.88	5	6	32.60	21	22	169.93
	3	10	11	0/0	1/9	6.31	4	5	34.88	4	5	32.70

Tabela 2 – Desempenho dos métodos STRQN-SR1, STRN e PLM.

Prb	κ	STRQN-SR1					STRN			PLM		
		IT	AF	S/C	QN/D	TE	IT	AF	TE	IT	AF	TE
23	1	68	69	0/0	1/67	673.46	6	7	149.39	F3	—	—
	2	180	181	0/0	1/179	1770.83	8	9	136.67	F3	—	—
	3	328	329	0/0	1/327	3240.75	9	10	140.18	F3	—	—
24	1	108	504	0/5	23/85	161.57	21	22	78.61	F3	—	—
	2	260	1667	0/3	13/247	343.92	6	7	23.03	5	6	132.10
	3	27	56	0/1	24/3	46.85	19	20	63.28	F3	—	—
25	1	19	20	0/0	18/1	408.56	19	20	1287.31	F3	—	—
	2	4	5	0/0	4/0	72.95	4	5	163.16	4	5	1642.47
	3	20	21	0/0	19/1	442.88	20	21	1334.14	F3	—	—
26	1	21	57	0/1	16/5	1.32	F1	—	—	12	17	4.39
	2	F2	—	—	—	—	F1	—	—	7	15	8.15
	3	F6	—	—	—	—	9	10	2.94	9	10	3.87
27	1	75	459	0/2	15/60	0.78	33	45	3.11	18	136	1.43
	2	10	34	0/1	8/2	0.28	6	8	0.50	10	66	2.02
	3	33	131	0/3	16/17	0.64	10	11	0.65	11	15	0.92
28	1	30	59	0/1	26/4	0.75	19	20	4.16	18	19	3.35
	2	33	67	0/1	29/4	0.77	16	17	3.19	15	16	3.15
	3	15	27	0/1	15/0	0.66	14	15	2.44	10	11	1.95
29	1	25	59	0/1	19/6	94.71	19	20	198.05	17	18	854.51
	2.5	22	42	0/1	20/2	81.69	18	19	209.22	15	16	748.59
	3	25	59	0/1	19/6	93.20	19	20	168.29	17	18	860.57
30	1	14	26	0/1	14/0	195.74	19	21	622.90	F3	—	—
	2	11	23	0/1	11/0	237.03	14	15	529.38	11	12	3501.92
	3	11	12	0/0	11/0	222.08	11	12	437.12	5	6	1225.51

O perfil de desempenho com relação ao tempo de execução dos solucionadores comparados é apresentado na Tabela 2:

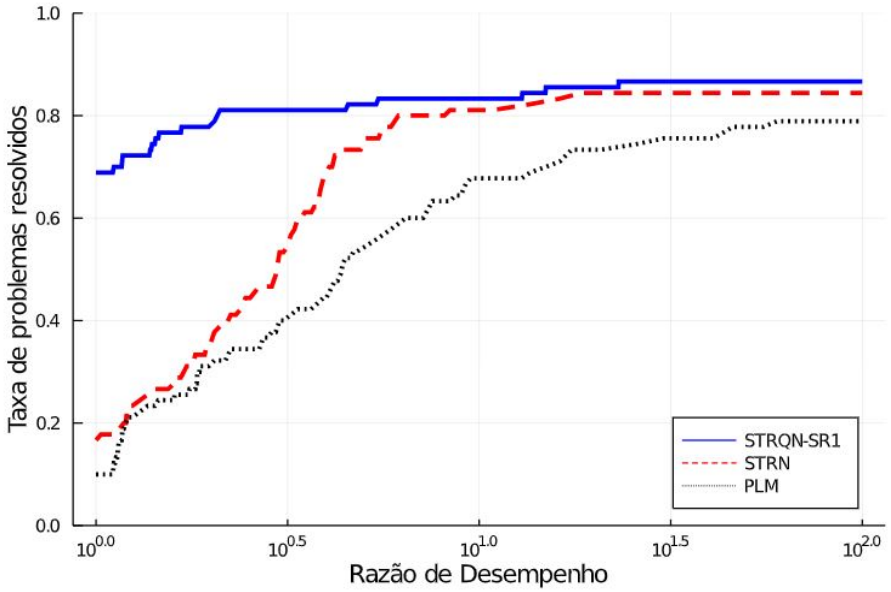

Figura 8 – Perfil de desempenho baseado no tempo de execução dos métodos STRQN-SR1, STRN e PLM.

Verificamos, de imediato, que as menores taxas de eficiência e de robustez foram do método PLM, já que resolveu 83.33% dos problemas testados e, dentre os que tiveram solução identificada por pelo menos um dos métodos, levou menor tempo de resolução do que os demais solucionadores em apenas 10% dos casos. Tal fato reforça a ideia de que, em geral, a estratégia de globalização bastante simples do método PLM, baseada no gradiente projetado, não é tão eficiente quanto a estratégia de pontos interiores usada nos métodos STRN e STRQN-SR1. Neste sentido, o processo iterativo do método PLM mostrou-se altamente custoso, do ponto de vista computacional, para problemas de grande dimensão, e fez com que a maior parte das falhas do método fosse declarada por exceder o tempo máximo de resolução estipulado. No entanto, vale salientar que o método PLM foi introduzido com o objetivo de resolver problemas mais gerais do que os considerados neste estudo, podendo ser utilizado em sistemas não lineares, quadrados ou indeterminados, com qualquer tipo de restrição convexa.

Apesar de ter um índice alto de robustez (85.56%) considerando os problemas testados, o método STRN foi o método intermediário no quesito eficiência, já que foi mais rápido que os demais solucionadores em apenas 16.67% dos problemas que tiveram alguma solução identificada. Conforme os dados apresentado na Tabela 2, em grande parte dos problemas com soluções interiores foi possível observar a convergência quadrática do método (garantida sob as devidas condições), principalmente em razão de uma quantidade inferior de iterações e de avaliações da função com relação ao método STRQN-SR1 (cuja convergência é superlinear).

Com um índice de robustez próximo ao do método STRN, o método STRQN-SR1 reali-

zou seu processo iterativo com sucesso em aproximadamente 87.78% dos problemas testados. Porém, diferentemente dos demais solucionadores, este teve destaque por sua alta taxa de eficiência, uma vez que levou menor tempo de resolução em 70% dos problemas que tiveram soluções identificadas. Apesar de não usufruir da mesma velocidade de convergência do método STRN, o método STRQN-SR1 combinado com poucos cálculos da matriz Jacobiana forneceu um compromisso satisfatório entre a precisão dos resultados e a redução do tempo computacional em relação ao método STRN.

Evidentemente, quando o método STRQN-SR1 é forçado a realizar muitos *step-backs* ou cálculos da Jacobiana no círculo interno, a eficiência do método tende cair consideravelmente. Com exceção do problema 8, a estratégia de *step-back* foi necessária apenas nos dois únicos problemas com solução identificada na fronteira da caixa (problemas 14 e 16), o que fez a quantidade de cálculos da matriz J ficar mais elevada que os demais casos e indicou a dificuldade do método STRQN-SR1 em identificar soluções nesta região da caixa. Por outro lado, nos problemas em que a solução estava mais distante da fronteira a convergência superlinear do método STRQN-SR1 para soluções interiores, garantida no Capítulo 3 sob as devidas condições, demonstrou-se presente. Tal fato foi observado na prática mediante um número ligeiramente maior de iterações e avaliações de função em relação ao método STRN, o que tornou o referido algoritmo relevante para a resolução de sistemas quadrados de equações não lineares restritos a caixas.

Dentre os problemas-teste utilizados, o único que não foi resolvido por nenhum dos métodos foi o problema 7. Este é caracterizado por um escalamento ruim, dificultando a identificação da única solução existente no conjunto viável $0 \leq x_i \leq 10$, $i = 1, \ldots, n$. Neste sentido, a presença de termos exponenciais atrelados com altas constantes em suas equações resultam em uma variação muito grande nos valores da função e das suas derivadas quando são efetuadas pequenas variações nas entradas do sistema linear presente nos métodos. Além disto, a matriz Jacobiana do problema fica muito próxima de ser singular em grande parte do conjunto viável, tornando o referido sistema extremamente difícil de resolver.

Conforme já destacado por Fletscher [10], problemas com alto grau de não linearidade podem ser vistos como dificultosos para métodos baseados em modelos quadráticos Gauss-Newton e Quase-Newton, já que nestes casos os resíduos relativos à concordância do modelo com a função objetivo costumam ser grandes. Como consequência, a convergência a uma solução do problema pode ser lenta, aumentando, inclusive, as chances de ocorrência de falhas ao longo do processo iterativo.

Por fim, embora o solucionador proposto tenha apresentado um desempenho satisfatório, ele ainda pode ser estudado e aprimorado do ponto de vista teórico e computacional. Além disto, destacamos que não existe um método numérico para sistemas não lineares quadrados restritos a caixas que seja ideal para todos os problemas. Nesta perspectiva, a escolha do melhor solu-

cionador depende diretamente de uma análise prévia da estrutura, comportamento e geometria dos sistemas e de suas respectivas restrições; requer conhecimento matemático e senso crítico do pesquisador, a fim de que realize as escolhas mais adequadas a partir da relação entre teoria e prática.

5 CONCLUSÃO

O desenvolvimento de um método de região de confiança escalado secante baseado na abordagem afim-escala proposta por Coleman e Li [6] para a resolução de sistemas quadrados de equações não lineares sujeitos a restrições de caixa mostra-se, naturalmente, uma tarefa complexa. Sabemos que a construção da matriz afim-escala é diretamente dependente da matriz Jacobiana do problema. Desta forma, fazem-se necessárias boas atualizações secantes para tal matriz; mais ainda, quando estas aproximações não são suficientemente boas, a necessidade de calcular a Jacobiana torna-se evidente, aumentando o custo computacional necessário para a resolução do problema.

Apesar das adversidades, a proposição do método STRQN, mostrou-se atrativa do ponto de vista teórico e computacional. Sob condições adequadas, a convergência global e superlinear alcançadas pelo método para soluções do sistema no interior da caixa evidenciaram sua viabilidade e relevância. Os resultados numéricos comprovaram a dificuldade do método em identificar soluções na ou próximas da fronteira da caixa, sendo necessário, em alguns casos, recorrer a estratégias que substituíssem a matriz Quase-Newton (secante) pela Jacobiana numérica.

Com relação à matriz Quase-Newton, a utilização de atualizações relativas à fórmula SR1 mostraram-se mais efetivas do que as tradicionais atualizações BFGS e Broyden, e, desta forma, evidenciaram boas aproximações em relação à Jacobiana exata. Para a maioria dos casos, a convergência superlinear foi observada na prática nas proximidades de soluções no interior da caixa. A posterior comparação do método STRQN-SR1 com o método de região de confiança afim-escala STRN e com o método projetivo PLM evidenciou sua alta taxa de eficiência e robustez perante a resolução de sistemas não lineares restritos a caixas.

Além de uma convergência, em geral, rápida, as vantagens do método STRQN se expandem com a redução de tempo computacional decorrente do evitamento do cálculo da Jacobiana numérica a cada iteração. Na maior parte dos problemas testados, este fator foi decisivo para que o método obtivesse êxito no processo de otimização em um tempo menor que os métodos STRN e PLM. Por outro lado, assim como enfrentado por estes dois solucionadores, a necessidade de resolução de um sistema linear de ordem n a cada iteração para a obtenção do passo mostrou-se um procedimento altamente custoso, principalmente em sistemas onde n é grande.

Neste contexto, a substituição da resolução do referido sistema linear por uma estratégia de atualizações sucessivas de B_k^{-1} ou de aproximações simples destas inversas a partir de técnicas de memória limitada mostra-se um desafio digno de estudos futuros que busquem aprimorar o método STRQN. Especificando-se ao campo teórico, a convergência do método para pontos estacionários do problema de minimização de $f(x)$ em Ω ainda poderá ser estudada e, se garantida sob as devidas condições, os resultados de convergência poderão ser aprimorados. Além disto, variações de raios de região de confiança e de atualizações secantes, implementação de estraté-

gias de não monotonia, e expansão do método para sistemas não lineares indeterminados com restrições de caixa são estudos que podem proporcionar resultados numéricos ainda melhores ao método STRQN.

REFERÊNCIAS

[1] S. Bellavia, M. Macconi e B. Morini. “An affine scaling trust-region approach to bound-constrained nonlinear systems”. *In*: *Applied Numerical Mathematics* 44 (2003), pp. 257–280.

[2] S. Bellavia e B. Morini. “An interior global method for nonlinear systems with simple bounds”. *In*: *Optimization Methods and Software* 20 (2005), pp. 1–22.

[3] S. Bellavia e S. Pieraccini. “On affine-scaling inexact dogleg methods for bound-constrained nonlinear systems”. *In*: *Optimization Methods and Software* 30 (2015), pp. 276–300.

[4] E. G. Birgin, N. Krejic e J. M. Martínez. “Globally convergent inexact quasi-Newton methods for solving nonlinear systems”. *In*: *Numerical Algorithms* 32 (2003), pp. 249–260.

[5] C.G. Broyden, J.E. Dennis e J.J. Moré. “On the Local and Superlinear Convergence of Quasi-Newtqn Methods”. *In*: *IMA Journal of Applied Mathematics* 12 (1973), pp. 223–245.

[6] T. F. Coleman e Y. Li. “An interior trust region approach for nonlinear minimization subject to bounds”. *In*: *SIAM Journal on Optimization* 6 (1996), pp. 418–445.

[7] J. E. Dennis e H. F. Walker. “Convergence Theorems for Least-Change Secant Update Methods”. *In*: *SIAM Journal on Numerical Analysis* 18 (1981), pp. 949–987.

[8] J.E. Dennis e R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. 1ª ed. Englewood Cliffs: Prentice Hall, 1983.

[9] E. D. Dolan e J. J. Moré. “Benchmarking optimization software with performance profiles”. *In*: *Mathematical Programming* 91(2) (2002), pp. 201–213.

[10] R. Fletscher. *Practical Methods of Optimization*. 2ª ed. New York: John Wiley e Sons, 1987.

[11] C.A. Floudas *et al. Handbook of test problems in local and global optimization*. 1ª ed. Dordrecht: Kluwer Academic Publishers, 1999.

[12] J. B. Francisco, N. Krejic e J. M. Martínez. “An interior-point method for solving box-constrained underdetermined nonlinear systems”. *In*: *Journal of Computational and Applied Mathematics* 177 (2005), pp. 67–88.

[13] J. Gao e D. Zhu. "An affine scaling derivative-free trust region method with interior backstracking technique for bounded-constrained nonlinear programming". *In*: *Journal of Systems Science and Complexity* 27(3) (2014), pp. 537–564.

[14] G.H. Golub e C.F. Van Loan. *Matrix computations*. 3ª ed. Baltimore: Johns Hopkins University Press, 1996.

[15] C. Kanzow e A. Klug. "An interior-point affine-scaling trust-region method for semismooth equations with box constraints". *In*: *Computational Optimization and Applications* 37 (2007), pp. 329–353.

[16] C. Kanzow e S. Petra. "Projected filter trust region methods for a semismooth least squares formulation of mixed complementarity problems". *In*: *Optimization Methods and Software* 22 (2007), pp. 713–735.

[17] C. Kanzow, N. Yamashita e M. Fukushima. "Levenberg–Marquardt methods for constrained nonlinear equations with strong local convergence properties". *In*: *Journal of Computational and Applied Mathematics* 172 (2004), pp. 375–397.

[18] C.T. Kelley. *Iterative Methods for Linear and Nonlinear Equations*. 1ª ed. Philadelphia: SIAM, 1995.

[19] M. Kimiaei. "A new class of nonmonotone adaptive trust-region methods for nonlinear equations with box constraints". *In*: *Calcolo* 54 (2017), pp. 769–812.

[20] R. Leveque. *Finite Difference Methods for ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems*. 1ª ed. Seattle: SIAM, 2007.

[21] D. Li e D. Zhu. "An affine scaling interior trust-region method combining with line search filter technique for optimization subject to bounds on variables". *In*: *Numerical Algorithms* 77 (2017), pp. 1159–1182.

[22] E.L. Lima. *Curso de Análise - Vol. 2*. 1ª ed. Rio de Janeiro: IMPA, 2014.

[23] M. Macconi, B. Morini e M. Porcelli. "A Gauss–Newton method for solving bound-constrained underdetermined nonlinear systems". *In*: *Optimization Methods and Software* 24 (2009), pp. 219–235.

[24] L. Marini, B. Morini e M. Porcelli. "Quasi-Newton methods for constrained nonlinear systems: complexity analysis and applications". *In*: *Computational Optimization and Applications* 71 (2018), pp. 147–170.

[25] J.M. Martínez e S.A. Santos. *Métodos Computacionais de Otimização*. 1ª ed. Campinas: IMECC-UNICAMP, 1998.

[26] B. Morini e M. Porcelli. "TRESNEI, a Matlab trust-region solver for systems of nonlinear equalities and inequalities". *In*: *Computational Optimization and Applications* 51 (2010), pp. 27–49.

[27] J. Nocedal e S.J. Wright. *Numerical Optimization*. 2ª ed. New York: Springer, 2006.

[28] F. R. Oliveira. "Methods for Constrained Nonlinear Systems: Inexact Newton-like Conditional Gradient and Levenberg-Marquardt with Inexact Projections". Tese (Doutorado em Matemática). Goiânia: PPGIME - UFG, 2019, p. 76.

[29] J.M. Ortega e W.C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. 1ª ed. London: Academic Press, 1970.

[30] M. Porcelli. "On the convergence of an inexact Gauss–Newton trust-region method for nonlinear least-squares problems with simple bounds". *In*: *Optimization Letters* 7 (2013), pp. 447–465.

[31] R. B. Schnabel e P. D. Frank. "Tensor methods for nonlinear equations". *In*: *SIAM Journal on Numerical Analysis* 21 (1984), pp. 815–843.

[32] V. E. Shamanski. "A modification of Newton's method". *In*: *Ukrain Mat. Z.* 19 (1967), pp. 133–138.

[33] X. Wang e Y. Yuan. "A trust region method based on a new affine scaling technique for simple bounded optimization". *In*: *Optimization Methods and Software* 28(4) (2011), pp. 871–888.

[34] N. Yamashita e M. Fukushima. "On the rate of convergence of the Levenberg-Marquardt Method". *In*: *Computing* 15 (2001), pp. 239–249.

[35] G. Yuan e Z. Wei. "A BFGS trust-region method for nonlinear equations". *In*: *Computing* 92 (2011), pp. 317–333.

[36] M. Zeng e H. Fu. "A Broyden Trust Region Quasi-Newton Method for Nonlinear Equations". *In*: *IAENG International Journal of Computer Science* 46 (2019), pp. 454–458.

[37] J. Zhang e Y. Wang. "A new trust region method for nonlinear equations". *In*: *Mathematical Methods of Operation Research* 58 (2003), pp. 283–298.

[38] D. Zhu. “Affine scaling interior Levenberg–Marquardt method for bound-constrained semismooth equations under local error bound conditions”. *In*: *Journal of Computational and Applied Mathematics* 219 (2008), pp. 198–215.

APÊNDICE A – PERFIS DE DESEMPENHO

Os perfis de desempenho foram desenvolvidos, inicialmente, por Dolan e Moré [9], com o objetivo de proporcionar em um único gráfico a representação da eficiência e robustez de um conjunto S, de n_s algoritmos, em um conjunto P, de n_p problemas. Para que compreendamos sua estruturação, definimos $t_{p,s}$ como o tempo de processamento da resolução do problema $p \in P$ pelo algoritmo $s \in S$. Além disto, assumimos que pelo menos um algoritmo resolve o problema p, e, nos casos em que o algoritmo não consegue resolver o problema p, definimos $t_{p,s} = \infty$.

Com as definições iniciais feitas, passamos a considerar a razão de desempenho de um método s em um problema p, dada por

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s} : s \in S\}}.$$

Por consequência desta definição, temos que o(s) método(s) mais eficiente(s) para um determinado problema tem razão de desempenho um; em contrapartida, os demais métodos têm razão maior que um.

A razão de desempenho é um indicador adequado para analisar o desempenho dos algoritmos, separadamente, em cada problema. Porém, para obter uma avaliação geral do desempenho dos solucionadores em uma classe de problemas, precisamos definir o desempenho do método s como uma probabilidade empírica construída a partir da série de tempos de razão de desempenho $\{r_{p,s} : p \in P\}$.

Tomemos $\tau_M > 1$ e $\tau \in [1, \tau_M]$. Definimos, então,

$$\rho_s(\tau) = \frac{1}{n_P}\#\{p \in P : r_{p,s} \leq \tau\},$$

onde # representa a cardinalidade do conjunto em questão.

Assim, $\rho_s(\tau)$ representa a probabilidade do solucionador $s \in S$ resolver um problema $p \in P$, levando em consideração que a razão de desempenho $r_{p,s}$ seja menor ou igual que τ. Os métodos que alcançarem uma probabilidade alta para menores valores de τ são considerados eficientes. Além disso, para valores não muito altos de τ, quanto mais próximo de 1 for o valor de $\rho_s(\tau)$ mais robusto é considerado o algoritmo s. Finalmente, o gráfico denominado perfil de desempenho é a função de distribuição cumulativa de $\rho_s(\tau)$, cuja finalidade principal consiste em indicar, de forma sintetizada, quais programas podem ser considerados os mais robustos e quais os mais eficientes para o conjunto de problemas S.

Convém ressaltar que a escolha de τ_M deve ser feita com cautela, uma vez que um valor pequeno demais pode não capturar o comportamento total do programa com relação ao conjunto de problemas testes. Considerando que o intervalo $[1, \tau_M]$ pode ser grande, é usual traçarmos o perfil de desempenho em escala logarítmica. Definimos, neste caso,

$$\rho_s(\tau) = \frac{1}{n_P}\#\{p \in P : \log\left(r_{p,s}\right) \leq \tau\},$$

ampliando, desta forma, a região abrangida pelo gráfico.

APÊNDICE B – TABELA DE DESEMPENHO DOS MÉTODOS STRQN

Tabela 3 – Desempenho dos métodos STRQN-BFGS, STRQN-SR1 e STRQN-Broyden.

Prb	κ	STRQN-BFGS					STRQN-SR1					STRQN-Broyden				
		IT	AF	S/C	QN/D	TE	IT	AF	S/C	QN/D	TE	IT	AF	S/C	QN/D	TE
1	1	9	24	0/1	9/0	0.10	9	19	0/0	7/2	0.07	9	19	0/0	7/2	0.08
	2	9	17	0/0	7/2	0.07	8	13	0/0	7/1	0.05	9	14	0/0	8/1	0.06
	3	11	12	0/0	11/0	0.06	9	10	0/0	9/0	0.05	11	12	0/0	11/0	0.06
2	1	F4	—	—	—	—	37	91	0/1	24/13	0.16	F4	—	—	—	—
	2	F4	—	—	—	—	34	82	0/1	27/7	0.14	52	214	0/2	35/17	0.32
	3	F4	—	—	—	—	42	90	0/1	29/13	0.18	65	235	0/2	46/19	0.32
3	1	8	21	0/0	6/2	0.07	6	28	0/1	5/1	0.08	F4	—	—	—	—
	2	F6	—	—	—	—	F4	—	—	—	—	F4	—	—	—	—
	3	F4	—	—	—	—	F4	—	—	—	—	F4	—	—	—	—
4	1	22	121	1/2	4/18	0.22	11	30	0/1	5/6	0.10	7	11	0/0	4/3	0.07
	2	16	98	0/3	9/7	0.17	13	24	0/0	12/1	0.07	9	10	0/0	9/0	0.05
	3	10	40	0/1	7/3	0.10	9	45	0/2	7/2	0.12	7	33	0/2	6/1	0.10
5	1	F6	—	—	—	—	10	17	0/0	6/4	0.07	15	16	1/0	8/7	0.10
	2	11	54	0/1	3/8	0.26	7	8	0/0	6/1	0.04	8	31	0/1	5/3	0.07
	3.5	13	35	0/1	9/4	0.11	13	32	0/1	10/3	0.10	14	33	0/1	11/3	0.11
6	1	F6	—	—	—	—	F4	—	—	—	—	F4	—	—	—	—
	2	25	82	0/3	10/15	0.12	14	38	0/2	9/5	0.09	21	26	1/0	10/11	0.10
	3	F4	—	—	—	—	19	36	0/0	12/7	0.10	F4	—	—	—	—
7	1	F2	—	—	—	—	F2	—	—	—	—	F1	—	—	—	—
	2	F2	—	—	—	—	F2	—	—	—	—	F1	—	—	—	—
	3	F2	—	—	—	—	F2	—	—	—	—	F1	—	—	—	—
8	1	36	43	0/0	33/3	0.06	35	40	0/0	30/5	0.05	28	73	0/2	25/3	0.07
	2	76	88	0/0	40/36	0.14	94	597	0/1	9/85	0.54	F4	—	—	—	—
	3	106	107	0/0	36/70	0.28	108	121	1/1	53/55	0.27	F4	—	—	—	—
9	1	4	5	0/0	3/1	0.01	4	5	0/0	3/1	0.01	4	5	0/0	3/1	0.01
	2	17	18	0/0	15/2	0.02	17	18	0/0	15/2	0.02	17	18	0/0	15/2	0.02
	3	24	25	0/0	22/2	0.03	24	25	0/0	22/2	0.03	24	25	0/0	22/2	0.03
10	1	9	10	0/0	9/0	0.08	6	18	0/1	6/0	0.11	10	11	0/0	10/0	0.10
	2	4	5	0/0	4/0	0.07	3	4	0/0	3/0	0.06	3	4	0/0	3/0	0.06
	3	13	14	0/0	13/0	0.09	13	22	0/0	12/1	0.12	14	15	0/0	14/0	0.10
11	1	97	110	0/0	25/72	1.68	91	118	0/0	17/74	1.60	85	124	0/1	11/74	1.95

Tabela 3 – Desempenho dos métodos STRQN-BFGS, STRQN-SR1 e STRQN-Broyden.

Prb	κ	STRQN-BFGS					STRQN-SR1					STRQN-Broyden				
		IT	AF	S/C	QN/D	TE	IT	AF	S/C	QN/D	TE	IT	AF	S/C	QN/D	TE
	2	2	3	0/0	2/0	0.26	2	3	0/0	2/0	0.25	2	3	0/0	2/0	0.25
	3	107	192	0/0	26/81	1.94	83	119	0/1	16/67	2.05	96	176	0/1	12/84	2.16
12	1	28	48	0/1	27/1	125.04	36	161	0/2	23/13	187.25	33	88	0/2	29/4	185.53
	2	30	206	0/1	4/26	180.95	14	49	0/1	9/5	99.39	29	140	0/1	13/16	167.88
	3	51	72	0/1	50/1	187.70	23	35	0/1	23/0	93.10	23	35	0/1	23/0	93.19
13	1	51	65	0/1	40/11	0.98	47	134	0/0	18/29	0.82	73	240	0/1	22/51	1.29
	2	11	12	0/0	7/4	0.37	11	12	0/0	8/3	0.33	11	14	0/0	7/4	0.38
	3	46	88	0/0	32/14	0.75	42	131	0/0	14/28	0.90	65	224	0/1	18/47	1.26
14	1	F6	—	—	—	—	35	36	20/0	31/4	18.33	31	32	8/0	28/3	9.50
	2	34	35	23/0	27/7	22.31	28	29	9/0	22/6	8.99	29	30	8/0	28/1	9.67
	3	33	65	6/2	23/10	8.98	35	36	4/0	24/11	5.35	36	37	12/0	34/4	12.37
15	1	124	545	0/3	18/106	7.90	21	63	0/2	14/7	2.71	78	343	0/2	29/49	5.22
	2	68	261	0/2	25/43	4.80	39	130	0/2	24/15	3.72	125	622	0/2	31/94	7.91
	3	F2	—	—	—	—	24	64	0/2	19/5	3.01	45	176	0/2	25/20	3.81
16	1	F4	—	—	—	—	24	70	4/3	20/4	2.20	F4	—	—	—	—
	2	F4	—	—	—	—	F4	—	—	—	—	F4	—	—	—	—
	3	F4	—	—	—	—	F4	—	—	—	—	F4	—	—	—	—
17	1	34	46	0/1	34/0	16.48	18	34	0/1	17/1	11.62	26	67	0/1	22/4	14.60
	2	66	110	0/1	60/6	32.80	F4	—	—	—	—	64	289	0/1	22/42	32.61
	3	F4	—	—	—	—	24	37	0/1	23/1	12.84	30	59	0/1	27/3	16.55
18	1	26	38	0/1	15/11	17.38	24	25	0/0	24/0	12.29	F4	—	—	—	—
	2	25	26	0/0	14/11	11.90	26	27	0/0	26/0	12.33	F4	—	—	—	—
	3	26	27	0/0	16/10	12.77	27	28	0/0	27/0	12.28	F4	—	—	—	—
19	1	7	8	0/0	3/4	3.29	7	8	0/0	3/4	3.19	7	8	0/0	3/4	3.29
	2	17	18	0/0	16/1	3.61	17	18	0/0	16/1	3.57	17	18	0/0	16/1	3.64
	3	19	62	0/1	5/14	6.80	19	62	0/1	5/14	6.71	19	62	0/1	5/14	6.89
20	1	5	10	0/0	3/2	3.19	5	10	0/0	3/2	3.19	5	10	0/0	3/2	3.22
	2	32	33	0/0	28/4	4.26	33	50	0/1	28/5	6.94	33	50	0/1	28/5	7.83
	3	21	44	0/1	7/14	10.96	21	44	0/1	7/14	11.32	21	44	0/1	7/14	11.32
21	1.5	20	21	0/0	16/4	6.45	20	21	0/0	16/4	6.22	20	21	0/0	16/4	6.86
	2	77	93	0/1	12/65	9.34	77	93	0/1	12/65	9.19	77	93	0/1	13/64	9.05
	3	26	27	0/0	12/14	6.46	26	27	0/0	12/14	7.10	27	28	0/0	13/14	6.57
22	1	30	31	0/0	1/29	13.42	30	31	0/0	1/29	13.85	30	31	0/0	1/29	12.97

Tabela 3 – Desempenho dos métodos STRQN-BFGS, STRQN-SR1 e STRQN-Broyden.

Prb	κ	STRQN-BFGS					STRQN-SR1					STRQN-Broyden				
		IT	AF	S/C	QN/D	TE	IT	AF	S/C	QN/D	TE	IT	AF	S/C	QN/D	TE
	2	20	21	0/0	1/19	9.24	20	21	0/0	1/19	9.88	20	21	0/0	1/19	10.77
	3	10	11	0/0	1/9	6.51	10	11	0/0	1/9	6.31	9	10	0/0	1/8	5.98
23	1	70	71	0/0	5/65	764.42	68	69	0/0	1/67	673.46	63	64	0/0	1/62	616.03
	2	174	175	0/0	3/171	1653.12	180	181	0/0	1/179	1770.83	180	181	0/0	1/179	1781.16
	3	315	316	0/0	1/314	3004.73	328	329	0/0	1/327	3240.75	328	329	0/0	1/327	3106.79
24	1	74	401	0/7	21/53	142.88	108	504	0/5	23/85	161.57	253	1117	0/2	21/232	305.42
	2	17	114	0/4	10/7	54.86	260	1667	0/3	13/247	343.92	179	1208	0/3	11/168	238.53
	3	36	100	0/2	24/12	67.86	27	56	0/1	24/3	46.85	222	1034	0/1	70/152	294.75
25	1	19	20	0/0	18/1	542.28	19	20	0/0	18/1	408.56	19	20	0/0	18/1	424.64
	2	5	6	0/0	5/0	159.85	4	5	0/0	4/0	72.95	4	5	0/0	4/0	85.13
	3	20	21	0/0	19/1	610.50	20	21	0/0	19/1	442.88	20	21	0/0	19/1	591.47
26	1	66	97	0/1	61/5	3.23	21	57	0/1	16/5	1.32	54	148	0/1	34/20	2.20
	2	72	444	0/1	3/69	2.87	F2	—	—	—	—	419	3451	0/1	80/339	12.89
	3	F6	—	—	—	—	F6	—	—	—	—	F4	—	—	—	—
27	1	108	652	0/1	25/83	1.19	75	459	0/2	15/60	0.78	117	632	0/3	26/91	1.07
	2	F6	—	—	—	—	10	34	0/1	8/2	0.28	32	154	0/1	13/19	0.42
	3	37	261	0/7	11/26	1.00	33	131	0/3	16/17	0.64	66	254	0/2	26/40	0.66
28	1	31	60	0/1	27/4	0.84	30	59	0/1	26/4	0.75	60	208	0/1	28/32	1.05
	2	26	44	0/1	25/1	0.73	33	67	0/1	29/4	0.77	241	1309	0/1	31/210	3.11
	3	15	27	0/1	15/0	0.62	15	27	0/1	15/0	0.66	18	41	0/1	15/3	0.74
29	1	25	75	0/2	17/8	100.59	25	59	0/1	19/6	94.71	27	55	0/1	20/7	93.56
	2.5	33	168	0/5	14/19	157.24	22	42	0/1	20/2	81.69	22	45	0/1	19/3	93.02
	3	22	40	0/1	21/1	83.76	25	59	0/1	19/6	93.20	27	55	0/1	20/7	102.75
30	1	15	30	0/1	14/1	201.53	14	26	0/1	14/0	195.74	30	39	0/1	28/2	364.21
	2	9	21	0/1	9/0	209.09	11	23	0/1	11/0	237.03	29	114	0/1	18/11	659.05
	3	13	67	0/1	8/5	349.04	11	12	0/0	11/0	222.08	12	13	0/0	12/0	230.35

Legenda:

IT - Quantidade de iterações realizadas;

AF - Quantidade de avaliações da função F realizadas (não inclui as avaliações necessárias para calcular J);

S - Quantidade de *step-backs* realizados;

C - Quantidade de cálculos de J no círculo interno realizados;

QN - Quantidade de passos Quase-Newton tomados;

D - Quantidade de passos Dogleg tomados;

TE - Tempo de execução do algoritmo.

APÊNDICE C – TABELA DE FALHAS DOS MÉTODOS STRQN

Tabela 4 – Falhas dos métodos STRQN.

Método	Prb	κ	Falha	$\|F(x_0)\|$	$\|F(x^*)\|$	$\left\|D(x^*)^{-1}\nabla f(x^*)\right\|$
STRQN-BFGS	2	1	F4	3.9×10^{4}	9.7×10^{3}	5.6×10^{7}
		2	F4	3.1×10^{5}	1.2×10^{0}	4.7×10^{2}
		3	F4	1.0×10^{6}	1.4×10^{0}	7.5×10^{2}
	3	2	F6	2.1×10^{5}	1.0×10^{-3}	2.0×10^{-5}
		3	F4	4.7×10^{5}	3.9×10^{5}	1.1×10^{11}
	5	1	F6	2.4×10^{1}	1.2×10^{1}	9.4×10^{1}
	6	1	F6	1.3×10^{0}	1.6×10^{-1}	1.4×10^{-1}
		3	F4	1.6×10^{0}	7.5×10^{-1}	7.1×10^{-1}
	7	1	F2	4.1×10^{2}	1.1×10^{1}	3.3×10^{3}
		2	F2	1.2×10^{3}	1.2×10^{1}	2.5×10^{1}
		3	F2	2.9×10^{3}	1.2×10^{1}	4.5×10^{0}
	14	1	F6	6.0×10^{3}	5.2×10^{3}	6.8×10^{5}
	15	3	F2	7.0×10^{3}	5.5×10^{-3}	4.8×10^{-4}
	16	1	F4	2.0×10^{3}	4.8×10^{-1}	3.9×10^{0}
		2	F4	1.1×10^{4}	2.2×10^{1}	9.7×10^{3}
		3	F4	3.4×10^{4}	4.0×10^{1}	2.6×10^{4}
	17	3	F4	6.6×10^{4}	2.1×10^{2}	4.4×10^{3}
	26	3	F6	1.7×10^{3}	8.3×10^{2}	5.0×10^{4}
	27	2	F6	2.1×10^{5}	3.2×10^{4}	7.2×10^{8}
STRQN-SR1	3	2	F4	2.1×10^{5}	1.3×10^{-1}	4.4×10^{4}
		3	F4	4.7×10^{5}	3.9×10^{5}	1.1×10^{11}
	6	1	F4	1.3×10^{0}	7.5×10^{-1}	5.0×10^{-1}
	7	1	F2	4.1×10^{2}	6.0×10^{0}	9.7×10^{0}
		2	F2	1.2×10^{3}	6.8×10^{0}	8.3×10^{1}
		3	F2	2.9×10^{3}	1.2×10^{1}	2.3×10^{0}
	16	2	F4	1.1×10^{4}	2.2×10^{1}	9.8×10^{3}
		3	F4	3.4×10^{4}	3.9×10^{1}	2.7×10^{4}
	17	2	F4	1.7×10^{4}	2.6×10^{1}	3.1×10^{2}
	26	2	F2	5.5×10^{1}	2.7×10^{0}	4.1×10^{-3}
		3	F6	1.7×10^{3}	2.6×10^{1}	5.0×10^{4}
STRQN-Broyden	2	1	F4	3.9×10^{4}	5.7×10^{-1}	1.6×10^{2}
	3	1	F4	5.2×10^{4}	6.3×10^{-1}	9.1×10^{4}

Tabela 4 – Falhas dos métodos STRQN.

Método	Prb	κ	Falha	$\|F(x_0)\|$	$\|F(x^*)\|$	$\left\|D(x^*)^{-1}\nabla f(x^*)\right\|$
		2	F4	2.1×10^{5}	3.2×10^{-3}	1.2×10^{3}
		3	F4	4.7×10^{5}	3.9×10^{5}	1.1×10^{11}
	6	1	F4	1.3×10^{0}	3.0×10^{-1}	3.5×10^{-1}
		3	F4	1.6×10^{0}	7.5×10^{-1}	7.1×10^{-1}
	7	1	F1	4.1×10^{2}	4.3×10^{0}	3.0×10^{0}
		2	F1	1.2×10^{3}	6.3×10^{0}	1.0×10^{1}
		3	F1	2.9×10^{3}	1.2×10^{1}	8.2×10^{-1}
	8	2	F4	8.3×10^{14}	8.6×10^{-2}	2.6×10^{-1}
		3	F4	1.0×10^{23}	4.2×10^{-2}	1.3×10^{-1}
	16	1	F4	2.0×10^{3}	9.3×10^{0}	6.2×10^{2}
		2	F4	1.1×10^{4}	2.2×10^{1}	1.3×10^{4}
		3	F4	3.4×10^{4}	4.0×10^{1}	1.0×10^{4}
	18	1	F4	1.3×10^{3}	1.2×10^{-2}	1.3×10^{-4}
		2	F4	5.0×10^{4}	3.4×10^{-3}	1.5×10^{-5}
		3	F4	1.1×10^{5}	1.6×10^{-3}	3.9×10^{-6}
	26	3	F6	1.7×10^{3}	8.3×10^{2}	5.0×10^{4}